



HAL
open science

The influence of semantic processing and response latency on the SNARC effect

Daniele Didino, Christina Breil, André Knops

► **To cite this version:**

Daniele Didino, Christina Breil, André Knops. The influence of semantic processing and response latency on the SNARC effect. *Acta Psychologica*, 2019, 196, pp.75-86. 10.1016/j.actpsy.2019.04.008 . hal-02454395

HAL Id: hal-02454395

<https://hal.science/hal-02454395>

Submitted on 24 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

1 Title: **The influence of semantic processing and response latency on the SNARC effect**

2

3

4

5 Daniele Didino^{a*}, Christina Breil^a, André Knops^{b,c}

6

7

8

9 a. Department of Psychology, Faculty of Life Sciences, Humboldt-Universität zu Berlin,
10 Berlin, Germany.

11 b. CNRS UMR 8240, Laboratory for the Psychology of Child Development and Education,
12 Paris, France.

13 c. University Paris Descartes, Sorbonne Paris Cité, Paris, France.

14

15

16

17 * Corresponding author:

18 Daniele Didino

19 Department of Psychology, Humboldt-Universität zu Berlin, Rudower Chaussee 18, 12489
20 Berlin, Germany. E-mail: daniele.didino@gmail.com

21

22

23 Competing interests: The authors declare that they have no conflict of interest.

24

25

26

27

28

29

30

1. INTRODUCTION

In their seminal paper, Dehaene, Bossini and Giraux (1993) asked participants to classify the parity of numbers by pressing a left- or right-sided key (i.e., parity judgement task). They discovered that relative small numbers elicit faster left-sided responses and large numbers faster right-sided responses. Since then, this Spatial-Numerical Association of Response Codes effect (SNARC; for reviews see Fisher & Shaki, 2014; van Dijck et al., 2015; Wood et al., 2008) has been replicated multiple times.

Originally, the SNARC effect has been explained within the framework of the *mental number line (MNL) account* (Dehaene et al., 1993). The MNL is a semantic long-term memory representation in which numbers are coded on a unidimensional conceptual space, horizontally organized in ascending order from left to right: small numbers are encoded on the left side and large numbers on the right side (Hubbard et al., 2005). This spatial organization is thought to emerge from cultural practices such as writing/reading direction (Göbel, 2015; Göbel, Maier, & Shaki, 2015; Göbel, Shaki, & Fischer, 2011; for a right-to-left spatial orientation of the MNL in non-Western cultures see Shaki, Fischer, & Göbel, 2012; Shaki, Fischer, & Petrusic, 2009). According to the MNL account, the SNARC effect emerges from an isomorphism (i.e., a direct mapping) between the position of a number on this semantic representation and the left-right coordinates of the external response locations. Therefore, a congruent mapping between conceptual (e.g., small numbers) and external space (e.g., left-sided responses) generates faster response latencies compared to an incongruent mapping (e.g., small numbers and right-sided responses), and vice versa for large numbers.

In the majority of the tasks used to investigate the SNARC effect (e.g., parity judgement task) the magnitude of the number is irrelevant for the decision process (for a list of tasks see Wood et al., 2008). Therefore, the MNL account proposes that the SNARC effect emerges from the interference produced by an automatic activation of the semantic representation of numbers (Dehaene et al., 1993; Nuerk, Wood, & Willmes, 2005). An important question pertains to whether and how the amount of semantic processing (i.e., processing of abstract properties of numbers such as magnitude or parity) required by the task modulates the level of interference generated by the MNL. In line with the MNL account, one would expect that deeper semantic processing of a number leads to a stronger activation of the MNL and in turn more interference (i.e., a stronger SNARC effect) is expected. Wood and colleagues (2008) evaluated this hypothesis in a meta-analysis which was consistent with the MNL account. Tasks requiring deeper semantic processing (e.g., parity judgement and magnitude classification) were associated with a stronger SNARC effect compared to tasks asking to discriminate non-semantic features of the numbers (e.g., color judgement) (see table 1 and

1 figure 2 in Wood et al., 2008). However, to the best of our knowledge, no study tested this
2 hypothesis in a within-participant design.

3 The assumption of a direct mapping between the position of a number on the MNL and
4 response location coordinates has been challenged by the dual route model (Gevers et al.,
5 2006, 2010; Santens & Gevers, 2008) and the working memory (WM) account (Abrahamse,
6 van Dijck, & Fias, 2016; Fias & van Dijck, 2016; van Dijck & Fias, 2011; van Dijck, Gevers, &
7 Fias, 2009; van Dijck et al., 2012, 2014). Neither of these accounts predicts that the strength
8 of the SNARC effect is affected by the level of semantic processing required by a task.

9 The *dual route model* (Gevers et al., 2006, 2010; Santens & Gevers, 2008; see also the
10 polarity correspondence account for a similar idea not restricted to number processing, Proctor
11 and Cho, 2006) introduces an intermediate level of processing, between the MNL and the
12 response stage, that codes numbers into binary categories such as magnitude (i.e., small vs.
13 large), parity (i.e., odd vs. even) or other dichotomous features of the stimulus. As the name
14 suggests, this model includes two routes of parallel information processing. The unconditional
15 route codes numbers based on their magnitude and is automatically activated regardless of
16 the task requirements. This route activates the long-term preexisting links between magnitude
17 and space coordinates, that is the associations small–left and large–right. The conditional
18 route codes numbers into a binary category based on task-specific requirements. It activates
19 short-term links between numbers and an arbitrary mapping based on task requirements, such
20 as small–right and large–left (for magnitude) or odd–right and even–left (for parity). The
21 SNARC effect emerges from the congruence, or lack of it, between the response sides
22 activated by the two routes. On the one hand, the conditional route activates a response side
23 based on task-specific mapping. On the other hand, the unconditional route can cooperate or
24 compete to activate the same or the opposite response side, respectively. If the task-related
25 mapping is consistent with the long-term number-space links the two routes cooperate,
26 otherwise they converge on opposite spatial response codes. The response selection process
27 takes longer and response latency is slower when the routes diverge and this generates the
28 SNARC effect. Since both routes are always simultaneously activated independent of the task,
29 the level of semantic processing required by the task should not influence the SNARC effect.
30 According to the dual-route model, however, the strength of the SNARC effect is influenced by
31 the duration of the number processing (Gevers et al., 2006). Namely, the strength of the
32 SNARC effect would increase along with response latency, because longer latency provides
33 the unconditional route with more time to interfere with response selection. Finally, the model
34 also predicts that in magnitude comparison or classification tasks the SNARC effect takes a
35 step-like shape (Gevers et al., 2006), indicating a categorical decision. In tasks that do not
36 require to evaluate the magnitude of a number (e.g., parity judgement task), the SNARC effect
37 is stronger for the extremes of the numerical interval presented in the experiment and thus it

1 takes a linear shape (for a review see Wood et al., 2008; but see Schroeder, Nuerk & Plewnia,
2 2017c). On the other hand, in a magnitude classification task, where the numbers have to be
3 compared to a fixed standard (e.g., the numbers 1 to 9 being classified as smaller or larger
4 than 5), the SNARC effect interacts with the distance effect (i.e., the time required to compared
5 two numbers is inversely proportional to their numerical distance; Moyer & Landauer, 1967;
6 Verguts, Fias, & Stevens, 2005). Therefore, the numbers closer to the standard require more
7 time to be classified and thus the unconditional route has more time to interfere with the
8 response selection. This increases the strength of the SNARC effect for numbers in the center
9 of the interval and thus the effect acquires a categorical rather than linear shape (see figures
10 2 and 3 in Gevers et al., 2006).

