

A multilab registered replication of the attentional SNARC effect

**Lincoln J Colling<sup>1</sup>, Dénes Szűcs<sup>1</sup>, Damiano De Marco<sup>1,12</sup>, Krzysztof Cipora<sup>2</sup>, Rolf Ulrich<sup>2</sup>, Hans-Christoph Nuerk<sup>2</sup>, Mojtaba Soltanlou<sup>2</sup>, Donna Bryce<sup>2</sup>, Sau-Chin Chen<sup>3</sup>, Philipp Alexander Schroeder<sup>4</sup>, Dion T Henare<sup>5</sup>, Christine K Chrystall<sup>5</sup>, Paul M Corballis<sup>5</sup>, Daniel Ansari<sup>6</sup>, Celia Goffin<sup>6</sup>, H Moriah Sokolowski<sup>6</sup>, Peter JB Hancock<sup>7</sup>, Ailsa E Millen<sup>7</sup>, Stephen RH Langton<sup>7</sup>, Kevin J Holmes<sup>8</sup>, Mark S Saviano<sup>8</sup>, Tia A Tummino<sup>8</sup>, Oliver Lindemann<sup>9</sup>, Rolf A Zwaan<sup>9</sup>, Jiří Lukavský<sup>10</sup>, Adéla Becková<sup>11</sup>, Marek A Vranka<sup>11</sup>, Simone Cutini<sup>12</sup>, Irene Cristina Mammarella<sup>12</sup>, Claudio Mulatti<sup>12</sup>, Raoul Bell<sup>13</sup>, Axel Buchner<sup>13</sup>, Laura Mieth<sup>13</sup>, Jan Philipp Röer<sup>14,2</sup>, Elise Klein<sup>15</sup>, Stefan Huber<sup>15</sup>, Korbinian Moeller<sup>15,2</sup>, Brenda Ocampo<sup>16</sup>, Juan Lupiáñez<sup>17</sup>, Javier Ortiz-Tudela<sup>17</sup>, Juanma De la fuente<sup>17</sup>, Julio Santiago<sup>17</sup>, Marc Ouellet<sup>17</sup>, Edward M Hubbard<sup>18</sup>, Elizabeth Y Toomarian<sup>18</sup>, Remo Job<sup>19</sup>, Barbara Treccani<sup>19</sup>, & Blakeley B McShane<sup>20</sup>**

<sup>1</sup> Department of Psychology, University of Cambridge

<sup>2</sup> Department of Psychology, University of Tübingen

<sup>3</sup> Department of Human Development and Psychology, Tzu-Chi University

<sup>4</sup> Department of Psychiatry and Psychotherapy, University of Tübingen

<sup>5</sup> School of Psychology, University of Auckland

<sup>6</sup> Department of Psychology & Brain and Mind Institute, The University of Western Ontario

<sup>7</sup> Psychology, Faculty of Natural Sciences, University of Stirling, Stirling, UK

<sup>8</sup> Department of Psychology, Colorado College

<sup>9</sup> Department of Psychology, Education & Child Studies, Erasmus University Rotterdam, Netherlands

- 21 <sup>10</sup> Institute of Psychology of the Czech Academy of Sciences
- 22 <sup>11</sup> Department of Psychology, Faculty of Arts, Charles University
- 23 <sup>12</sup> Department of Developmental Psychology, University of Padova
- 24 <sup>13</sup> Department of Experimental Psychology, Heinrich Heine University Düsseldorf
- 25 <sup>14</sup> Department of Psychology and Psychotherapy, Witten/Herdecke University
- 26 <sup>15</sup> Leibniz-Institut für Wissensmedien, Tübingen
- 27 <sup>16</sup> School of Psychology, The University of Queensland
- 28 <sup>17</sup> Research Center for Mind, Brain, and Behavior, University of Granada
- 29 <sup>18</sup> Department of Educational Psychology, University of Wisconsin-Madison
- 30 <sup>19</sup> Department of Psychology and Cognitive Science, University of Trento
- 31 <sup>20</sup> Kellogg School of Management, Northwestern University

32 Author Note

33 Correspondence concerning this article should be addressed to **Dénes Szűcs**, Downing Street,  
34 CB2 3EB, Cambridge, UK. E-mail: ds377@cam.ac.uk

## Abstract

35

36 The attentional Spatial-Numerical Association of Response Codes (att-SNARC) effect (Fischer et al.,  
37 2003; Nature Neuroscience)—the finding that participants are quicker to detect left-side targets when  
38 the targets are preceded by small numbers and quicker to detect right-side targets when they are  
39 preceded by large numbers—has been used as evidence for *embodied* number representations and to  
40 allow for strong claims about the link between number and space (e.g., a mental number line). We  
41 attempted to replicate Study 2 of Fischer et al. (2003) by collecting data from 1105 participants across  
42 seventeen labs. Across all 1105 participants and four ISI conditions, the proportion of times the  
43 direction of the observed effect was consistent with the original effect was 0.50. Further, the effects we  
44 observed both within and across labs were minuscule and incompatible with those observed in Fischer  
45 et al. (2003). Given this, we conclude that we have *failed* to replicate the effect reported by Fischer et al.  
46 (2003). In addition, our analysis of several participant-level moderators (finger counting preferences,  
47 reading/writing direction experience, handedness, and mathematics fluency and mathematics anxiety)  
48 revealed no substantial moderating effects. Our results demonstrate that the att-SNARC effect cannot be  
49 used as evidence to support the strong claims about the link between number and space discussed above

50 A multilab registered replication of the attentional SNARC effect

51 **Introduction**

52 A foundational issue in cognitive science is the question of how we *represent* concepts. Classical  
53 approaches to cognitive science, exemplified by Fodor’s (1975) “language of thought” and Newell and  
54 Simon’s (1976) “physical symbol systems” hypothesis, view representations as abstract or amodal and  
55 as distinct from sensorimotor processing. In contrast to these traditional views, a range of other views  
56 that go under labels such as “embodied”, “situated”, or “grounded” cognition argue that representations  
57 (i) are intimately linked to sensorimotor processing (see, e.g., Wilson, 2002, for an overview); (ii) are  
58 analogue rather than symbolic; and (iii) represent by in some sense resembling their targets (e.g., see  
59 Gładziejewski & Miłkowski, 2017; Williams & Colling, 2018).

60 One area of research that has provided a wealth of empirical findings valuable for debates about  
61 the nature of concept representation has been numerical cognition. In fact, Fischer and Brugger (2011)  
62 have referred to numerical cognition as the “prime example of embodied cognition”. In particular,  
63 Fischer and Brugger (2011) point to tasks examining spatial-numerical associations to make their case.

64 Researchers have long reasoned that numbers might be represented in a spatially organised  
65 manner (Galton, 1880), for example, as a *mental number line* (e.g., Restle, 1970). Key support for this  
66 notion comes from a series of nine experiments conducted by Dehaene, Bossini, and Giraux (1993). In  
67 these experiments, Dehaene et al. (1993) asked participants to judge whether the parity of a number was  
68 odd or even, finding that responses to large numbers were faster when pressing a right-hand key relative  
69 to a left-hand key while the opposite was true for small numbers. They labelled this number magnitude  
70 by response side interaction the Spatial-Numerical Association of Response Codes (SNARC) effect.

71 In these parity judgement experiments, there was no standard with which to compare the  
72 presented number. Consequently, whether a particular number was responded to quicker with the left  
73 hand or the right hand was not determined by the absolute magnitude of the number, but by the relative  
74 magnitude of the number within a stimulus set. Thus, the number five was responded to more quickly

75 with the left hand when appearing in a set of numbers ranging from four to nine but more quickly with  
76 the right hand when appearing in a set of numbers ranging from zero to five (e.g., Dehaene et al., 1993;  
77 Fias, Brysbaert, Geypens, & d'Ydewalle, 1996).

78 Dehaene et al. (1993) reported that the effects were neither dependent on the handedness of  
79 participants nor the hand used to make the response. Instead, they tracked the side of space of the  
80 response, with responses to small numbers being quicker with the right hand when the participants'  
81 hands were crossed (see, however, Wood, Nuerk, & Willmes, 2006). Nonetheless, Dehaene et al. (1993)  
82 did report that the effect was dependent on reading/writing direction. Specifically, while they initially  
83 found the effect in French participants who had experience reading/writing from left to right, they did not  
84 replicate it in a follow-up experiment with Iranian participants who had experience reading/writing from  
85 right to left (see Shaki, Fischer, and Petrusic (2009) and Zebian (2005)). Together, the results of the nine  
86 experiments reported in Dehaene et al. (1993) were taken to support the idea of a mental number line  
87 with numbers of increasing magnitude associated with the left-to-right axis of external space.

88 While SNARC effects appear to be robust (see Wood, Willmes, Nuerk, and Fischer (2008) and  
89 Toomarian and Hubbard (2018) for recent reviews), the great range of findings has resulted in some  
90 debate about the underlying mechanism(s) that produce them. One such debate is concerned with  
91 whether the SNARC effect is produced by early, *response-independent* mechanisms or whether  
92 processes at the stage of *response selection* are responsible. According to theories that place the origin  
93 of the SNARC effect at an early stage, the mere observation of the number should be sufficient to  
94 activate the spatial code because the spatial code is intimately connected to the numerical representation.  
95 Consequently, these theories make the strongest claims about the link between number and space.  
96 Theories that place the origin of the SNARC effect at the response selection stage, however, make  
97 weaker claims about the connection between number and space. As Pecher and Boot (2011) note, if the  
98 response selection stage gives rise to the SNARC effect, then no underlying spatial-numerical  
99 representation need be assumed.

100 Most recent work has tended to support the notion that the response selection stage is the locus of

101 the SNARC effect. In particular, Keus and colleagues have used both behavioural (Keus & Schwarz,  
102 2005) and psychophysiological (Keus, Jenks, & Schwarz, 2005) evidence to argue in favour of a later,  
103 response-related origin of the SNARC effect. Further support comes from a computational model that  
104 relies on task-dependent conceptual coding of the number at a stage distinct from the numerical  
105 representation itself (Gevers, Verguts, Reynvoet, Vaessens, & Fias, 2006).

106 Additional accounts that break the link between number, space, and the SNARC effect are  
107 so-called response polarity-related accounts. Specifically, Proctor and Cho (2006) argue that on binary  
108 classification tasks, items in the task set are coded as being positive or negative in polarity. Response  
109 selection can then be facilitated when there is a structural overlap between the polarity of the item (the  
110 number in the case of the SNARC effect) and the response. As with the model from Gevers et al. (2006),  
111 the account of Proctor and Cho (2006) does not require any perceptual or conceptual overlap between  
112 the stimulus and the response dimensions for the SNARC effect to occur. That is, these accounts do not  
113 rely on the notion of a mental number line or sensorimotor-linked representations. A range of empirical  
114 findings support these types of accounts. For example, Santens and Gevers (2008) found that  
115 SNARC-like effects can be produced when left-right responses are replaced with unimanual close-far  
116 responses, with small numbers associated with close responses and large numbers associated with far  
117 responses. Further, Landy, Jones, and Hummel (2008) found that verbal “Yes” and “No” responses on a  
118 parity judgement task were facilitated by large numbers and small numbers respectively.