11 Both the MNL account and the dual-route model assume that the association between
12 numbers and space coordinates are an intrinsic property of the long-term memory
13 representation of number. The *working memory (WM) account* instead assumes that no spatial
14 information is co-represented together with numbers and that the long-term representation only
15 includes serial order information (Abrahamse, van Dijck, & Fias, 2016; Fias & van Dijck, 2016;
16 van Dijck & Fias, 2011; van Dijck, Gevers, & Fias, 2009; van Dijck et al., 2012, 2014). The
17 number-space association would emerge from a temporary binding of numbers to a spatially
18 oriented template in working memory. Therefore, the WM account proposes that the SNARC
19 effect is produced by this temporary left-to-right orientation that numbers (considered as verbal
20 items) take only after being transferred to a spatial working memory template. Evidence in
21 favor of the coexistence of long-term space-number associations (i.e., spatially oriented MNL)
22 and temporary associations between space and order position in working memory (Ginsburg
23 & Gevers, 2015; Huber et al., 2016; Schroeder, Nuerk & Plewnia, 2017c) can be reconciled
24 with the WM account working assuming that multiple serially ordered sequences of numbers
25 can be simultaneously activated in working memory (Abrahamse, van Dijck, & Fias, 2016).
26 Reverse SNARC-like effect for ordinal non-numerical sequences (e.g., weekdays) under
27 transcranial direct current stimulation (Schroeder, Nuerk & Plewnia, 2017a) also challenges
28 the WM account, but mixed evidence emerged for such effect (Schroeder, Nuerk & Plewnia,
29 2017b). Since the spatial orientation is not a feature of the long-term memory representation
30 but rather emerges from short-term association in working memory, the level of processing
31 required by the task should not affect the strength of the SNARC effect. Therefore, the WM
32 account also predicts no influence of the amount of semantic processing on the strength of the
33 SNARC effect.

34 If one strictly defines “semantic processing” as reflecting the access to and the manipulation
35 of numerical magnitude, the fact that the SNARC effect was originally discovered in a parity
36 judgement task (Dehaene et al., 1993), which does not require magnitude processing, already
37 suggests that the amount of semantic processing required by the task does not influence the

1 strength of this effect. Wood and colleagues' review (2008), which showed that the size of the
2 SNARC effect in magnitude classification and parity judgement tasks is comparable, also
3 provides evidence for a lack of influence of semantic processing. However, although unrelated
4 with numerical magnitude, parity is actually an intrinsic property of both non-symbolic
5 numerosities (reflecting the possibility to divide a set of objects in two subsets with equal
6 cardinality) and symbolic numbers (reflecting the property of being divisible by 2 with no
7 remainder, in the natural numbers set). Therefore, a parity judgement task requires explicitly
8 to process an abstract semantic (unrelated to magnitude) property of the number. Since in
9 both parity judgement and magnitude classification tasks participants are asked to evaluate an
10 intrinsic property of the number, it is plausible to assume that these two tasks do not allow to
11 clearly disentangle the depth of the semantic processing and thus to measure properly its
12 influence on the SNARC effect. Therefore, to investigate the effect of semantic processing on
13 the SNARC effect, it is more convenient to compare tasks requiring deeper semantic
14 processing (e.g., parity judgement and magnitude classification) with tasks in which response
15 selection is not related to semantic features of the numbers (e.g., color judgement and
16 phoneme detection, see description below). In fact, although to some extent the mere
17 presentation of a number may suffice to automatically activate its semantic representation
18 (Dehaene et al., 1998; Koechlin et al., 1999; Naccache & Dehaene, 2001), the automatic
19 processing of semantic information is not compulsory (Kunde, Kiesel & Hoffmann, 2003;
20 Maxfield, 1997). Therefore, we assumed that tasks requiring to assess non-numerical
21 properties of the stimulus (e.g., color judgement and phoneme detection) will evoke a weaker
22 activation of the semantic representation compared to tasks that directly require the
23 discrimination of magnitude or parity.

24 The present study aimed to evaluate whether and how the amount semantic processing
25 required by a task affects the strength of the SNARC effect, as predicted by the MNL approach.
26 Participants performed two tasks requiring semantic number processing (magnitude
27 classification and parity judgement) and two tasks requiring the discrimination of non-semantic
28 features of the numbers (color judgement and phoneme detection). According to the MNL
29 account, a stronger SNARC effect is expected for the magnitude classification task and parity
30 judgement task compared to the other two tasks. According to the dual-route model, the
31 SNARC effect should be modulated by response latency rather than by the amount of semantic
32 processing. Moreover, this model predicts a categorical shape of the SNARC effect for the
33 magnitude classification task and a linear shape for the other tasks. Finally, similar to the dual-
34 route model, the WM account does not predict that tasks requiring deeper semantic processing
35 (magnitude classification and parity judgement) generate a stronger SNARC effect compared
36 to those based on more peripheral features of the numbers (color judgement or phoneme
37 detection).

1 Further predictions can also be made on the relationship between the SNARC effects
2 evoked in the different tasks¹ [DD1]. The MNL account assumes the same origin for the SNARC
3 effect regardless of the task. In fact, independently of the task requirements, a common
4 semantic representation is automatically (or deliberately) activated and interferes with the
5 response selection. Given the common origin of this interference, in addition to the expected
6 relationship between SNARC effect and amount of semantic processing, the MNL account also
7 predicts strong correlations between the strength of this effect in the different tasks. The dual-
8 route model, on the other hand, assumes that the SNARC effect originates from the
9 competition/cooperation between two routes of parallel information processing. In other words,
10 in different tasks the unconditional route competes or cooperates with distinct conditional
11 routes, which reflect short-term task-specific mappings. Therefore, the dual-route model does
12 not expressly entail a correlation between the SNARC effects of the different tasks. The WM
13 account assumes that the SNARC effect emerges from the temporary spatial orientation that
14 numbers acquire after being loaded in a working memory template. The binding of numbers to
15 this spatial template is both context- and task-specific. Therefore, similar to the dual-route
16 model, the WM account also does not expressly entail a correlation between the SNARC
17 effects evoked by the different tasks.

18

19

20 **2. METHOD**

21

22 **2.1. Participants**

23 Thirty-two German-speaking participants took part in the study (22 female; mean age (SD)
24 = 26.7 (4.4), range = 18–35). All participants had normal or corrected-to-normal vision and
25 gave informed consent to participate for course credits or 8€. All participants were right-handed
26 and had no experience with right-to-left writing/reading languages. The study was approved
27 by the Ethics committee at the Department of Psychology of Humboldt-Universität zu Berlin
28 (Nr. 2017–29).

29

30 **2.2. Materials, tasks and design**

31 The same set of numbers (Arabic digits ranging from 1 to 9, excluding 5) was used in all
32 tasks. Each participant performed four tasks. In the *magnitude classification task*, participants
33 were asked to decide if the presented number was smaller or larger than the fixed standard 5.
34 In the *parity judgement task*, were asked to judge if the number was even or odd. In the *color*
35 *judgement task*, participants classified the color of the font: numbers 2, 3, 6 and 8 were always
36 printed in red (RGB: 255, 0, 0) and numbers 1, 4, 7 and 9 always in green (RGB: 0, 255, 0). In

¹ We would like to thank Wim Gevers for pointing out this idea.