119 Finally, still other researchers have argued in favour of a working memory account of the SNARC  
120 effect. For example, in the task reported by van Dijck and Fias (2011), participants performed a  
121 fruit/vegetable classification after having been encouraged to store the stimuli as an ordered set in  
122 working memory. This was done by presenting participants with a sequence of fruit and vegetable  
123 names (displayed in the centre of the screen) before the classification task and then testing them on the  
124 order of the items. A spatial response-compatibility effect emerged with participants responding faster  
125 to items early in the sequence with their left hand and items later in the sequence with the right hand.  
126 van Dijck and Fias (2011) argue that this working memory account can also explain why SNARC-like

127 effect emerge for other kinds of ordinal sequences such as months of the year (Gevers, Reynvoet, &  
128 Fias, 2003) or days of the week (Gevers, Reynvoet, & Fias, 2004) as well as why spatial-numerical  
129 associations can be moderated by giving participants instructions to associate numbers with positions on  
130 a clock-face (1–5 on the right and 6–10 on the left) rather than on a ruler (1–5 on the left and 6–10 on  
131 the right; Bächtold, Baumüller, & Brugger, 1998)

132         Given that several competing accounts of the SNARC effect exist and that many of the accounts  
133 do not require a mental number line, one may doubt whether spatial-numerical associations provide  
134 evidence for anything like “embodied” number representations or number representation that are  
135 intimately linked with space. However, there is evidence that does support an early,  
136 response-independent locus for the SNARC effect and thus does provide support for the notion of a  
137 mental number-line and spatially-linked number representation—the modified version of the Posner  
138 (1980) attentional cueing task developed by Fischer, Castel, Dodd, and Pratt (2003). In this study,  
139 participants were asked to detect the appearance of lateralised targets. The target, a white circle, was  
140 preceded by either a small number (1 or 2) or a large number (8 or 9). Importantly, the digit did not  
141 predict the subsequent location of the target, that is, it was not task-relevant. Instead, the task was  
142 merely to press a single response button when the target appeared regardless of whether it appeared on  
143 the left or the right. Importantly, not requiring a spatially lateralised response negates the possibility of  
144 any response-related effects. The finding from this paradigm was consistent with the SNARC effect, as  
145 participants were quicker to detect left-side targets when the targets were preceded by small numbers  
146 and quicker to detect right-side targets when they were preceded by large numbers, at least for digits and  
147 targets that were separated by an inter-stimulus interval (ISI) between 250 and 1000 ms. This  
148 finding—named the attentional SNARC (att-SNARC) effect—suggests that viewing numbers alone was  
149 able to cue spatial attention either to the left or the right depending on the magnitude of the number.

150         Because the att-SNARC effect argues strongly in favour of an early, response-independent locus  
151 for the cause of the SNARC effect, the att-SNARC effect plays a crucially important role in adjudicating  
152 debates about the origin of the SNARC effect and the nature of number representations. As a result, the

153 original finding has been extremely influential (e.g., cited 704 times according to Google Scholar as of  
154 12 September 2019). However, subsequent attempts to replicate the effect have produced mixed results.

155 Galfano, Rusconi, and Umiltà (2006) report a statistically significant effect for right-side targets  
156 when the data was collapsed across two ISI conditions of 500 and 800 ms using a one-tailed test  
157 [Estimate = 6.5 ms;  $t(25) = 1.75$ ;  $p = .046$  (reported as  $p = .04$ )]. They also report a statistically  
158 significant effect for left-side targets collapsed across the two ISI conditions, but the claimed statistical  
159 significance reflected a reporting error [Estimate = 5.5 ms;  $t(25) = 1.59$ ;  $p = .062$  (reported as  $p = .04$ )].  
160 Finally, they report an overall estimate (collapsed across the left and right target locations) of 8 ms for  
161 the 500 ms ISI condition and 4 ms for the ISI 800 ms condition, but the reporting is such that the  
162 corresponding variances or test statistics for these estimates cannot be obtained.

163 In addition, Dodd, Van der Stigchel, Leghari, Fung, and Kingstone (2008) report a statistically  
164 significant effect when the data was collapsed across three ISI conditions between 250 and 750 ms and  
165 across both left and right target locations, but again the claimed statistical significance reflected a  
166 reporting error [Estimate = 5.5 ms;  $F(1,29) = 4.05$ ;  $p = .054$  (reported as  $p < .05$ )]. At the level of  
167 individual inter-stimulus intervals, they report statistically significant effects at 500 ms for right-side  
168 targets [Estimate = 6 ms;  $t(29) = 2.34$ ;  $p = .013$ ] and left-side targets [Estimate = 16 ms;  $t(29) = 2.48$ ;  $p$   
169 = .010]. Finally, they report estimates of 6 ms for the 250 ms ISI condition, 11 ms for the 500 ms ISI  
170 condition, and -0.5 ms for the 750 ms ISI condition (collapsed across left and right target locations), but  
171 the reporting is such that the corresponding variances or test statistics for these estimates cannot be  
172 obtained.

173 Ristic, Wright, and Kingstone (2006) also report a statistically significant effect when the data  
174 was collapsed across six ISI conditions between 350 and 800 ms and across right and left side targets  
175 [Estimate = 3.79 ms;  $F(1,17) = 5.48$ ;  $p = .032$ ]. Although it is possible to obtain point estimates for each  
176 of the six inter-stimulus intervals [350 ms ISI = 11.24 ms; 400 ms ISI = 2.81 ms; 500 ms ISI = -1.44 ms;  
177 600 ms ISI = 6.17 ms; 700 ms ISI = 6.05 ms; 800 ms ISI = -2.17 ms] (collapsed across left and right  
178 target locations), the reporting is such that the corresponding variances or test statistics for these



179 estimates cannot be obtained.

180 Several other failed replications have also been reported. Zanolie and Pecher (2014) report two  
181 experiments that failed to find a statistically significant effect when collapsed across four ISIs between  
182 250 and 750 ms and across left and right side targets [Experiment 1: No estimates reported;  $F(1,19) =$   
183  $0.03, p = .863$ ; Experiment 2: No estimates reported;  $F(1,23) = 0.13, p = .772$ ]. Ranzini, Dehaene,  
184 Piazza, and Hubbard (2009) also failed to find a statistically significant effect when collapsed across  
185 three ISIs between 300 and 500 ms and across left and right side targets [Estimate = 3 ms;  $F(1,14) = 4.1,$   
186  $p = .06$ ]. Salillas, El Yagoubi, and Semenza (2008) failed to find a statically signifiant effect at a 400 ms  
187 ISI when collapsed across left and right side targets [Estimate = 2 ms;  $F(1,11) = 1.3, p = .28$ ]. **More**  
188 **recently, van Dijck, Abrahamse, Acar, Ketels, and Fias (2014) failed to find an effect collapsed across**  
189 **four ISIs between 250 and 1000 ms and left and right side targets [Experiment 1: No estimates reported]**  
190 **and three ISIs between 100 and 700 ms [Experiment 2: No estimates reported;  $F(1,28) = 2.94, p = .097$ ].**  
191 **While Fattorini, Pinto, Rotondaro, and Doricchi (2015) failed to find an effect collapsed across two ISI**  
192 **of 500 and 750 ms and across left and right side targets [Experiment 1: No estimates reported;  $F(1,59) =$**   
193  **$1.69, p = .20$ ] and four ISIs between 250 and 1000 ms [Experiment 2: No estimates reported;  $F(1,31) =$**   
194  **$1.5, p = .22$ ]. The final two studies by van Dijck et al. (2014) and Fattorini et al. (2015) are particularly**  
195 **notable for their large sample size.**

196 **It should be noted that alternative accounts of the effect reported by Fischer et al. (2003) have**  
197 **been suggested. These alternative accounts include, for example, accounts based on working memory**  
198 **(van Dijck et al., 2014). Similarly, manipulations that make explicit associations between number and**  
199 **space have also been able to produce att-SNARC-like effect (e.g., Fattorini et al., 2015, Experiment 3).**  
200 However, because of these modifications, the findings of these studies have different theoretical  
201 implications to the att-SNARC and, therefore, they will not be discussed here. Instead, the focus of the  
202 present work will be on the att-SNARC as originally proposed.

203 In sum, prior studies have demonstrated—at best—only qualified and partial success at  
204 replicating Fischer et al. (2003). That said, one might argue that failure to replicate Fischer et al. (2003),

205 reflects more the definition of replication employed—namely one based on statistical  
206 significance—than any true failure to replicate the scientific hypothesis as opposed to the statistical  
207 hypothesis examined by Fischer et al. (2003). As we discuss in greater depth below, we are sympathetic  
208 to this view and prefer alternative operationalisations of replication.

209 One component of such a better approach to assessing replication might involve synthesising the  
210 evidence across all published studies of the effect via meta-analysis in order to estimate, for example, an  
211 overall average effect size, the heterogeneity in effect sizes across studies, and the effects of potential  
212 moderators at the study-level or otherwise. However, this is complicated because (i) the statistical  
213 significance of a study's results typically impacts whether or not the study is published therefore  
214 resulting a set of published studies that is not representative and (ii) meta-analytic results are biased  
215 when the set of studies analysed is not representative (McShane, Böckenholt, & Hansen, 2016;  
216 Ioannidis, 2008).

217 Given this, the Registered Replication Report (RRR) format that we pursue here provides an ideal  
218 means of assessing the att-SNARC effect because results from all participating labs are included in the  
219 meta-analysis regardless of the results. Further, pre-registration of the primary hypotheses and statistical  
220 analyses further mitigates many potential biases.

221 An additional benefit of an RRR is that it allows for the investigation of potential moderator  
222 variables not previously considered thereby shedding light on mechanism and perhaps also the current  
223 mixed record of replication success. Consequently, in addition to replicating the experimental protocol  
224 of Fischer et al. (2003), we investigate several variables that could potentially moderate the att-SNARC  
225 effect including finger counting habits, reading/writing direction, handedness, mathematics ability, and  
226 mathematics anxiety (see Fischer (Fischer, 2006; Fischer, 2008), Fischer and Knops (2014), Georges,  
227 Hoffmann, and Schiltz (2016), and Shaki et al. (2009) for details and conjectures).

## Methods

228

### Design

229

#### Sample size

230

231 Each participating lab was required to provide a target sample size and stopping rule on a  
232 lab-specific OSF page (<https://osf.io/7zyxj/>), with labs agreeing to a minimum target sample size of  
233 sixty participants. We chose sixty as the minimum because it provides more than adequate power (0.92  
234 using a one-tailed test at  $\alpha = 0.05$ ) assuming an effect size on the standardised Cohen's  $d$  scale of 0.4,  
235 about the midpoint of previously published estimates. This corresponds to a raw effect size of 6 ms  
236 assuming a between-participant standard deviation of 15 ms, again about the midpoint of previously  
237 published estimates.

238 Due to time constraints, not all labs were able to reach the minimum target of sixty (see Table 1  
239 for sample sizes achieved by each lab). However, again assuming an effect size of 0.4, we would expect  
240 to see a statistically significant effect in 93% of the labs (i.e., about sixteen) given the sample sizes  
241 actually achieved. Given this, if 0.4 is a reasonable estimate of the effect size and there are no  
242 substantial moderators of the effect, we would expect statistically significant effects not only at the  
243 meta-analytic level but also at the level of the individual lab.

### Materials

244

245 The participating labs all had: (i) a testing station, such as a room or a cubicle, where participants  
246 could undertake the experiment without distraction; (ii) a computer for presenting stimuli and recording  
247 responses; (iii) a chin rest or similar device to ensure that the participant remained a set distance from  
248 the computer monitor; and (iv) a tape measure for use in the screen calibration process. Five labs also  
249 optionally made use of an eye-tracker to record participants' eye movements during the replication task;  
250 see the lab-specific OSF pages for details.

251 Additionally, an instruction booklet detailing how to perform the setup and calibration procedure

252 and the finger counting assessment was provided. The materials were initially written in English. The  
253 experiment was also conducted in German, Dutch, Czech, Spanish, Italian, and Chinese, reflecting the  
254 predominant language in the locale of the individual labs; for these labs, the English language  
255 instructions were translated into the new language and then independently back-translated into English  
256 to ensure accuracy.

257 All materials including translations are available on OSF (see <https://osf.io/7zyxj/>). To perform  
258 analyses, we used R (Version 3.5.1; R Core Team, 2018) and the R packages *bindrcpp* (Version 0.2.2;  
259 Müller, 2018), *checkmate* (Version 1.8.5; Lang, 2017), *dplyr* (Version 0.7.6; Wickham, François, Henry,  
260 & Müller, 2018), *forcats* (Version 0.3.0; Wickham, 2018a), *forestplot* (Version 1.7.2; Gordon & Lumley,  
261 2017), *ggplot2* (Version 3.0.0; Wickham, 2016), *glue* (Version 1.3.0; Hester, 2018), *kableExtra* (Version  
262 0.9.0; Zhu, 2018), *knitr* (Version 1.20; Xie, 2015), *lme4* (Version 1.1.18.1; Bates, Mächler, Bolker, &  
263 Walker, 2015), *magick* (Version 1.9; Ooms, 2018), *magrittr* (Version 1.5; Bache & Wickham, 2014),  
264 *Matrix* (Version 1.2.14; Bates & Maechler, 2018), *nlme* (Version 3.1.137; Pinheiro, Bates, DebRoy,  
265 Sarkar, & R Core Team, 2018), *papaja* (Version 0.1.0.9842; Aust & Barth, 2018), *purrr* (Version 0.2.5;  
266 Henry & Wickham, 2018), *pwr* (Version 1.2.2; Champely, 2018), *R.matlab* (Version 3.6.2; Bengtsson,  
267 2018), *readr* (Version 1.1.1; Wickham, Hester, & Francois, 2017), *reticulate* (Version 1.10; Allaire,  
268 Ushey, & Tang, 2018), *stringr* (Version 1.3.1; Wickham, 2018b), *tibble* (Version 1.4.2; Müller &  
269 Wickham, 2018), *tidyr* (Version 0.8.1; Wickham & Henry, 2018), and *tidyverse* (Version 1.2.1;  
270 Wickham, 2017).