1 the *phoneme detection task*, participants decided whether or not any of the phonemes /s/, /z/
 2 and /ts/ was included in the German number word corresponding to the presented Arabic digit
 3 (for a similar task see Fias, 2001; Fias et al., 1996). This rule was selected to create an arbitrary
 4 mapping (i.e., not related to semantic features) between numbers and response sides.
 5 Therefore, numbers were divided into two groups based on the presence (1 = /a:ns/, 2 = /tsva:/,
 6 6 = /zεks/, 7 = /zi:bn/) or the absence (3 = /dra:/, 4 = /fi:ɐ/, 8 = /axt/, 9 = /no:n/) of these
 7 phonemes. Participants were explicitly instructed to base their decision on the phonological
 8 feature and ignore the orthography of the word.

9 Four task sequences were created using the following rules: (1) each task must be
 10 presented in each position of the sequence only once; (2) each task can follow the other tasks
 11 only once. The four sequences were: Magnitude–Color–Parity–Phonology; Parity–
 12 Magnitude–Phonology–Color; Phonology–Parity–Color–Magnitude; Color–Phonology–
 13 Magnitude–Parity. Each sequence was administered to 8 participants. Each task included two
 14 blocks, in which the response mapping was reversed (**Table 1**). Fifteen participants started all
 15 tasks with the response mapping A and 17 participants with the response mapping B (the
 16 sample size in the two groups is not equal because of a technical problem). In total, participants
 17 performed 8 blocks (4 tasks × 2 response mappings). Following the recommendations of
 18 Cipora and Wood (2017), in each block each number was repeated 20 times. Numbers were
 19 pseudo-randomly presented with the constraint that the same digit could not be presented on
 20 two consecutive trials. Sixteen practice trials preceded each block (each number presented
 21 twice in a randomized order) to familiarize with the block-specific response mapping and
 22 feedback was provided (no feedback was presented during the test blocks).

23

24 **Table 1.** The two response mappings used in the tasks. Response mappings A and B
 25 were presented in two different blocks in counterbalanced order across participants.

	Response mapping	Left key	Right key
Magnitude task	A	1 2 3 4	6 7 8 9
	B	6 7 8 9	1 2 3 4
Parity task	A	1 3 7 9	2 4 6 8
	B	2 4 6 8	1 3 7 9
Phonological task	A	1 2 6 7	3 4 8 9
	B	3 4 8 9	1 2 6 7
Color task	A	2 3 6 8	1 4 7 9
	B	1 4 7 9	2 3 6 8

26

27 **2.3. Procedure**

28 The same procedure was used in all tasks. Stimulus presentation and response collection
 29 were implemented in Matlab, using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997).
 30 Stimuli were presented in the center of the monitor and were 22 mm high and 15 mm wide.
 31 Participants sat at approximately 50 cm from the monitor (visual angle: 2.5° × 1.7°).

1 Each trial started with a fixation mark (#) presented for 600 ms, followed by an Arabic
2 number, which remained on the screen until the response or for 1300 ms. Participants were
3 instructed to press a key on the left (“left-control” with the left-hand) or right (“enter” on the
4 numpad with the right-hand) side of the keyboard according to the block-specific response
5 mapping (**Table 1**). The two keys were approximately 40 cm apart. Following the offset of the
6 presented number, the next trial began after an intertrial-interval of 500 ms consisting of a
7 black screen. All stimuli were printed in white against a black background, except for numbers
8 in the color judgement which could be red (RGB: 255,0,0; numbers 2, 3, 6 and 8) or green
9 (RGB: 0, 255, 0; 1, 4, 7, 9).

10 Since participants had to perform eight blocks with different response mappings, a small
11 sheet showing the response mapping was placed under the monitor to help them to remember
12 the block-specific instructions. Participants were asked to respond as fast and accurately as
13 possible.

16 3. RESULTS

17
18 All analyses^[DD2] were performed using R-project software (R Core Team, 2015) and
19 RStudio software (RStudio Team, 2015). Participants had a very high accuracy in all tasks
20 (magnitude classification: mean = 0.97, SD = 0.03, range = 0.82–1.0; parity judgement: mean
21 = 0.95, SD = 0.03, range = 0.87–0.99; color discrimination: mean = 0.97, SD = 0.03, range =
22 0.88–0.997; phoneme detection: mean = 0.94, SD = 0.04, range = 0.84–0.99). Accuracy data
23 are likely affected by a ceiling effect and thus will not be further analyzed.

24 Trials with incorrect responses and with reaction times (RTs) shorter than 200 ms (N = 9)
25 were excluded from the analyses. For each participant and task combination, trials with RT
26 more than 2.5 SD from the participant’s mean were considered outliers and excluded from the
27 analysis (2.9% of the trials). A preliminary analysis on RTs showed that the order of the tasks
28 was not significant and did not interact with other variables. Therefore, order was excluded
29 from the following analyses. All ANOVAs were Greenhouse-Geisser corrected (Greenhouse &
30 Geisser, 1959) when the assumption of sphericity was violated; uncorrected degrees of
31 freedom and epsilon values (ϵ_{GG}) are reported. Effect sizes are reported following the
32 recommendation of Lakens (2013).

33 We also calculated the dRT as mean RTs for the right hand minus mean RTs for the left
34 hand, separately for each target number, task, and participant (see Fias et al., 1996; see also
35 Pinhas, Tzelgov, & Ganor-Stern, 2012; Tzelgov, Zohar-Shai, & Nuerk, 2013). The dRTs were
36 analyzed by means of a regression model: $dRT \sim \beta_0 + \beta_1 \times \text{Number}$ (Fias et al., 1996). The

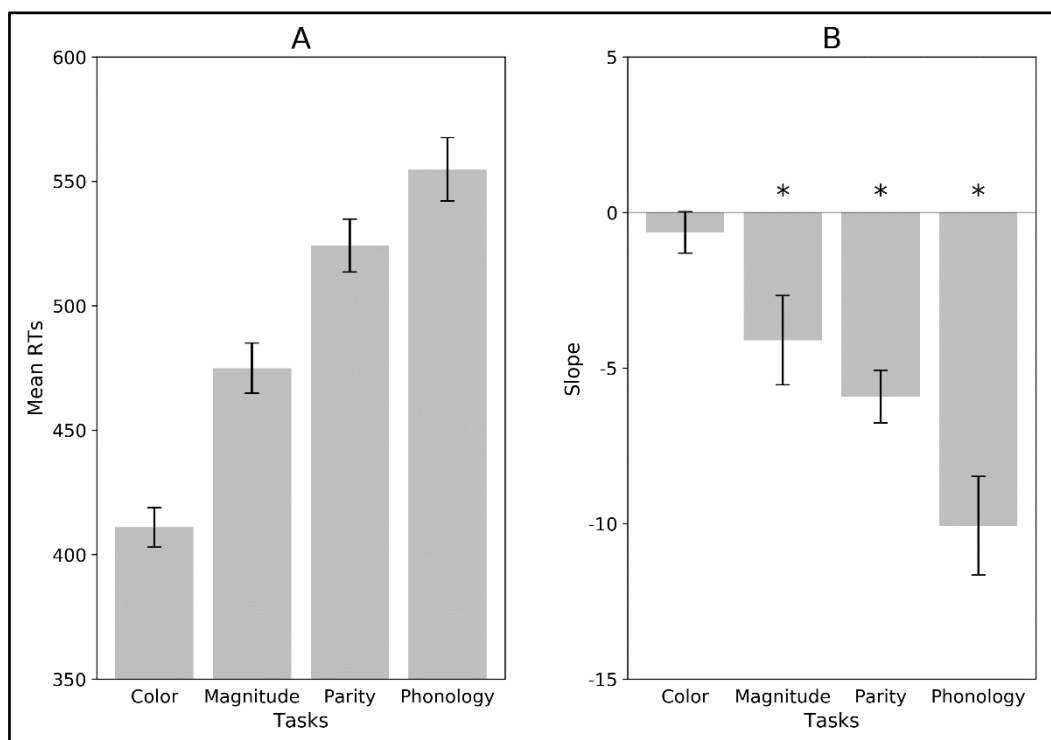
1 slope (β_1) was interpreted as a measure of the SNARC effect: larger negative values
2 correspond to a stronger effect.

3

4 3.1. Reaction times

5 To investigate the SNARC effect, we first analyzed mean RTs with a repeated measure
6 ANOVA with the within-subject factors task (magnitude, parity, phonology, color), target
7 number (1, 2, 3, 4, 6, 7, 8, 9) and response hand (right vs. left). The main effect of number was
8 significant [$F(7, 217) = 21.47, p < 0.001, \epsilon_{GG} = 0.70, \text{generalized } \eta^2 = 0.02$]. Mean RTs (in
9 ms) were 478 (SD = 52), 477 (57), 501 (53), 501 (56), 501 (58), 493 (52), 485 (53), 493 (49),
10 for the numbers 1, 2, 3, 4, 6, 7, 8, 9, respectively. The main effect of task was significant [$F(3,$
11 $93) = 127.67, p < 0.001, \epsilon_{GG} = 0.87, \text{generalized } \eta^2 = 0.4$]. Mean RTs were 411 (SD = 45),
12 475 (57), 524 (60), 555 (72), for color, magnitude, parity and phonology tasks, respectively
13 (**Figure 1-A**). As shown in **Table 2**, all the comparisons between mean RTs in the different
14 tasks were significant. Moreover, number significantly interacted with task [$F(21, 651) = 12.30,$
15 $p < 0.001, \epsilon_{GG} = 0.43, \text{generalized } \eta^2 = 0.03$]. Number and response hand also significantly
16 interacted [$F(7, 217) = 16.03, p < 0.001, \epsilon_{GG} = 0.58, \text{generalized } \eta^2 = 0.02$], indicating a
17 SNARC effect. Crucially, the significant three-way interaction number x response hand x task
18 indicates that the SNARC effect varied as a function of task [$F(21, 651) = 3.01, p = 0.01, \epsilon_{GG}$
19 $= 0.28, \text{generalized } \eta^2 = 0.007$] (**Figure 2**).

20



21

22 **Figure 1.** Mean RTs (panel A) and mean slopes (panel B) across the tasks. Magnitude: magnitude classification
23 task; Parity: parity judgement task; Phonology: phoneme detection task; Color: color judgement task. Error bars
24 represent the standard error of the mean. Mean RTs of the tasks are all significantly different from each other (all

1 $p_s < 0.001$). Except for the difference between magnitude and parity, all the other mean slopes are significantly
 2 different from each other (all $p_s < 0.05$). In panel B, asterisks indicate that the mean slope is significantly different
 3 from zero (all $p_s < 0.05$).

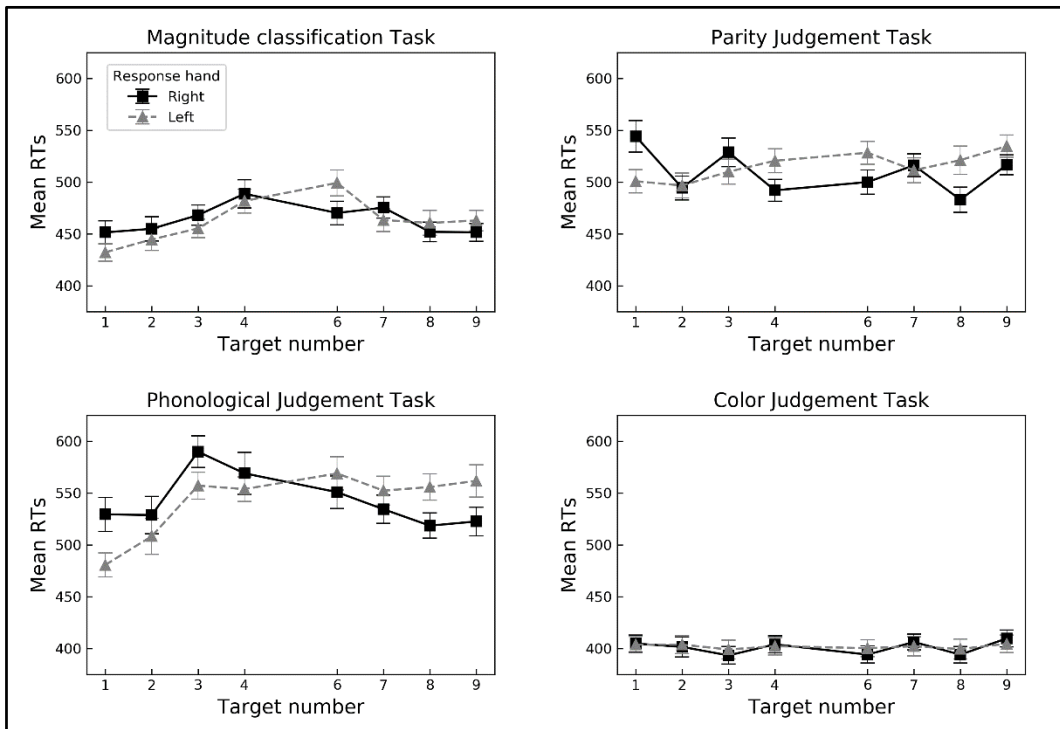
4
 5

6 **Table 2.** T-tests (one-tailed) comparing the tasks (slope and RTs) against each other and assessing whether
 7 the slope distributions were different from zero.

Tasks	Mean	SD	<i>t</i>	df	<i>p</i> -value	Cohen's <i>d_z</i>	Hedges <i>g_{av}</i>
Reaction times							
Magnitude vs. Parity	–	–	-8.97	31	< 0.001	1.58	0.82
Magnitude vs. Phonology	–	–	-9.86	31	< 0.001	1.74	1.19
Magnitude vs. Color	–	–	8.33	31	< 0.001	1.47	1.20
Parity vs. Phonology	–	–	-3.62	31	0.01	0.64	0.44
Parity vs. Color	–	–	14.41	31	< 0.001	2.55	2.08
Phonology vs. Color	–	–	15.41	31	< 0.001	2.72	2.33
Slopes							
Magnitude	-4.11	8.10	-2.87	31	0.018	0.51	0.70
Parity	-5.92	4.79	-6.98	31	< 0.001	1.23	1.70
Phonology	-10.07	8.95	-6.36	31	< 0.001	1.12	1.55
Color	-0.64	3.75	-0.97	31	0.273	0.17	0.24
Magnitude vs. Parity	–	–	1.12	31	0.273	0.20	0.27
Magnitude vs. Phonology	–	–	3.75	31	0.002	0.66	0.68
Magnitude vs. Color	–	–	-2.47	31	0.029	0.44	0.54
Parity vs. Phonology	–	–	2.59	31	0.029	0.46	0.56
Parity vs. Color	–	–	-4.90	31	< 0.001	0.87	1.20
Phonology vs. Color	–	–	-5.89	31	< 0.001	1.04	1.34

8 All *p*-values have been corrected with Holm's method. For the calculation of the effect sizes (Cohen's *d_z* and
 9 Hedges *g_{av}*) refers to Lakens (2013).