## 271 **Procedure**

272 We employed an experimental paradigm based on Experiment 2 of Fischer et al. (2003). We  
273 chose Experiment 2 over Experiment 1 because it had fewer ISI conditions and because the results were  
274 statistically significant in a greater proportion of conditions. Before starting data collection, each lab  
275 performed a monitor calibration procedure using a supplied calibration script which involved measuring  
276 the viewing distance and the size of standard stimuli presented on the screen; see OSF for details. After  
277 participants provided informed consent, they were seated in front of a computer monitor with their

278 heads placed into a chin rest that was located a fixed distance from the monitor (set during the  
279 calibration procedure) and then data collection commenced.

280 The standard trial structure, which is identical to that of Fischer et al. (2003) and which does not  
281 contain timing modifications for the eye-tracker (see below for details), is shown in Figure 1. The initial  
282 display of each trial contained a centrally presented white fixation point on a black background ( $0.2^\circ$   
283 diameter), and two white boxes ( $1^\circ \times 1^\circ$ ) presented on either side of the fixation point. The centres of  
284 the boxes were located  $5^\circ$  from the centre of the fixation point. This initial display was shown for 500  
285 ms. Following the initial display, a digit (1, 2, 8, or 9;  $0.75^\circ$  height) was presented at a fixed duration of  
286 300 ms. After the digit was removed, the fixation point reappeared for a variable duration (250 ms, 500  
287 ms, 750 ms, or 1000 ms). This was followed by a circular white target ( $0.7^\circ$  diameter) appearing in  
288 either the left- or right-side box on target trials or no target appearing on catch trials.

289 Target trials ended after a response was made or 1000 ms after target onset, whichever came first.  
290 Catch trials ended 1000 ms after the digit was removed. Trials automatically advanced and were  
291 separated by an inter-trial interval of 1000 ms.

292 Participants responded by pressing the spacebar with the preferred hand. Participants who  
293 responded before the target appeared or who responded on a catch trial were presented with the warning  
294 “Too quick! Please wait until the target appears in a box before pressing SPACE” [English text] and the  
295 trial ended. Participants who failed to respond on a target trial were presented with the warning “Too  
296 slow! Please press SPACE as soon as the target appears”. Participants who erred on more than 5% of  
297 trials were excluded from analyses.

298 Participants performed a total of 800 trials (640 target trials and 160 catch trials), split into five  
299 blocks of 160 trials each with 128 target trials and 32 catch trials per block; each block contained an  
300 equal number of trials for each ISI, digit, and target location, and these were presented in a random order.

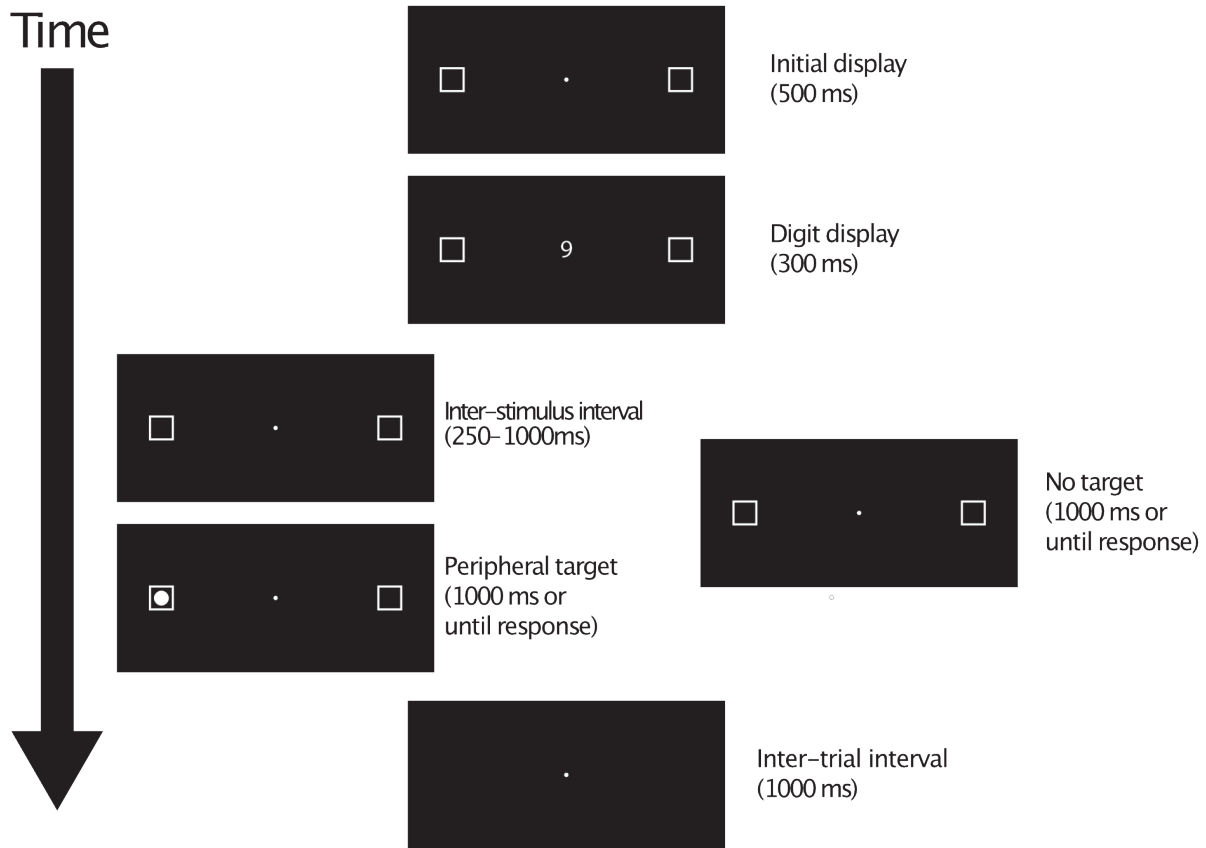


Figure 1. Outline of the trial structure for target trials and catch trials.

301 **Eye-tracking protocol**

302 Code implementing an eye-tracking protocol using an EyeLink 1000 eye-tracker was provided to  
 303 all labs (and is available at <https://github.com/ljcolling/FischerRRR-eyetracking>). For labs using an  
 304 eye-tracker other than an EyeLink 1000, deviations from the standard protocol are listed on the  
 305 lab-specific OSF page. The standard nine-point grid was used for calibration and validation at the start  
 306 of each block or when required during a block. The start of trials was triggered after the detection of 500  
 307 ms of stable fixation within a 2° box centred on the fixation point. If the system could not detect a stable  
 308 fixation within a 2000 ms time window, the calibration process was repeated. After the digit was  
 309 presented, and before the target appeared, the gaze position was monitored and any deviations outside a

310 1° box centred on the fixation point were recorded. Any deviations towards the lateral boxes that  
311 exceeded 2° resulted in the trial being marked as contaminated. These trials were excluded from  
312 primary analyses; however, they were analysed separately to attempt to determine any possible effect of  
313 eye movements on the results.

### 314 **Finger counting**

315 The finger counting assessment was derived from the task developed by Lucidi and Thevenot  
316 (2014). Participants were asked to read aloud four sentences while counting the number of syllables in  
317 each. As reading aloud prevents prevents participants from verbalising counting, most participants  
318 would need to resort to finger counting while sounding out the syllables. For each sentence, the  
319 experimenter recorded the first finger and first hand the participant used. While most participants used  
320 their fingers for the task, some participants did not use their fingers and instead adopted a different  
321 strategy. Participant who failed to engage in finger counting after two sentences were prompted to do so.  
322 Details of the prompting were recorded in lab logs. See OSF for details.

323 The results of the finger counting task were used to place participants into one of five groups:  
324 left-starters, right-starters, left-prefer, right-prefer, and no group. The finger counting group was  
325 determined not only by participants' hand preferences but also by how consistently they engaged in  
326 finger counting. The left- (right-)starter group was defined as those who counted using a hand on all  
327 four occasions and used the left (right) hand on at least three of them. The left (right)-prefer group was  
328 defined as those who counted using a hand on two or three occasions and used the left (right) hand on at  
329 least two of them. The no group group was defined as all other participants (for example, those who did  
330 not count on their fingers, those who only counted on their fingers once, and those who counted an equal  
331 number of times with each hand).

### 332 **Reading/writing direction**

333 Reading and writing direction was determined with a simple three option questionnaire asking if  
334 participants had experience with languages that are written exclusively from left to right (e.g., English

335 and German), not exclusively left to right (e.g., Hebrew), or both types (see <https://osf.io/he5za/> for  
336 details). This was used to cluster participants into two groups: exclusively left-to-right readers/writers  
337 and not exclusively left-to-right readers/writers.

### 338 **Handedness**

339 To assess handedness, we used the 10-item questionnaire from Nicholls, Thomas, Loetscher, and  
340 Grimshaw (2013). In labs conducting the experiment in a language other than English, the questionnaire  
341 was translated and some questions were replaced with more culturally appropriate versions when  
342 required (see <https://osf.io/he5za/> for details).

### 343 **Mathematics assessment**

344 To assess mathematics fluency, we used the short mathematics assessment employed by Tibber  
345 et al. (2013). This test is adapted from the Mathematics Calculation Subtest (WJ-RCalc) of the  
346 Woodcock-Johnson III Tests of Cognitive Abilities (Woodcock & Johnson, 1989). It contains  
347 twenty-five multiple choice mathematics questions requiring addition, subtraction, multiplication, and  
348 division. Participants had thirty seconds to select the response on each trial, with the timing controlled  
349 by the computer software. A countdown timer was stationed in the top left of the screen to inform  
350 participants of the time remaining. The twenty-five questions were split into five levels of five  
351 questions. Two errors on a single level or errors on consecutive levels terminated the test. The final  
352 score was the total number of correct answers.

### 353 **Mathematics anxiety**

354 Mathematics anxiety was assessed using the Abbreviated Math Anxiety Scale (AMAS; Hopko,  
355 Mahadevan, Bare, & Hunt, 2003). The AMAS contains nine questions that ask participants to rate (on a  
356 one to five scale) how anxious they would feel during particular events including thinking of an  
357 upcoming mathematics test, sitting a mathematics examination, and listening to a mathematics lecture.  
358 In labs conducting the experiment in a language other than English, the AMAS was translated (see



359 <https://osf.io/dhnf8/> for details). The final score was the sum of the individual ratings, with scores  
360 ranging from nine (low anxiety) to forty-five (high anxiety).

### 361 **Exit questionnaire**

362 An exit questionnaire that asked participants to describe the purpose of the experiment was used  
363 to determine whether participants could guess the purpose of the experiment. Participants who correctly  
364 guessed the purpose of the experiment, as judged by the experimenter, were excluded from primary  
365 analyses; however, they were analysed separately to determine whether this moderated the effect.

### 366 **Exclusion criteria**

367 Participants whose reaction time data contained more than 5% catch trial errors, who correctly  
368 guessed the purpose of the experiment or who who did not undertake all additional assessments were  
369 excluded from the analysis as per our pre-registration plan (see <https://osf.io/6a2ny/>).

### 370 **Analysis**

371 The dependent variables of interest were the congruency effect at each of the four ISI conditions  
372 (i.e., 250 ms, 500 ms, 750 ms, and 1000 ms). This is defined as the average difference in response time  
373 between congruent and incongruent targets, with congruent targets being defined as left targets preceded  
374 by low digits and right targets preceded by high digits and incongruent targets being defined as left  
375 targets preceded by high digits and right targets preceded by low digits. A positive value for the  
376 congruency effect indicates that participants were faster at responding to congruent targets relative to  
377 incongruent targets, and a negative value indicates the reverse.