10
 11



12 **Figure 2.** Mean RTs across numbers as a function of response hand (right-hand in black, left-hand in grey),
 13 separately for each task. Error bars represent the standard error of the mean.

14
 15
 16

1 3.2. Slopes

2 To further investigate how the SNARC effect varies in the different tasks², the slope³ was
3 analyzed with a repeated measures ANOVA with task (magnitude, parity, phonology, color) as
4 within-subject factor. Slope was significantly influenced by task [$F(3, 93) = 13.78, p < 0.001,$
5 $\epsilon_{GG} = 0.87,$ generalized $\eta^2 = 0.21$] (**Figure 1-B**). Two sets of t-tests were performed to further
6 explore the effect of task on slope (**Table 2**). First, one-sample t-tests, separately computed
7 for each task, were used to assess whether the slope distributions were significantly less than
8 zero. As shown in **Figure 1-B** and **Table 2**, the slopes for the magnitude classification task,
9 the parity judgement task, and the phoneme detection task were significantly different from
10 zero. The slope distribution for the color judgement task was not significantly different from
11 zero, suggesting the absence of a SNARC effect.

12 Second, two-sample t-tests were used to compare the slope distributions for the different
13 tasks. Except for the magnitude vs. parity comparison, all comparisons were significant (**Table**
14 **2**). These results confirm that there was no SNARC effect for the color judgement task, a
15 medium effect for the magnitude classification and the parity judgement tasks, and the
16 strongest effect for the phoneme detection task. It is worth noting that the pattern of results for
17 the slopes (i.e., the strength of the SNARC effect) mirrors the RT distributions (see panels A
18 and B in **Figure 1**). The similarity between the results of RTs and of slopes suggests that the
19 SNARC effect is due to the differences in latency rather than to the level of semantic
20 processing. This is also suggested by the unexpected pattern in the slopes: the phoneme
21 detection showed the strongest SNARC effect and the magnitude classification and parity
22 judgement tasks a smaller effect (**Figure 1-B**). The relationship between the slope distributions
23 of the different tasks were further investigated with Pearson's correlation coefficient (**Table 3**
24 and **Figure 3**). In line with previous studies ($\rho = 0.08$ in Georges, Hoffmann, & Schiltz, 2017
25 and $\rho = 0.2$ in Georges, Hoffmann, & Schiltz, 2018), the magnitude classification and the parity
26 judgement tasks did not correlate. The only significant positive correlation emerged between
27 the magnitude classification and phoneme detection tasks. Although this correlation is

² Split-half reliability was computed for each task. Trials were odd-even half-split based on the order of appearance, separately for each participant, block (response mapping A vs. B) and relevant category (i.e., magnitude classification task: smaller vs. larger than 5; parity judgement task: even vs. odd; phoneme detection task: with vs. without the phonemes; color judgement task: red vs. green). Then, the participant's slope was calculated for each of the two halves. Finally, the Pearson correlation coefficients (r_S) between the slopes of the two halves of each task were calculated and corrected with the Spearman-Brown formula: $r_{SB} = 2r_S / (1 + r_S)$. The Spearman-Brown corrected correlation coefficients were: $r_{SB} = 0.86$ for the magnitude classification task; $r_{SB} = 0.83$ for the parity judgement task; $r_{SB} = 0.79$ for the phoneme detection task; $r_{SB} = 0.58$ for the color judgement task. These coefficients are comparable to or higher than those reported in previous studies (Cipora & Nuerk, 2013; Georges, Hoffmann & Schiltz, 2017, 2018). These results indicated excellent reliability for the magnitude classification task, the parity judgement task and the phoneme detection task, and poor reliability for the color judgement task, which may be due to the absence of a SNARC effect for this task.

³ We also calculated the slopes based on the z-standardized RTs and the same results were found.

1 consistent with the MNL account, it does not fully support the expected strong correlations
 2 between the SNARC effects.

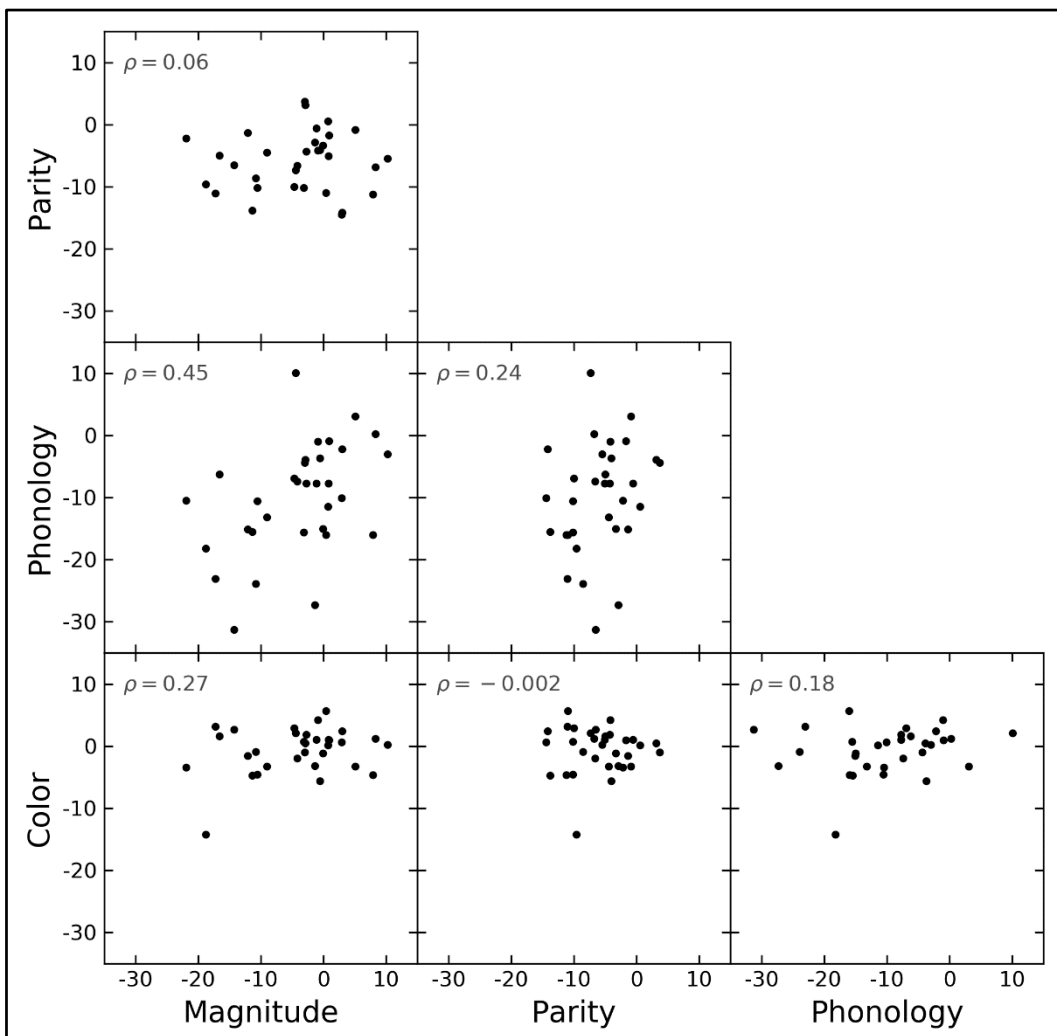
3

4 **Table 3.** Pearson's correlation coefficients between slope distributions of the four tasks
 5 (below the diagonal) and between slope and RT distributions (the diagonal).