378 We analysed our data via multilevel multivariate meta-analytic models (McShane & Böckenholt,  
379 2018). Such models have at least two advantages over the standard random effects meta-analytic model.  
380 First, they better account for the dependence between our multiple dependent variables (i.e., the  
381 congruency effect at each of the four ISI conditions). Second, rather than assuming a simple two-level  
382 structure, with participants nested within labs, they can account for more complex nesting structures

383 such as participants nested within with moderator groups (e.g., left-starters, right-starters, etc.) and  
384 moderator groups nested within within labs. In short, the standard approach necessitates treating several  
385 variance components as zero, thereby making unwarranted independence assumptions.

386 For each analysis, we consider several simplifications to the equal allocation multilevel  
387 multivariate compound symmetry specification detailed in McShane and Böckenholt (2018); we also  
388 consider an equal variance version of the single correlation equal allocation multilevel multivariate  
389 compound symmetry specification that, using the notation of that paper, sets the  $\sigma_{d,d}$  equal for all  
390 dependent variables  $d$  (i.e., the congruency effect at each of the four ISI conditions). We chose among  
391 the six specifications via the Akaike Information Criterion (AIC; Akaike, 1974).

392 In analysing the effect of moderators, it would be ideal to consider them jointly within a single  
393 model. However, this would require a sufficient number of participants in each moderator group.  
394 Specifically, a minimum number of five participants is necessary to compute a  $4 \times 4$  covariance matrix  
395 of full rank (i.e., corresponding to the congruency effect at each of the four ISI conditions) as required.  
396 Therefore, the decision on whether to consider all the moderators jointly within a single model or  
397 separately in different models was left until the sample sizes were known.

398 Unfortunately, data sparsity prevented us from considering all the moderators jointly in a single  
399 model: when considered jointly, many combinations of moderators (e.g., finger counting,  
400 reading/writing direction, handedness) result in either zero or very few participants per moderator group;  
401 indeed, this is also the case for some moderators (i.e., reading/writing direction and handedness) when  
402 considered alone as can be seen in Supplementary Tables A4 and A6 respectively. Consequently, we  
403 consider each moderator separately analysing only moderator groups with a minimum of five  
404 participants. All analyses were pre-registered (see <https://osf.io/6a2ny/>) and carried out in accordance  
405 with this plan.

406 For models featuring no moderators (Model 1) or discrete moderators (finger counting,  
407 reading/writing direction, and handedness; Models 2–4 respectively), we analysed the data at the

408 moderator group level as per McShane and Böckenholt (2018). For the model featuring continuous  
409 moderators (mathematics fluency and mathematics anxiety; Model 5), we analysed the data at the  
410 participant level using an analogous specification (see below for details). Our motivation for  
411 considering these moderators and predictions follow as applicable.

412 **Model 1: No Moderators.** Fischer et al. (2003) suggests a positive congruency effect. The  
413 purpose of Model 1 was to assess this by replicating the analysis performed by Fischer et al. (2003);  
414 consequently, it did not account for any moderators.

415 **Model 2: Finger counting.** Recent work suggests that spatial-numerical compatibility effects  
416 in general (Fischer, 2008)—including attentional cueing effects in response to numbers (Fischer &  
417 Knops, 2014)—might be moderated by finger counting behaviour, specifically being stronger among  
418 those who start finger counting on the left hand and weaker or possibly even reversed among those who  
419 start finger counting on the right hand. The purpose of Model 2 was to assess this and consequently it  
420 took account of the finger counting moderator.

421 This model used only data from participants who consistently engaged in finger counting and  
422 consistently started on the same hand, that is, participants categorised as left-starters or right-starters.  
423 We restricted the analysis to these two groups principally because, if the finger counting moderator is to  
424 have an effect, then we would expect it to be most prominent in those whose finger counting is clear and  
425 unambiguous.

426 **Model 3: Reading/writing direction.** Recent work suggests that the congruency effect might  
427 be weaker or possibly even reversed among those who have experience with languages that are not  
428 read/written exclusively from left to right (Fischer, 2008; Shaki et al., 2009). The purpose of Model 3  
429 was to assess this and consequently it took account of the reading/writing direction moderator.  
430 Specifically, participants were placed into two groups based on the reading/writing questionnaire: those  
431 who read/wrote exclusively left to right and those who did not.

432 **Model 4: Handedness.** The purpose of Model 4 was to assess whether handedness moderates  
433 the congruency effect and consequently it took account of the handedness moderator. Specifically,  
434 participants were classified as left-handed or right-handed according to the handedness questionnaire.

435 **Model 5: Mathematics fluency and mathematics anxiety.** Recent work suggests that  
436 numerical abilities (Fischer, 2006) and mathematics anxiety (Georges et al., 2016) may influence the  
437 strength of spatial-numerical associations. The purpose of Model 5 was to assess this and consequently  
438 it jointly took account of both mathematics fluency and mathematics anxiety as measured by the maths  
439 test and AMAS respectively.

440 Specifically, we fit a multilevel model to the participant-level congruency effects at each of the  
441 four ISI conditions; fixed effects were included for the full set of ISI Condition  $\times$  Maths test  $\times$  AMAS  
442 interactions and random effects were included for (i) each participant, (ii) each Lab  $\times$  ISI Condition  
443 (with equal variance and zero correlation), and (iii) independently each Lab  $\times$  Maths test, Lab  $\times$   
444 AMAS, and Lab  $\times$  Maths test  $\times$  AMAS.

445 **Secondary analyses.** The purpose of our secondary analyses was to assess whether insight into  
446 the purpose of the experiment or eye movements moderate the congruency effect. Specifically, Model 1  
447 was refit separately to data from participants who correctly guessed the purpose of the experiment and to  
448 data from eye movement contaminated trials from participants with contaminated trials at each ISI  $\times$   
449 congruency condition.

450

## Results

### 451 Replication operationalisation

452 The common definition of replication employed in practice is that a subsequent study is  
453 considered to have successfully replicated a prior study if either both failed to attain statistical  
454 significance or both attained statistical significance and were directionally consistent. This definition has  
455 been applied analogously in large-scale replication projects like the present one by comparing the results  
456 of a meta-analysis of the replication studies to the original study in terms of statistical significance.

457 However, the null hypothesis significance testing paradigm upon which this operationalisation of  
458 replication is based has been the subject of no small amount of criticism over the decades (see, for  
459 example, Rozenboom, 1960; Meehl, 1978; Cohen, 1994; Gelman, Carlin, Stern, & Rubin, 2003;

460 McShane & Gal, 2016; McShane & Gal, 2017) and recent calls to abandon it abound (Amrhein,  
461 Trafimow, & Greenland, 2019; McShane, Gal, Gelman, Robert, & Tackett, 2019; Wasserstein, Schirm,  
462 & Lazar, 2019; Amrhein, Greenland, & McShane, 2019). Further, recent work discussing alternative  
463 statistical paradigms specifically in the context of replication (Colling & Szűcs, 2018) has called for a  
464 better understanding of how statistical inference relates to scientific inference. A key point is that any  
465 assessment of whether a theory is supported by data depends on whether the magnitude of the observed  
466 effect is consistent with the theory (Gelman & Carlin, 2014). Consequently, in assessing replication, we  
467 distinguish between *statistical hypotheses* and *scientific hypotheses* and focus on that latter. Specifically,  
468 in discussing our results, we do so in light of the scientific hypothesis advanced by Fischer et al. (2003).

#### 469 **Exclusions**

470 In total, seventeen labs contributed data from a total of 1267 participants; 162 were excluded as  
471 per our pre-registered criteria leaving a total of 1105. See Table 1 for details of the number of  
472 participants collected by each lab, the number analysed, and the number excluded based on each  
473 criterion; the technical error category includes those participants that were excluded for having  
474 incomplete data due to, for example, equipment failure, experimenter error, or other technical errors.

475 Five labs used an eye-tracker for at least some of their participants. See Table A11 for details of  
476 the number of participants tested with an eye-tracker, number of participants analysed in our secondary  
477 analysis of eye movement contaminated trials, and number of eye movement contaminated trials at each  
478 ISI  $\times$  congruency condition for each lab.

#### 479 **Preliminary analyses**

480 Across all 1105 participants and four ISI conditions, the congruency effect we observed had a  
481 mean of 0.24 ms and a standard deviation of 12.48 ms. In addition, across all 1105 participants, it had a  
482 mean of -0.07 ms and a standard deviation of 13.45 ms at the 250 ms ISI condition, a mean of 0.94 ms  
483 and a standard deviation of 12.42 ms at the 500 ms ISI condition, a mean of -0.02 ms and a standard  
484 deviation of 12.12 ms at the 750 ms ISI condition, and a mean of 0.10 ms and a standard deviation of

Table 1

*Total number of participants, number analysed, number excluded for reasons of technical error, number excluded for more than 5% catch trial errors, and number excluded for guessing the purpose of the experiment for each lab.*

Lab	Total Participants	Analysed Participants	Technical Error	Catch Trial Error	Guessed Purpose
Ansari	68	60	2	6	0
Bryce	68	61	0	3	4
Chen	62	60	1	1	0
Cipora	93	82	1	3	7
Colling (Szűcs)	72	65	4	3	0
Corballis	68	64	2	2	0
Hancock	66	54	5	6	1
Holmes	77	60	3	8	6
Lindemann	50	47	0	1	2
Lukavský	62	61	1	0	0
Mammarella	126	103	15	1	7
Mieth	124	93	2	8	21
Moeller	77	63	13	1	0
Ocampo	60	59	0	0	1
Ortiz-Ouellet-Lupiáñez-Santiago	60	54	3	2	1
Toomarian	74	61	4	7	2
Treccani	60	58	0	1	1

485 11.84 ms at the 1000 ms ISI condition. Further, the correlation between conditions had a mean of 0.00  
 486 (and a mean of 0.03 in magnitude) across the six possible pairs of conditions.

487 In terms of sign, across all 1105 participants and four ISI conditions, the proportion of times the  
488 congruency effect we observed was positive was 0.50. In addition, across all 1105 participants, this  
489 proportion was 0.49 at the 250 ms ISI condition, 0.53 at the 500 ms ISI condition, 0.48 at the 750 ms ISI  
490 condition, and 0.50 at the 1000 ms ISI condition. Further, the proportion of times the number of positive  
491 congruency effects per participant was equal to zero, one, two, three, and four was respectively 0.06,  
492 0.26, 0.36, 0.26, and 0.06. All of these results are compatible with the relevant binomial distribution  
493 with probability parameter one-half (i.e., the distribution of the number of heads on tosses of a fair coin).

#### 494 **Primary analyses**

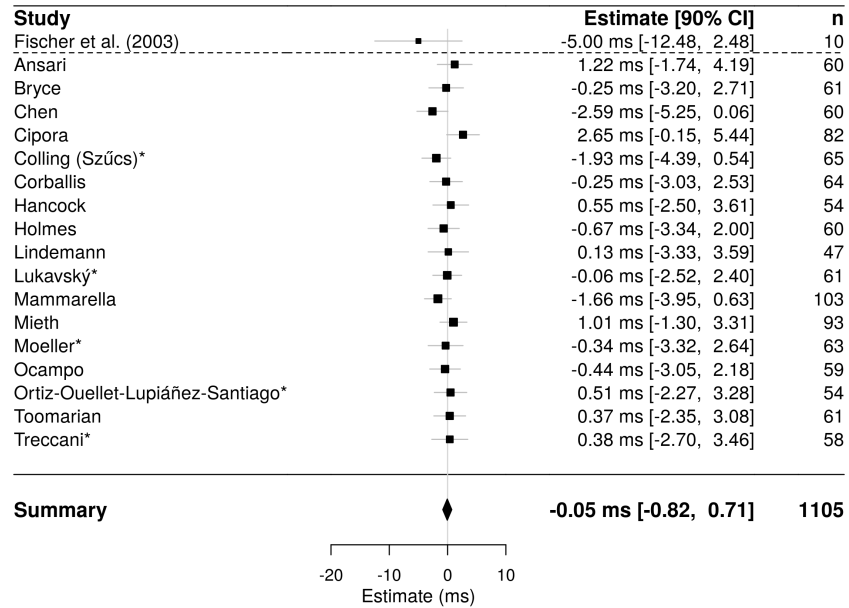
495 **Model 1: No moderators.** The purpose of Model 1 was to replicate the analysis performed by  
496 Fischer et al. (2003), and thus it did not account for any moderators. Model 1 was fit to data from 1105  
497 participants from seventeen labs. We summarise the results from Study 2 of Fischer et al. (2003) along  
498 with results from each lab and from Model 1 in Figure 2.