	Magnitude	Parity	Phonology	Color
Magnitude	-0.31 ⁺ (-0.59, 0.04)	-	-	-
Parity	0.06 (-0.30, 0.40)	-0.63* (-0.80, -0.36)	-	-
Phonology	0.45* (0.12, 0.69)	0.24 (-0.11, 0.55)	-0.36* (-0.63, -0.02)	-
Color	0.27 (-0.09, 0.57)	-0.002 (-0.35, 0.35)	0.18 (-0.18, 0.50)	-0.30* (-0.60, 0.06)

The 95% confidence intervals are presented in parentheses. ⁺ $p < 0.1$, * $p < 0.05$.

6
7



8

9 **Figure 3.** Scatter plots of the correlation between the SNARC effects of the four tasks. Dots represent the
 10 participant's mean slope. Magnitude: magnitude classification task; Parity: parity judgement task; Phonology:
 11 phoneme detection task; Color: color judgement task. Pearson's correlation coefficients (ρ) are reported in the upper
 12 right corner.

13

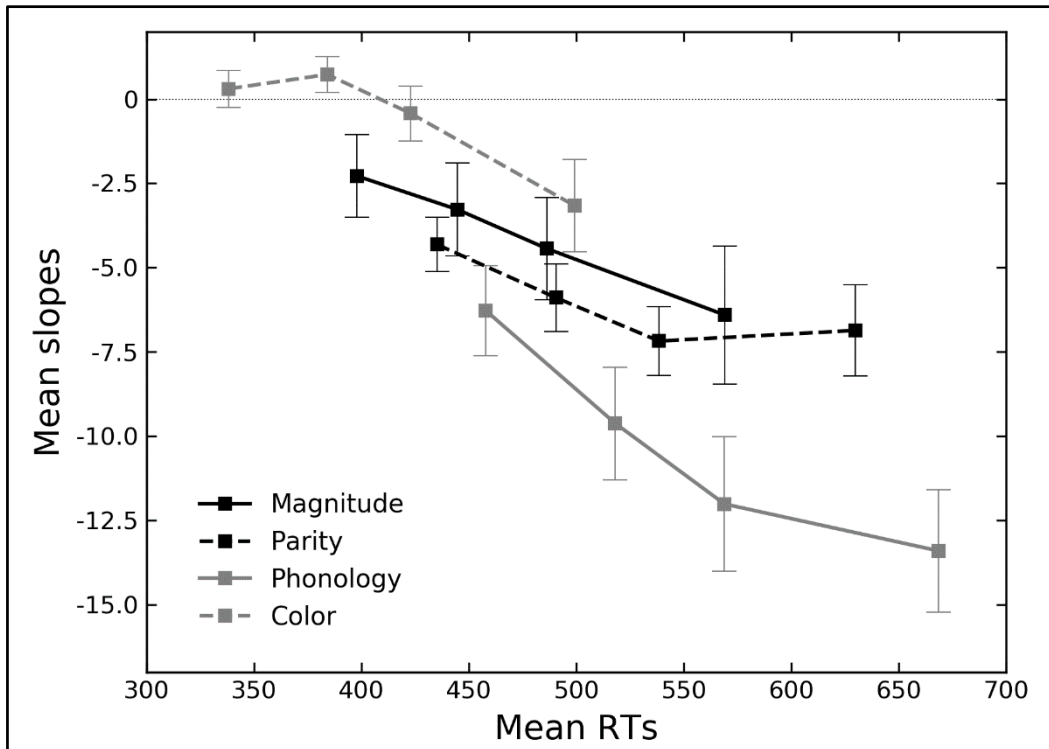


Figure 4. Mean slopes across tasks and bins as a function of RTs. Magnitude: magnitude classification task; Parity: parity judgement task; Phonology: phoneme detection task; Color: color judgement task. For each task, the mean slopes of the four bins are presented: from fastest RTs (leftmost point) to slowest RTs (rightmost point). Error bars represent the standard error of the mean.

The absence of a SNARC effect for the color judgement task may be due to either the very short RTs (i.e., the number processing did not have time to interfere with the response selection) or the fact that a minimum number processing is required in order to produce the SNARC effect and in this task numbers do not reach a sufficient level of processing. To further investigate this possibility, for each participant and task, RTs were rank ordered and divided into 4 bins (Gevers et al., 2006; Ratcliff, 1979). In each bin, slopes were recalculated following the same procedure described above. **Figure 4** shows the mean slope distributions across tasks and the bins as a function of RTs and **Table 4** reports mean RTs, mean slopes and one-sample t-tests assessing whether the slope distributions were significantly less than zero, separately computed for each bin and task. For the parity judgement tasks and the phoneme detection task, a robust SNARC effect emerged in all the bins. The magnitude classification task showed a SNARC effect only in the two slowest bins, a tendency in bin 2 and no effect for the fastest RTs. For the color judgement task, a tendency toward significance emerged only for the bin with longest RTs. All the tasks showed the same patterns: mean slope and effect size increased along with response latency. To confirm this pattern, for each task, repeated measure correlation coefficients (Bakdash & Marusich, 2017, 2018) were calculated between mean slopes and mean RTs. All tasks showed a significant negative correlation between slope and response latency: $r = -0.48$ (95% confidence interval = $-0.62, -0.30$; $p < 0.001$) for the

1 magnitude classification task; $r = -0.32$ (95% CI = $-0.49, -0.13$; $p < 0.01$) for the parity
 2 judgement task; $r = -0.55$ (95% CI = $-0.68, -0.40$; $p < 0.001$) for the phoneme detection task;
 3 $r = -0.46$ (95% CI = $-0.61, -0.28$; $p < 0.001$) for the color judgement task. Furthermore, a
 4 negative correlation also emerged between the overall mean slopes and overall mean RTs
 5 (**Table 3** and **Figure 5**). In fact, both parity judgement and phoneme detection tasks showed
 6 a significant negative correlation between overall mean slopes and overall mean RTs, and
 7 magnitude classification and color judgement tasks showed a tendency for a negative
 8 correlation. These negative correlations further support the idea that the SNARC effect
 9 increases along with response latency.

10

11

Table 4. Statistics for the four bins and t-tests assessing whether the slopes differed from zero.