499 The effects we observed both within and across labs were minuscule and incompatible with those  
500 observed in Fischer et al. (2003). Specifically, Fischer et al. (2003) estimated an effect of -5.00 ms at the  
501 250 ms ISI condition, 18.00 ms at the 500 ms ISI condition, 23.00 ms at the 750 ms ISI condition, and  
502 11.00 ms at the 1000 ms ISI condition. In contrast, Model 1 estimates an effect of -0.05 ms, 1.06 ms,  
503 0.19 ms, and 0.18 ms at each of the four respective ISI conditions.

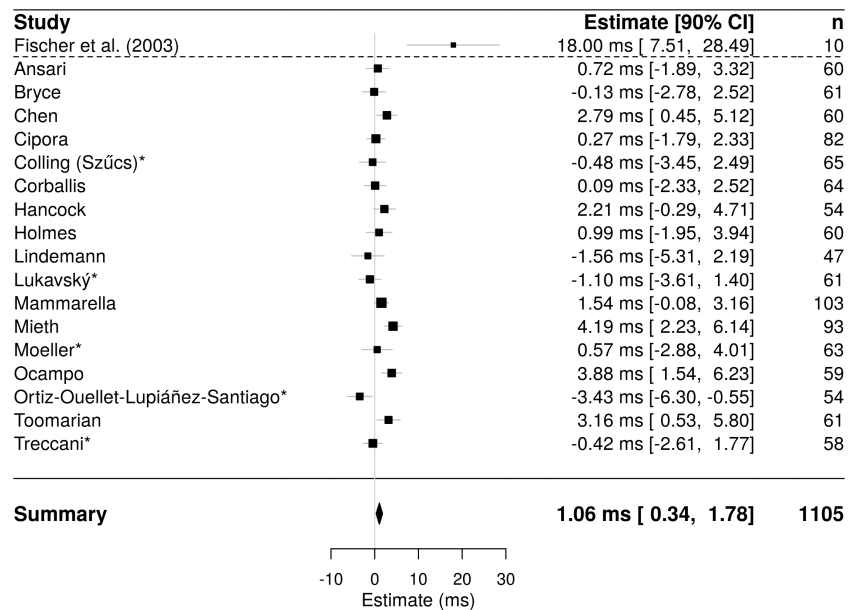
504 Given this in tandem with the results of our preliminary analyses, we conclude that we have *failed*  
505 to replicate the effect reported by Fischer et al. (2003).

506 Another major finding was that the effects we observed were highly consistent not only across ISI  
507 conditions but also—perhaps more surprisingly—across labs. Recent work has found that, contrary to  
508 both substantive and statistical expectations, large-scale replications projects like the present one tend to  
509 show a nontrivial degree of heterogeneity across labs (McShane, Tackett, Böckenholt, & Gelman, 2019).  
510 In contrast, we estimate heterogeneity across labs at 1.02 ms and thus practically unimportant for most  
511 purposes. This suggests that, at least across the labs involved in the present project, there are unlikely to

512 be lab-level moderators driving our results. See Table 1 and Supplementary Table A1 for additional  
 513 details.

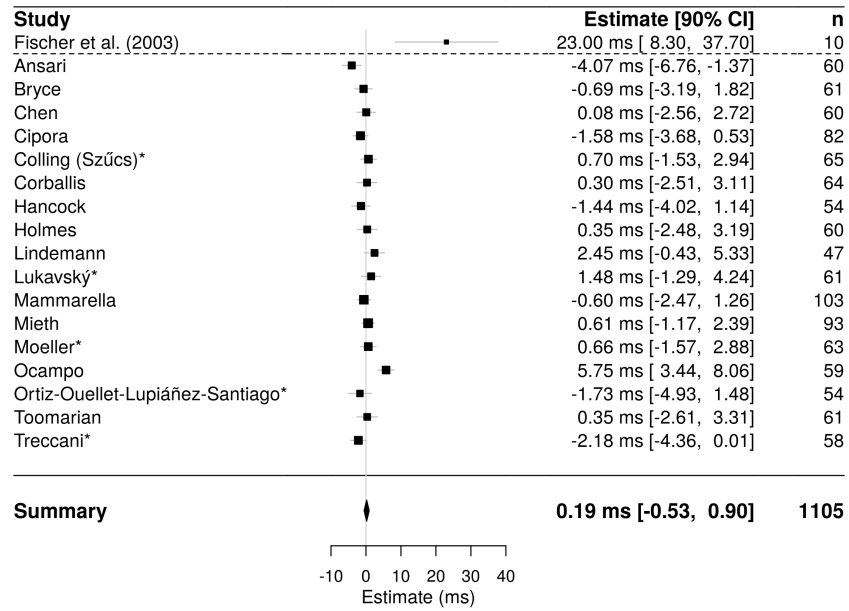


(a) 250 ms ISI Condition

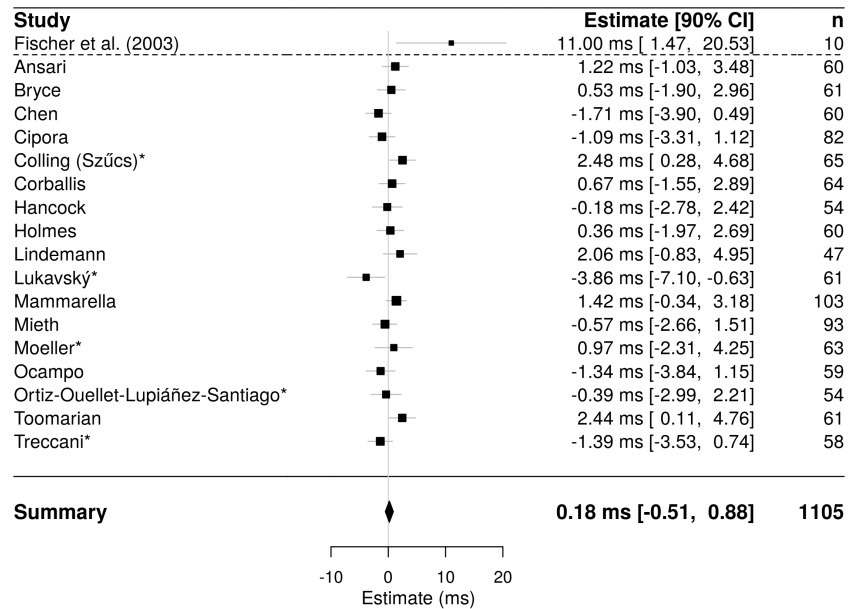


(b) 500 ms ISI Condition





(c) 750 ms ISI Condition



(d) 1000 ms ISI Condition

Figure 2. Summary of Results from Study 2 of Fischer et al. (2003), Each Lab, and Model 1. The effects we observed both within and across labs were miniscule—around 1 ms—and incompatible with those of around 20 ms observed in Fischer et al. (2003). They were also highly consistent not only across ISI conditions but also—perhaps more surprisingly—across labs with the latter suggesting there are unlikely to be lab-level moderators driving our results Labs using an eye-tracker are marked with an asterisk.

514 **Model 2: Finger counting.** Model 2 was fit to data from 343 left-starter participants from  
515 seventeen labs and 482 right-starter participants from seventeen labs. We summarize the results from  
516 Model 2 along with the results from Study 2 of Fischer et al. (2003) as well as Model 1, Model 3, and  
517 Model 4 in Figure 3. While the evidence presented above suggests a stronger congruency effect among  
518 left-starters and a weaker or possibly even reversed effect among right-starters, as can be seen in Figure  
519 3, finger counting had no substantial impact on the results: we observed minuscule effects for each ISI  
520 condition and finger counting group and minuscule differences between the two finger counting groups  
521 at each ISI condition. See Supplementary Table A2 and Supplementary Table A3 for additional details.

522 **Model 3: Reading/writing direction.** Model 3 was fit to data from 1014 exclusively  
523 left-to-right readers/writers from seventeen labs and 76 not exclusively left-to-right readers/writers from  
524 eight labs. While the evidence presented above suggests a weaker or possibly even reversed congruency  
525 effect among those who have experience with languages that are not read/written exclusively from left  
526 to right, as can be seen in Figure 3, reading/writing direction had no substantial impact on the results:  
527 we observed minuscule effects for each ISI condition and reading/writing direction group and minuscule  
528 differences between the two reading/writing direction groups at each ISI condition. See Supplementary  
529 Table A4 and Supplementary Table A5 for additional details.

530 **Model 4: Handedness.** Model 4 was fit to data from 69 left-handed participants from nine labs  
531 and 1007 right-handed participants seventeen labs. As can be seen in Figure 3, handedness had no  
532 substantial impact on the results: we observed minuscule effects for each ISI condition and handedness  
533 group and minuscule differences between the two handedness groups at each ISI condition. See  
534 Supplementary Table A6 and Supplementary Table A7 for additional details.

535 **Model 5: Mathematics fluency and mathematics anxiety.** Model 5 was fit to data from 1105  
536 participants from seventeen labs. While the evidence presented above suggests mathematics fluency and  
537 mathematics anxiety might moderate congruency effects, we observed no substantial moderating effects.  
538 See Table 1 and Supplementary Table A8 for additional details.

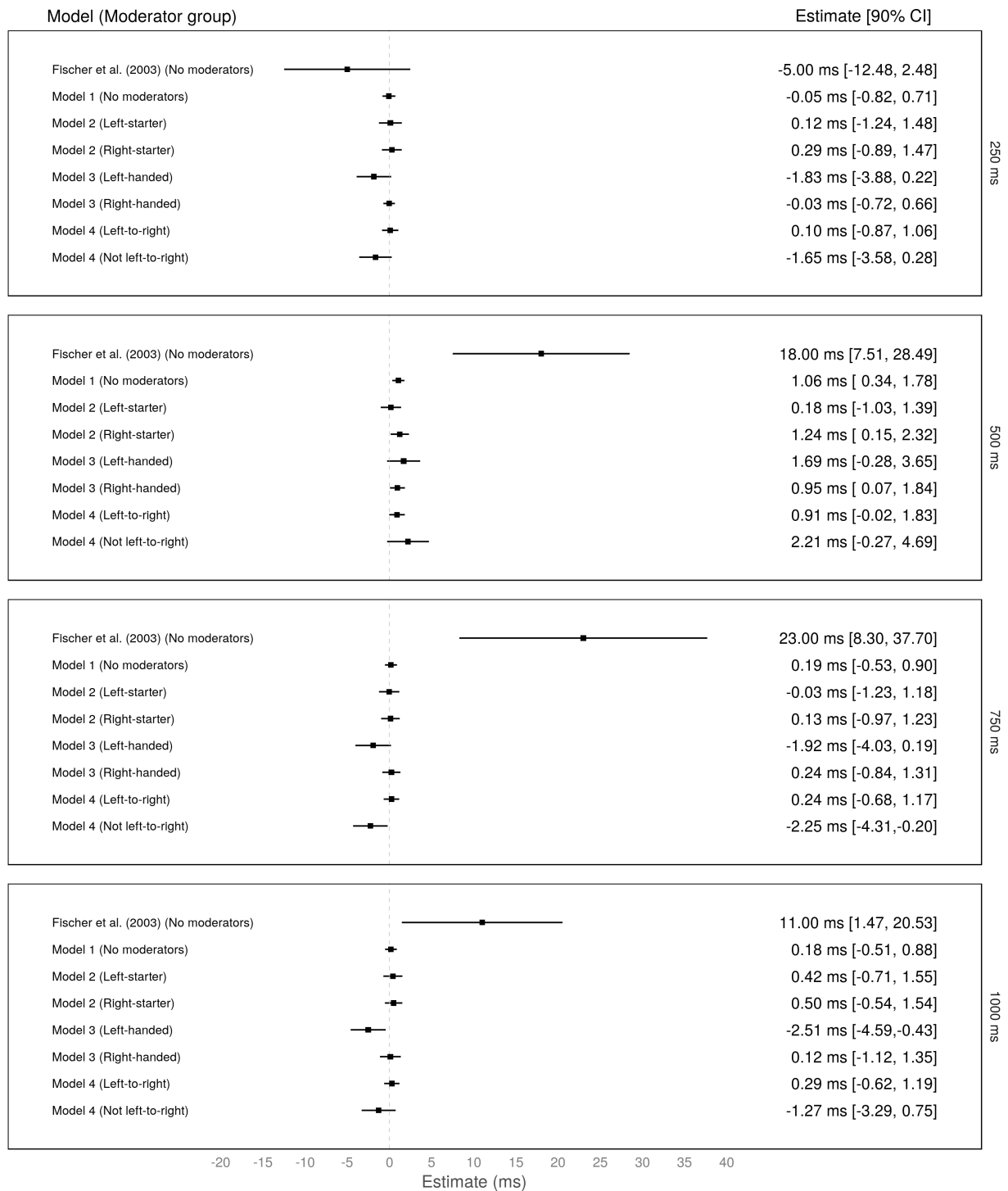


Figure 3. Summary of Results from Study 2 of Fischer et al. (2003) and Models 1–4. The effects we observed were minuscule and incompatible with those observed in Fischer et al. (2003). They were also highly consistent across ISI conditions.

### 539 **Secondary analyses**

540 Model 1 was refit separately to data from 41 participants from four labs who correctly guessed the  
541 purpose of the experiment and to data from 10468 eye movement contaminated trials from 132  
542 participants from five labs with contaminated trials at each ISI  $\times$  congruency condition. These analyses  
543 yielded nothing of substantive interest. See Supplementary Materials for details.

### 544 **Discussion**

545 The att-SNARC effect (Fischer et al., 2003) has been used to argue for an early,  
546 response-independent, and automatic origin of the SNARC effect. If the SNARC effect is produced by  
547 early mechanisms, this would provide good evidence for “embodied” number representations and allow  
548 for strong claims about the link between number and space (e.g., a mental number line).

549 We attempted to replicate Study 2 of Fischer et al. (2003) by collecting data from 1105  
550 participants across seventeen labs. Across all 1105 participants and four ISI conditions, the proportion  
551 of times the congruency effect we observed was positive was 0.50. Further, the effects we observed both  
552 within and across labs were miniscule and incompatible with those observed in Fischer et al. (2003).  
553 Given this, we conclude that we have *failed* to replicate the effect reported by Fischer et al. (2003).

554 The effects we observed were also highly consistent not only across ISI conditions but  
555 also—perhaps more surprisingly—across labs. The latter suggests there are unlikely to be lab-level  
556 moderators driving our results. In addition, our analysis of several participant-level moderators (finger  
557 counting preferences, reading/writing direction experience, handedness, and mathematics fluency and  
558 mathematics anxiety) revealed no substantial moderating effects.

559 We conclude with two important points. First, one might, on the basis of the common definition  
560 of replication employed in practice, object that we have successfully replicated Fischer et al. (2003), at  
561 least at the 500 ms ISI condition. In response, we argue this illustrates one major flaw of that definition:  
562 our result at the 500 ms ISI condition is manifestly incompatible with the analogous result of Fischer  
563 et al. (2003). In addition, we view a difference of about 1 ms, even if “real”, as too small for any



589 of the original authors, in particular Martin Fischer and Jay Pratt. We also note this project would not  
590 have been possible without editor Alex Holcombe's patient and thoughtful help at every step of the  
591 process.

592

### **Author contributions**

593 LJC and DS proposed the study. LJC programmed the experiments. LJC and BBM conducted the  
594 analyses. LJC wrote an initial manuscript. LJC and BBM wrote revised and final manuscripts. All  
595 authors critically reviewed the final manuscript.

## References

596

597 Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic*  
598 *Control*, 19(6), 716–723. doi:10.1109/TAC.1974.1100705

599 Allaire, J., Ushey, K., & Tang, Y. (2018). *Reticulate: Interface to 'python'*. R package version 1.10.  
600 Retrieved from <https://CRAN.R-project.org/package=reticulate>

601 Amrhein, V., Greenland, S., & McShane, B. B. (2019). Retire statistical significance. *Nature*, 7748(567),  
602 305–307. doi:10.1038/d41586-019-00857-9

603 Amrhein, V., Trafimow, D., & Greenland, S. (2019). Inferential statistics as descriptive statistics: There  
604 is no replication crisis if we don't expect replication. *The American Statistician*, 73(sup1),  
605 262–270. doi:10.1080/00031305.2018.1543137

606 Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. R package version  
607 0.1.0.9842. Retrieved from <https://github.com/crsh/papaja>

608 Bache, S. M., & Wickham, H. (2014). *Magrittr: A forward-pipe operator for r*. R package version 1.5.  
609 Retrieved from <https://CRAN.R-project.org/package=magrittr>

610 Bächtold, D., Baumüller, M., & Brugger, P. (1998). Stimulus-response compatibility in representational  
611 space. *Neuropsychologia*, 36(8), 731–735. doi:10.1016/S0028-3932(98)00002-5

612 Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4.  
613 *Journal of Statistical Software*, 67(1), 1–48. doi:10.18637/jss.v067.i01

614 Bates, D., & Maechler, M. (2018). *Matrix: Sparse and dense matrix classes and methods*. R package  
615 version 1.2-14. Retrieved from <https://CRAN.R-project.org/package=Matrix>

616 Bengtsson, H. (2018). *R.matlab: Read and write mat files and call matlab from within r*. R package  
617 version 3.6.2. Retrieved from <https://CRAN.R-project.org/package=R.matlab>

618 Champely, S. (2018). *Pwr: Basic functions for power analysis*. R package version 1.2-2. Retrieved from  
619 <https://CRAN.R-project.org/package=pwr>

620 Cohen, J. (1994). The earth is round ( $p < .05$ ). *American Psychologist*, 49, 997–1003.  
621 doi:10.1037/0003-066X.49.12.997

- 622 Colling, L. J., & Szűcs, D. (2018). Statistical inference and the replication crisis. *Review of Philosophy*  
623 *and Psychology*, 1–27. doi:10.1007/s13164-018-0421-4
- 624 Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number  
625 magnitude. *Journal of Experimental Psychology: General*, 122(3), 371–396.  
626 doi:10.1037//0096-3445.122.3.371
- 627 Dodd, M. D., Van der Stigchel, S., Leghari, M. A., Fung, G., & Kingstone, A. (2008). Attentional  
628 SNARC: There’s something special about numbers (let us count the ways). *Cognition*, 108(3),  
629 810–818. doi:10.1016/j.cognition.2008.04.006
- 630 Fattorini, E., Pinto, M., Rotondaro, F., & Doricchi, F. (2015). Perceiving numbers does not cause  
631 automatic shifts of spatial attention. *Cortex*, 73, 298–316. doi:10.1016/j.cortex.2015.09.007
- 632 Fias, W., Brysbaert, M., Geypens, F., & d’Ydewalle, G. (1996). The importance of magnitude  
633 information in numerical processing: Evidence from the SNARC effect. *Mathematical Cognition*,  
634 2(1), 95–110. doi:10.1080/135467996387552
- 635 Fischer, M. H. (2006). The Future for Snarc Could Be Stark... *Cortex*, 42(8), 1066–1068.  
636 doi:10.1016/S0010-9452(08)70218-1
- 637 Fischer, M. H. (2008). Finger counting habits modulate spatial-numerical associations. *Cortex*, 44(4),  
638 386–392. doi:10.1016/j.cortex.2007.08.004
- 639 Fischer, M. H., & Brugger, P. (2011). When digits help digits: Spatial-numerical associations point to  
640 finger counting as prime example of embodied cognition. *Frontiers in Psychology*, 2, 260.  
641 doi:10.3389/fpsyg.2011.00260
- 642 Fischer, M. H., Castel, A. D., Dodd, M. D., & Pratt, J. (2003). Perceiving numbers causes spatial shifts  
643 of attention. *Nature Neuroscience*, 6(6), 555–556. doi:10.1038/nn1066
- 644 Fischer, M. H., & Knops, A. (2014). Attentional cueing in numerical cognition. *Frontiers in Psychology*,  
645 5(325), 426. doi:10.3389/fpsyg.2014.01381
- 646 Fodor, J. A. (1975). *The language of thought*. Cambridge, MA: Harvard Unvieristy Press.
- 647 Galfano, G., Rusconi, E., & Umiltà, C. (2006). Number magnitude orients attention, but not against  
648 one’s will. *Psychonomic Bulletin & Review*, 13(5), 869–874. doi:10.3758/BF03194011



- 649 Galton, F. (1880). Visualised numerals. *Nature*, *21*, 252–256. doi:10.1038/021252a0
- 650 Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2003). *Bayesian data analysis*. Chapman and Hall/CRC:  
651 Boca Raton, FL.
- 652 Gelman, A., & Carlin, J. (2014). Beyond power calculations: Assessing Type S (sign) and Type M  
653 (magnitude) errors. *Perspectives in Psychological Science*, *9*(6), 641–651.  
654 doi:10.1177/1745691614551642
- 655 Georges, C., Hoffmann, D., & Schiltz, C. (2016). How Math Anxiety Relates to Number–Space  
656 Associations. *Frontiers in Psychology*, *7*(33), 143. doi:10.3389/fpsyg.2016.01401
- 657 Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially  
658 organized. *Cognition*, *87*(3), B87–B95. doi:10.1016/S0010-0277(02)00234-2
- 659 Gevers, W., Reynvoet, B., & Fias, W. (2004). The Mental Representation of Ordinal Sequences is  
660 Spatially Organised: Evidence from Days of the Week. *Cortex*, *40*(1), 171–172.  
661 doi:10.1016/S0010-9452(08)70938-9
- 662 Gevers, W., Verguts, T., Reynvoet, B., Vaessens, B., & Fias, W. (2006). Numbers and space: A  
663 computational model of the SNARC effect. *Journal of Experimental Psychology: Human  
664 Perception and Performance*, *32*(1), 32–44. doi:10.1037/0096-1523.32.1.32
- 665 Gładziejewski, P., & Miłkowski, M. (2017). Structural representations: Causally relevant and different  
666 from detectors. *Biology & philosophy*, *32*(3), 337–355. doi:10.1007/s10539-017-9562-6
- 667 Gordon, M., & Lumley, T. (2017). *Forestplot: Advanced forest plot using 'grid' graphics*. R package  
668 version 1.7.2. Retrieved from <https://CRAN.R-project.org/package=forestplot>
- 669 Henry, L., & Wickham, H. (2018). *Purrr: Functional programming tools*. R package version 0.2.5.  
670 Retrieved from <https://CRAN.R-project.org/package=purrr>
- 671 Hester, J. (2018). *Glue: Interpreted string literals*. R package version 1.3.0. Retrieved from  
672 <https://CRAN.R-project.org/package=glue>
- 673 Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The abbreviated math anxiety scale  
674 (AMAS): Construction, validity, and reliability. *Assessment*, *10*(2), 178–182.  
675 doi:10.1177/1073191103252351

- 676 Ioannidis, J. P. A. (2008). Why most discovered true associations are inflated. *Epidemiology*, *19*(5),  
677 640–648. doi:10.1097/EDE.0b013e31818131e7
- 678 Keus, I. M., Jenks, K. M., & Schwarz, W. (2005). Psychophysiological evidence that the SNARC effect  
679 has its functional locus in a response selection stage. *Cognitive Brain Research*, *24*(1), 48–56.  
680 doi:10.1016/j.cogbrainres.2004.12.005
- 681 Keus, I. M., & Schwarz, W. (2005). Searching for the functional locus of the SNARC effect: Evidence  
682 for a response-related origin. *Memory & Cognition*, *33*(4), 681–695. doi:10.3758/BF03195335
- 683 Landy, D. H., Jones, E. I., & Hummel, J. E. (2008). Why spatial-numerical associations aren't evidence  
684 for a mental number line. In *Proceedings of the 30th annual conference of the cognitive science*  
685 *society* (pp. 257–362). Austin, TX.
- 686 Lang, M. (2017). checkmate: Fast argument checks for defensive r programming. *The R Journal*, *9*(1),  
687 437–445. doi:10.32614/RJ-2017-028
- 688 Lucidi, A., & Thevenot, C. (2014). Do not count on me to imagine how I act: behavior contradicts  
689 questionnaire responses in the assessment of finger counting habits. *Behavior Research Methods*,  
690 *46*(4), 1079–1087. doi:10.3758/s13428-014-0447-1
- 691 McShane, B. B., & Böckenholt, U. (2018). Multilevel multivariate meta-analysis with application to  
692 choice overload. *Psychometrika*, *83*(1), 255–271. doi:10.1007/s11336-017-9571-z
- 693 McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in  
694 meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on*  
695 *Psychological Science*, *11*(5), 730–749. doi:10.1177/1745691616662243
- 696 McShane, B. B., & Gal, D. (2016). Blinding us to the obvious? The effect of statistical training on the  
697 evaluation of evidence. *Management Science*, *62*(6), 1707–1718. doi:10.1287/mnsc.2015.2212
- 698 McShane, B. B., & Gal, D. (2017). Statistical significance and the dichotomization of evidence. *Journal*  
699 *of the American Statistical Association*, *112*(519), 885–895. doi:10.1080/01621459.2017.1289846
- 700 McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical  
701 significance. *American Statistician*, *73*(sup1), 235–245. doi:10.1080/00031305.2018.1527253