Bins	Mean RTs (SD)	Mean slope (SD)	<i>t</i>	df	<i>p</i> -value	Cohen's <i>d</i> _z	Hedges <i>g</i> _{av}
Magnitude							
Bin 1	398 (46)	-2.28 (6.95)	-1.86	31	> 0.1	0.33	0.45
Bin 2	445 (53)	-3.27 (7.83)	-2.37	31	0.073	0.42	0.58
Bin 3	486 (60)	-4.44 (8.59)	-2.92	31	0.022	0.52	0.71
Bin 4	569 (77)	-6.41 (11.55)	-3.14	31	0.015	0.55	0.77
Parity							
Bin 1	435 (47)	-4.32 (4.56)	-5.35	31	< 0.001	0.95	1.31
Bin 2	491 (55)	-5.89 (5.65)	-5.90	31	< 0.001	1.04	1.44
Bin 3	538 (61)	-7.18 (5.74)	-7.07	31	< 0.001	1.25	1.72
Bin 4	630 (80)	-6.86 (7.69)	-5.05	31	< 0.001	0.89	1.23
Phonology							
Bin 1	458 (56)	-6.28 (7.59)	-4.68	31	< 0.001	0.83	1.14
Bin 2	518 (66)	-9.62 (9.45)	-5.76	31	< 0.001	1.02	1.41
Bin 3	569 (76)	-12.01 (11.27)	-6.03	31	< 0.001	1.07	1.47
Bin 4	668 (92)	-13.40 (10.27)	-7.39	31	< 0.001	1.31	1.80
Color							
Bin 1	338 (37)	0.31 (3.10)	0.56	31	> 0.1	0.10	0.14
Bin 2	384 (40)	0.73 (3.04)	1.36	31	> 0.1	0.24	0.33
Bin 3	423 (45)	-0.42 (4.59)	-0.52	31	> 0.1	0.09	0.13
Bin 4	499 (61)	-3.16 (7.78)	-2.30	31	0.073	0.41	0.56

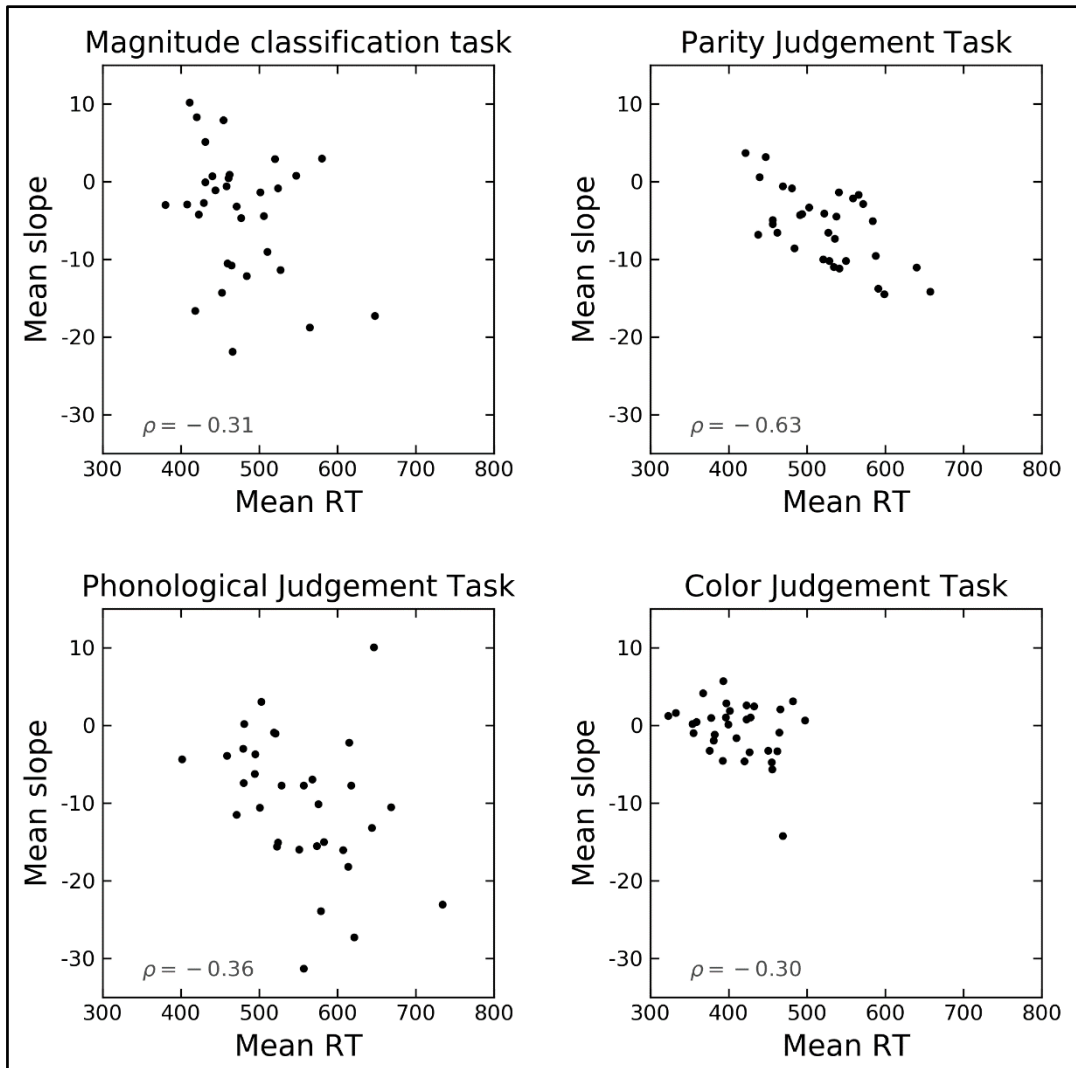
12

13

14

15

All *p*-values have been corrected with Holm's method. For the calculation of the effect sizes (Cohen's *d*_z and Hedges *g*_{av}) refers to Lakens (2013).



1
 2 **Figure 5.** Scatter plots of the correlation between the SNARC effects (mean slope) and response latency (mean
 3 RT) in the four tasks. Dots represent the participant's mean slope and RT. Pearson's correlation coefficients (ρ) are
 4 reported in the bottom left corner.

5
 6 **3.3. Shape of the SNARC effect**

7 According to the dual-route account (Gevers et al., 2006), the SNARC effect should have a
 8 categorical shape in the magnitude classification task and a linear shape in the other tasks. To
 9 test this prediction, we assessed whether mean dRTs were best fitted by a categorical predictor
 10 or a continuous predictor (for a similar analysis see, for example, Gevers et al., 2006). For
 11 each task, we performed a stepwise linear regression analysis (by AIC, with both forward and
 12 backward selection method) with mean dRT as dependent variable and two variables entered
 13 in the model as possible predictors: a categorical predictor (a binary variable with the values 1
 14 = "smaller than 5" and 2 = "larger than 5") and a continuous predictor (a continuous variable
 15 with the values 1–9, excluding 5). Mean dRTs were calculated aggregating across participants.

16 **Table 65** reports the models with the variables selected by the analysis. As expected according
 17 to the dual-route account, for the magnitude classification task the categorical predictor was

1 selected by the analysis, whereas for the parity judgement task and the phoneme detection
 2 task the continuous predictor provided the best fit (**Figure 6**). Due to the lack of a SNARC
 3 effect, no predictor was selected for the color judgement task (**Figure 6**).

Table 45. Results of the stepwise linear regression analysis.