- 702 McShane, B. B., Tackett, J. L., Böckenholt, U., & Gelman, A. (2019). Large scale replication projects in  
703 contemporary psychological research. *The American Statistician*, 73(sup1), 99–105.  
704 doi:10.1080/00031305.2018.1505655
- 705 Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress  
706 of soft psychology. *Journal of Consulting and Clinical Psychology*, 46(4), 806–834.  
707 doi:10.1037/0022-006X.46.4.806
- 708 Müller, K. (2018). *Bindrcpp: An 'rcpp' interface to active bindings*. R package version 0.2.2. Retrieved  
709 from <https://CRAN.R-project.org/package=bindrcpp>
- 710 Müller, K., & Wickham, H. (2018). *Tibble: Simple data frames*. R package version 1.4.2. Retrieved  
711 from <https://CRAN.R-project.org/package=tibble>
- 712 Newell, A., & Simon, H. (1976). Computer science as empirical inquiry: Symbols and search.  
713 *Communications of the ACM*, 19(3), 113–126. doi:10.1145/360018.360022
- 714 Nicholls, M. E. R., Thomas, N. A., Loetscher, T., & Grimshaw, G. M. (2013). The Flinders Handedness  
715 survey (FLANDERS): A brief measure of skilled hand preference. *Cortex*, 49(10), 2914–2926.  
716 doi:10.1016/j.cortex.2013.02.002
- 717 Ooms, J. (2018). *Magick: Advanced graphics and image-processing in r*. R package version 1.9.  
718 Retrieved from <https://CRAN.R-project.org/package=magick>
- 719 Pecher, D., & Boot, I. (2011). Numbers in space: Differences between concrete and abstract situations.  
720 *Frontiers in Psychology*, 2(121). doi:10.3389/fpsyg.2011.00121
- 721 Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2018). *nlme: Linear and nonlinear*  
722 *mixed effects models*. R package version 3.1-137. Retrieved from  
723 <https://CRAN.R-project.org/package=nlme>
- 724 Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32(1), 3–25.  
725 doi:10.1080/00335558008248231
- 726 Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for performance of  
727 speeded binary classification tasks. *Psychological Bulletin*, 132(3), 416–442.  
728 doi:10.1037/0033-2909.132.3.416

- 729 R Core Team. (2018). *R: A language and environment for statistical computing*. R Foundation for  
730 Statistical Computing. Vienna, Austria. Retrieved from <https://www.R-project.org/>
- 731 Ranzini, M., Dehaene, S., Piazza, M., & Hubbard, E. M. (2009). Neural mechanisms of attentional shifts  
732 due to irrelevant spatial and numerical cues. *Neuropsychologia*, *47*(12), 2615–2624.  
733 doi:10.1016/j.neuropsychologia.2009.05.011
- 734 Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology*, *83*(2,  
735 Pt.1), 274–278. doi:10.1037/h0028573
- 736 Ristic, J., Wright, A., & Kingstone, A. (2006). The number line effect reflects top-down control.  
737 *Psychonomic Bulletin & Review*, *13*(5), 862–868. doi:10.3758/BF03194010
- 738 Rozenboom, W. W. (1960). The fallacy of the null hypothesis significance test. *Psychological Bulletin*,  
739 *57*(5), 416–428. doi:10.1037/h0042040
- 740 Salillas, E., El Yagoubi, R., & Semenza, C. (2008). Sensory and cognitive processes of shifts of spatial  
741 attention induced by numbers: An erp study. *Cortex*, *44*(4), 406–413.  
742 doi:10.1016/j.cortex.2007.08.006
- 743 Santens, S., & Gevers, W. (2008). The SNARC effect does not imply a mental number line. *Cognition*,  
744 *108*(1), 263–270. doi:10.1016/j.cognition.2008.01.002
- 745 Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers  
746 contribute to the SNARC effect. *Psychonomic Bulletin & Review*, *16*(2), 328–331.  
747 doi:10.3758/PBR.16.2.328
- 748 Tibber, M. S., Manasseh, G. S. L., Clarke, R. C., Gagin, G., Swanbeck, S. N., Butterworth, B., ...  
749 Dakin, S. C. (2013). Sensitivity to numerosity is not a unique visuospatial psychophysical  
750 predictor of mathematical ability. *Vision Research*, *89*, 1–9. doi:10.1016/j.visres.2013.06.006
- 751 Toomarian, E. Y., & Hubbard, E. M. (2018). On the genesis of spatial-numerical associations:  
752 Evolutionary factors co-construct the mental number line. *Neuroscience and Biobehavioural*  
753 *Reviews*, *90*, 184–199. doi:10.1016/j.neubiorev.2018.04.010

- 754 van Dijck, J.-P., Abrahamse, E. L., Acar, F., Ketels, B., & Fias, W. (2014). A working memory account  
755 of the interaction between number and spatial attention. *Quarterly Journal of Experimental*  
756 *Psychology*, 67(8), 1500–1513. doi:10.1080/17470218.2014.903984
- 757 van Dijck, J.-P., & Fias, W. (2011). A working memory account for spatial–numerical associations.  
758 *Cognition*, 119(1), 114–119. doi:10.1016/j.cognition.2010.12.013
- 759 Wasserstein, R., Schirm, A., & Lazar, N. (2019). Moving to a world beyond  $p < 0.05$ . *The American*  
760 *Statistician*, 73(sup1), 1–19. doi:10.1080/00031305.2019.1583913
- 761 Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York.  
762 Retrieved from <http://ggplot2.org>
- 763 Wickham, H. (2017). *Tidyverse: Easily install and load the 'tidyverse'*. R package version 1.2.1.  
764 Retrieved from <https://CRAN.R-project.org/package=tidyverse>
- 765 Wickham, H. (2018a). *Forcats: Tools for working with categorical variables (factors)*. R package  
766 version 0.3.0. Retrieved from <https://CRAN.R-project.org/package=forcats>
- 767 Wickham, H. (2018b). *Stringr: Simple, consistent wrappers for common string operations*. R package  
768 version 1.3.1. Retrieved from <https://CRAN.R-project.org/package=stringr>
- 769 Wickham, H., François, R., Henry, L., & Müller, K. (2018). *Dplyr: A grammar of data manipulation*. R  
770 package version 0.7.6. Retrieved from <https://CRAN.R-project.org/package=dplyr>
- 771 Wickham, H., & Henry, L. (2018). *Tidyr: Easily tidy data with 'spread()' and 'gather()' functions*. R  
772 package version 0.8.1. Retrieved from <https://CRAN.R-project.org/package=tidyr>
- 773 Wickham, H., Hester, J., & François, R. (2017). *Readr: Read rectangular text data*. R package version  
774 1.1.1. Retrieved from <https://CRAN.R-project.org/package=readr>
- 775 Williams, D., & Colling, L. J. (2018). From symbols to icons: The return of resemblance in the cognitive  
776 neuroscience revolution. *Synthese*, 195(5), 1941–1967. doi:10.1007/s11229-017-1578-6
- 777 Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, 9(4), 625–636.  
778 doi:10.3758/BF03196322

- 779 Wood, G., Nuerk, H.-C., & Willmes, K. (2006). Crossed Hands and the Snarc Effect: A failure to  
780 Replicate Dehaene, Bossini and Giraux (1993). *Cortex*, *8*, 1069–1079.  
781 doi:10.1016/S0010-9452(08)70219-3
- 782 Wood, G., Willmes, K., Nuerk, H.-C., & Fischer, M. H. (2008). On the cognitive link between space and  
783 number: A meta-analysis of the SNARC effect. *Psychology Science*, *50*(4), 489–525.
- 784 Woodcock, R., & Johnson, M. (1989). *Woodcock Johnson—Revised-tests of academic achievement*.  
785 Chicago: The Riverside Publishing Company.
- 786 Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd). Boca Raton, Florida: Chapman and  
787 Hall/CRC.
- 788 Zanolie, K., & Pecher, D. (2014). Number-induced shifts in spatial attention: a replication study.  
789 *Frontiers in Psychology*, *5*(e85048), 667. doi:10.3389/fpsyg.2014.00987
- 790 Zebian, S. (2005). Linkages between number concepts, spatial thinking, and directionality of writing:  
791 The SNARC Effect and the REVERSE SNARC effect in English and Arabic monoliterates,  
792 biliterates, and illiterate Arabic speakers. *Journal of Cognition and Culture*, *5*(1–2), 165–190.  
793 doi:10.1163/1568537054068660
- 794 Zhu, H. (2018). *Kableextra: Construct complex table with 'kable' and pipe syntax*. R package version  
795 0.9.0. Retrieved from <https://CRAN.R-project.org/package=kableExtra>

## Supplementary Results

796 **Primary analyses**

797 **Model 1: No Moderators.** Model 1 was fit to data from 1105 participants from seventeen labs  
798 (see Table 1 for details). Of the six equal allocation multilevel multivariate compound symmetry  
799 (EAMMCS) model specifications, the *Equal Variance, Zero Correlation* specification was chosen by  
800 AIC. AIC; fixed effect estimates, standard errors, and  $z$ -statistics; and variance component estimates are  
801 shown in Supplementary Table A1.

802 **Model 2: Finger counting.** Model 2 was fit to data from 343 left-starter participants from  
803 seventeen labs and 482 right-starter participants from seventeen labs (see Supplementary Table A2 for  
804 details). Of the six EAMMCS model specifications, the *Equal Variance, Zero Correlation* specification  
805 was chosen by AIC. AIC; fixed effect estimates, standard errors, and  $z$ -statistics; and variance  
806 component estimates are shown in Supplementary Table A3.

807 **Model 3: Reading/writing direction.** Model 3 was fit to data from 1014 exclusively  
808 left-to-right readers/writers from seventeen labs and 76 not exclusively left-to-right readers/writers from  
809 eight labs (see Supplementary Table A4 for details). Of the six EAMMCS model specifications, the  
810 *Equal Variance, Zero Correlation* specification was chosen by AIC. AIC; fixed effect estimates,  
811 standard errors, and  $z$ -statistics; and variance component estimates are shown in Supplementary Table  
812 A5.

813 **Model 4: Handedness.** Model 4 was fit to data from 69 left-handed participants from nine labs  
814 and 1007 right-handed participants from seventeen labs (see Supplementary Table A6 for details). Of  
815 the six EAMMCS model specifications, the *Unequal Variance, Zero Correlation* specification was  
816 chosen by AIC. AIC; fixed effect estimates, standard errors, and  $z$ -statistics; and variance component  
817 estimates are shown in Supplementary Table A7.

818 **Model 5: Mathematics fluency and mathematics anxiety.** Model 5 was fit to data from 1105  
819 participants from seventeen labs (see Table 1). See the main text for model specification details, but

820 note that (i) for consistency with Model 1 we employed the *Equal Variance, Zero Correlation*  
821 specification for the Lab  $\times$  ISI Condition effects and (ii) the maths test and AMAS were centred and  
822 scaled by their respective means and standard deviations across the 1105 participants prior to estimation  
823 of the model. Fixed effect estimates, standard errors, and *t*-statistics and variance component estimates  
824 are shown in Supplementary Table A8.

## 825 **Secondary analyses**

826 **Purpose of experiment.** Data from several participants were not included in the primary  
827 analysis because they correctly guessed the purpose of the experiment (as assessed by the exit  
828 questionnaire). The data from these participants was analysed separately to determine whether insight  
829 into the purpose of the experiment moderated the effect. Specifically, Model 1 was refit to data from the  
830 41 participants from four labs who correctly guessed the purpose of the experiment (see Supplementary  
831 Table A9 for details). Of the six model EAMMCS model specifications, the *Equal Variance, Zero*  
832 *Correlation* specification was chosen by AIC. AIC; fixed effect estimates, standard errors, and  
833 *z*-statistics; and variance component estimates are shown in Supplementary Table A10.

834 **Eye-movement contaminated trials.** Data from individual trials that were contaminated with  
835 eye movements were also not included the primary analysis. The data from these trials was analysed  
836 separately to determine whether eye movements moderated the effect. Specifically, Model 1 was refit to  
837 data from 10468 eye movement contaminated trials from 132 participants from five labs with  
838 contaminated trials at each ISI  $\times$  congruency condition (see Supplementary Table A11 for details). Of  
839 the six EAMMCS model specifications, the *Fixed Effects* specification was chosen by AIC. AIC; fixed  
840 effect estimates, standard errors, and *z*-statistics; and variance component estimates are shown in  
841 Supplementary Table A12



Table A1

*Model 1 Estimates.*

(a) *AIC*

Specification	AIC
Fixed Effects	264.12
Equal Variance, Zero Correlation	259.66
Equal Variance, Single Correlation	261.64
Unequal Variance, Zero Correlation	261.04
Unequal Variance, Single Correlation	260.87
No Constraints	270.83

(b) *Fixed Effect Estimates*

ISI Condition	Estimate	Std. Err.	<i>z</i>
250 ms	-0.05	0.47	-0.11
500 ms	1.06	0.44	2.43
750 ms	0.19	0.43	0.43
1000 ms	0.18	0.42	0.44

(c) *Variance Component Estimates. Estimates are presented on the standard deviation scale.*

ISI Condition	Estimate
250 ms	1.02
500 ms	1.02
750 ms	1.02
1000 ms	1.02

Table A2

*Number of participants in each finger counting group for each of the seventeen labs.*

Lab	Left- Starter	Left- Prefer	No Group	Right- Prefer	Right- Starter
Ansari	23	2	2	3	30
Bryce	13	8	2	17	21
Chen	22	0	2	0	36
Cipora	19	9	5	18	31
Colling (Szűcs)	21	3	11	3	27
Corballis	18	3	5	4	34
Hancock	22	6	0	3	23
Holmes	14	2	1	8	35
Lindemann	22	1	4	1	19
Lukavský	12	7	2	16	24
Mammarella	30	8	6	23	36
Mieth	32	10	10	16	25
Moeller	23	0	6	0	34
Ocampo	27	0	2	0	30
Ortiz-Ouellet-Lupiáñez-Santiago	10	8	4	22	10
Toomarian	19	0	0	0	42
Treccani	16	7	4	6	25

Table A3

*Model 2 Estimates.*

(a) *AIC*

Specification	AIC
Fixed Effects	665.97
Equal Variance, Zero Correlation	637.31
Equal Variance, Single Correlation	639.00
Unequal Variance, Zero Correlation	638.57
Unequal Variance, Single Correlation	640.13
No Constraints	646.51

(b) *Fixed Effect Estimates*

ISI Condition	Finger counting group	Estimate	Std. Err.	<i>z</i>
250 ms	Right-starter	0.29	0.72	0.40
250 ms	Left-starter	0.12	0.83	0.14
500 ms	Right-starter	1.24	0.66	1.88
500 ms	Left-starter	0.18	0.74	0.24
750 ms	Right-starter	0.13	0.67	0.19
750 ms	Left-starter	-0.03	0.73	-0.04
1000 ms	Right-starter	0.50	0.63	0.79
1000 ms	Left-starter	0.42	0.69	0.61

(c) *Variance Component Estimates. Estimates are presented on the standard deviation scale. 39% of the variance is estimated to be at the lab-level and 61% at the group-level.*

ISI Condition	Estimate
250 ms	1.74
500 ms	1.74
750 ms	1.74
1000 ms	1.74

Table A4

*Number of participants in each of the reading/writing direction groups for each of the seventeen labs.*

Lab	Exclusively	Not exclusively
	Left-to-Right	Left-to-Right
Ansari	55	5
Bryce	59	2
Chen	39	21
Cipora	76	6
Colling (Szűcs)	55	10
Corballis	60	4
Hancock	53	1
Holmes	54	6
Lindemann	47	0
Lukavský	58	3
Mammarella	103	0
Mieth	79	14
Moeller	54	9
Ocampo	55	4
Ortiz-Ouellet-Lupiáñez-Santiago	54	0
Toomarian	56	5
Treccani	57	1

Table A5

*Model 3 Estimates.*

(a) *AIC*

Specification	AIC
Fixed Effects	495.58
Equal Variance, Zero Correlation	448.05
Equal Variance, Single Correlation	449.41
Unequal Variance, Zero Correlation	451.89
Unequal Variance, Single Correlation	453.44
No Constraints	457.83

(b) *Fixed Effect Estimates*

ISI Condition	Reading/Writing Direction	Estimate	Std. Err.	<i>z</i>
250 ms	Exclusively LTR	0.10	0.59	0.17
250 ms	Not exclusively LTR	-1.65	1.17	-1.41
500 ms	Exclusively LTR	0.91	0.56	1.62
500 ms	Not exclusively LTR	2.21	1.51	1.46
750 ms	Exclusively LTR	0.24	0.56	0.43
750 ms	Not exclusively LTR	-2.25	1.25	-1.80
1000 ms	Exclusively LTR	0.29	0.55	0.53
1000 ms	Not exclusively LTR	-1.27	1.23	-1.03

(c) *Variance Component Estimates. Estimates are presented on the standard deviation scale. 10% of the variance is estimated to be at the lab-level and 90% at the group-level.*

ISI Condition	Estimate
250 ms	1.71
500 ms	1.71
750 ms	1.71
1000 ms	1.71

Table A6

*Number of participants in each handedness group for each of the seventeen labs.*

Lab	Left- handed	Right- handed
Ansari	4	56
Bryce	4	57
Chen	5	55
Cipora	3	79
Colling (Szűcs)	7	58
Corballis	9	55
Hancock	6	48
Holmes	4	56
Lindemann	5	42
Lukavský	7	54
Mammarella	6	97
Mieth	14	79
Moeller	4	59
Ocampo	4	55
Ortiz-Ouellet-Lupiáñez-Santiago	3	51
Toomarian	10	51
Treccani	3	55

Table A7

*Model 4 Estimates.*

(a) *AIC*

Specification	AIC
Fixed Effects	598.41
Equal Variance, Zero Correlation	473.56
Equal Variance, Single Correlation	475.56
Unequal Variance, Zero Correlation	470.86
Unequal Variance, Single Correlation	472.48
No Constraints	480.12

(b) *Fixed Effect Estimates*

ISI Condition	Handedness Group	Estimate	Std. Err.	<i>z</i>
250 ms	Right-handed	-0.03	0.42	-0.07
250 ms	Left-handed	-1.83	1.25	-1.46
500 ms	Right-handed	0.95	0.54	1.76
500 ms	Left-handed	1.69	1.19	1.42
750 ms	Right-handed	0.24	0.65	0.37
750 ms	Left-handed	-1.92	1.28	-1.50
1000 ms	Right-handed	0.12	0.75	0.16
1000 ms	Left-handed	-2.51	1.27	-1.98

(c) *Variance Component Estimates. Estimates are presented on the standard deviation scale. 12% of the variance is estimated to be at the lab-level and 88% at the group-level.*

ISI Condition	Estimate
250 ms	0.01
500 ms	1.57
750 ms	2.19
1000 ms	2.71

Table A8

*Model 5 Estimates.*

(a) *Fixed Effect Estimates*

Effect	Estimate	Std. Err.	<i>t</i>
250 ms ISI	-0.03	0.44	-0.07
500 ms ISI	0.88	0.44	2.02
750 ms ISI	0.01	0.44	0.02
1000 ms ISI	0.21	0.44	0.48
250 ms ISI × Maths test	-0.15	0.42	-0.35
500 ms ISI × Maths test	-0.80	0.42	-1.90
750 ms ISI × Maths test	-0.24	0.42	-0.57
1000 ms ISI × Maths test	0.08	0.42	0.18
250 ms ISI × AMAS	-0.66	0.40	-1.66
500 ms ISI × AMAS	0.29	0.40	0.73
750 ms ISI × AMAS	-0.21	0.40	-0.54
1000 ms ISI × AMAS	-0.57	0.40	-1.44
250 ms ISI × Maths test × AMAS	-0.12	0.39	-0.30
500 ms ISI × Maths test × AMAS	-0.38	0.39	-0.98
750 ms ISI × Maths test × AMAS	-0.24	0.39	-0.63
1000 ms ISI × Maths test × AMAS	0.22	0.39	0.56

(b) *Variance Component Estimates. Estimates are presented on the standard deviation scale.*

ISI Condition	Estimate	Additional Effects	Estimate
250 ms	0.85	Participant	0.00
500 ms	0.85	Maths Test	0.61
750 ms	0.85	AMAS	0.33
1000 ms	0.85	Maths test × AMAS	0.50



Table A9

*Number of participants who correctly guessed the purpose of the experiment for each lab.*

Lab	<i>n</i>
Cipora	7
Holmes	6
Mammarella	7
Mieth	21

Table A10

*Model 1 Estimates (only participants who correctly guessed the purpose of the experiment).*

(a) *AIC*

Specification	AIC
Fixed Effects	80.21
Equal Variance, Zero Correlation	71.39
Equal Variance, Single Correlation	73.39
Unequal Variance, Zero Correlation	73.83
Unequal Variance, Single Correlation	75.83
No Constraints	85.42

(b) *Fixed Effect Estimates*

ISI Condition	Estimate	Std. Err.	<i>z</i>
250 ms	1.49	2.21	0.67
500 ms	0.36	2.32	0.16
750 ms	-0.68	2.17	-0.31
1000 ms	1.15	2.37	0.48

(c) *Variance Component Estimates. Estimates are presented on the standard deviation scale.*

ISI Condition	Estimate
250 ms	3.08
500 ms	3.08
750 ms	3.08
1000 ms	3.08

Table A11

*Number of participants tested with an eye-tracker, number of participants analysed in our secondary analysis of eye movement contaminated trials, and number of eye movement contaminated trials in the analysis (total number of eye movement contaminated trials) at each ISI × congruency condition for each lab.*

Lab	Participants	Analysed	Trial Type	250 ms	500 ms	750 ms	1000 ms
Colling (Szűcs)	52	18	Congruent	64 (88)	93 (133)	109 (173)	107 (162)
			Incongruent	71 (97)	95 (144)	103 (140)	95 (142)
Lukavský	61	29	Congruent	158 (182)	201 (240)	235 (278)	252 (292)
			Incongruent	146 (176)	202 (238)	231 (280)	233 (282)
Moeller	64	53	Congruent	593 (600)	723 (734)	774 (787)	851 (868)
			Incongruent	621 (635)	711 (729)	774 (802)	842 (858)
Ortiz-Ouellet-Lupiáñez-Santiago	28	18	Congruent	127 (135)	165 (177)	176 (186)	184 (197)
			Incongruent	130 (138)	147 (157)	167 (174)	160 (175)
Treccani	30	14	Congruent	89 (99)	113 (136)	129 (139)	133 (152)
			Incongruent	99 (109)	116 (126)	124 (144)	125 (141)

Table A12

*Model 1 Estimates (only eye movement contaminated trials).*

(a) *AIC*

Specification	AIC
Fixed Effects	120.28
Equal Variance, Zero Correlation	122.28
Equal Variance, Single Correlation	124.28
Unequal Variance, Zero Correlation	127.98
Unequal Variance, Single Correlation	129.75
No Constraints	139.65

(b) *Fixed Effect Estimates*

ISI Condition	Estimate	Std. Err.	<i>z</i>
250 ms	-5.35	6.27	-0.85
500 ms	-2.65	4.95	-0.54
750 ms	-5.52	3.98	-1.39
1000 ms	3.86	4.17	0.93

(c) *Variance Component Estimates. Estimates are presented on the standard deviation scale.*

ISI Condition	Estimate
250 ms	0
500 ms	0
750 ms	0
1000 ms	0