Selected predictor	β	SE	t	p-value	Adj. R ²
Magnitude classification task					0.57
Intercept	15.14	5.23	2.90	0.028	
Categorical	-23.59	7.39	-3.19	0.019	
Parity judgement task					0.37
Intercept	22.61	15.02	1.51	0.183	
Continuous	-5.92	2.63	-2.25	0.066	
Phoneme detection task					0.91
Intercept	52.41	6.58	7.96	< 0.001	
Continuous	-10.07	1.15	-8.72	< 0.001	
Color judgement task					–
Intercept	-1.73	2.00	-0.87	0.415	

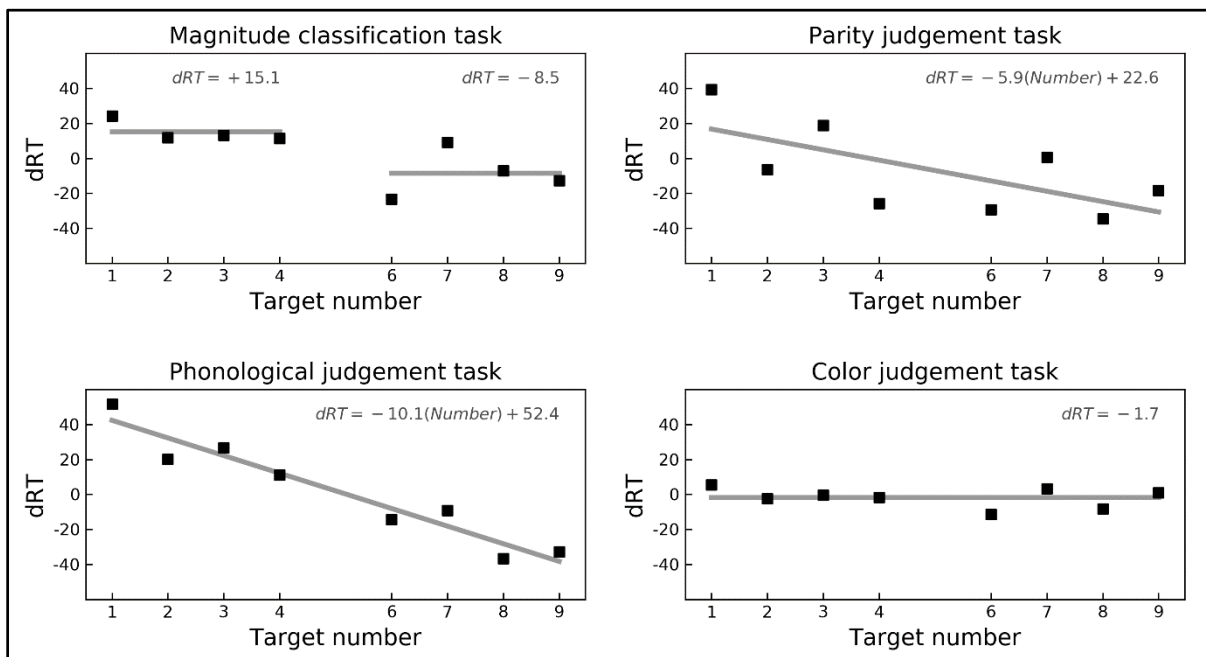


Figure 6. Mean dRTs across numbers, separately for each task. Gray lines represent the model selected in the stepwise analysis: a categorical predictor for the magnitude task; a continuous predictor for the parity and phonology tasks; a model with only intercept for the color task. Above each line, the corresponding model is reported.

4. DISCUSSION

The present study aimed to investigate whether the strength of the SNARC effect is influenced by the amount of semantic number processing required by a task. Participants performed two tasks requiring deep semantic number processing (magnitude classification task and parity judgement task) and two tasks requiring to evaluate non-semantic features of

1 the numbers (color judgement task and phoneme detection task). Results do not provide
2 evidence for the idea that the level of semantic processing influences the SNARC effect.
3 Conversely, the strength of the SNARC effect was proportional to the overall RTs (the slower
4 the RT, the larger the effect, see **Figure 1**), suggesting that response latency is the main factor
5 that affects the strength of the SNARC effect. The phoneme detection task had both the longest
6 RTs and the strongest SNARC effect (mean RTs: 555 ms, mean slope: -10.07). The
7 magnitude classification task (mean RTs: 475 ms, mean slope: -4.11) and the parity
8 judgement task (mean RTs: 524 ms, mean slope: -5.92) had faster RTs and showed a weaker
9 SNARC effect of comparable size. The color judgement task had the fastest RTs and showed
10 no evidence of a SNARC effect (mean RTs: 411 ms, mean slope: -0.64). The influence of
11 response latency was also confirmed by the analysis on the rank ordered RTs, which showed
12 that the strength of the SNARC effect was proportional to RTs. In fact, the color judgement
13 task showed a tendency toward significance when the slowest RTs were analyzed (bin 4, mean
14 RTs: 490 ms, mean slope: -3.16) and there was no evidence of a SNARC effect for the fastest
15 RTs of the magnitude classification task (bin 1, mean RTs: 398 ms, mean slope: -2.28). The
16 overall slopes also negatively correlated with overall RTs in the parity judgement and phoneme
17 detection tasks and a tendency emerged in the magnitude classification and color judgement
18 tasks. Although the correlation analyses reported in this paper may be affected by the low
19 sample size ($N = 32$), they are consistent with those reported by Cipora and colleagues (2019),
20 who analyzed the results from a massive sample ($N > 1,000$) in an online assessment of the
21 SNARC effect in a parity judgement task and found a negative correlation between slope and
22 RT. Considering the relationship between the SNARC effects evoked by the four tasks, a
23 positive correlation ($\rho = 0.45$) between slopes emerged only for the magnitude classification
24 and phoneme detection tasks. Finally, the analysis of the shape of the SNARC effect showed
25 that the dRTs of the magnitude classification task were best fit by a categorical predictor, while
26 those of the parity judgement task and phoneme detection task by a continuous predictor.

27 The MNL account explains the SNARC effect as the result of the association between
28 response locations (left vs. right) and the position of numbers on their spatially oriented
29 representation (small/large numbers on the left/right, respectively) (Dehaene et al., 1993;
30 Hubbard et al., 2005). Therefore, this account predicts that the deeper the access to the MNL,
31 the stronger the influence of the spatial organization of this mental representation should be.
32 Namely, tasks associated with deeper semantic processing (magnitude classification task and
33 parity judgement task) should show a stronger SNARC effect compared to tasks requiring no
34 semantic processing (phoneme detection task and color judgement task). In the latter two
35 tasks, the MNL is thought to be automatically activated (Dehaene et al., 1993; Nuerk, Wood,
36 & Willmes, 2005) and thus the space-number association should, to some extent, generate a
37 SNARC effect. However, it should be smaller in size compared to tasks that explicitly require

1 the processing of semantic aspects of numbers. Our results clearly point in another direction.
2 In fact, not only is the strength of the SNARC effect not proportional to the amount of semantic
3 processing, but also the pattern that emerges (i.e., phoneme detection > parity \approx magnitude >
4 color) cannot be explained by the MNL account. Consistent with a growing amount of evidence
5 showing that the SNARC effect is much more flexible than one would expect under the
6 hypothesis that it is generated from a long-term memory representation (Bächtold et al., 1998;
7 Fias & van Dijck, 2016; Fias et al., 1996; Georges, Hoffmann, & Schiltz, 2017, 2018; Shaki &
8 Fisher, 2008, 2018; van Dijck & Fias, 2011), our results provide additional empirical evidence
9 that the MNL account cannot adequately explain.

10 The dual-route model explains the SNARC effect as the result of the
11 competition/cooperation between two parallel information processing streams (Gevers et al.,
12 2006, 2010; Santens & Gevers, 2008). Since both the unconditional route (i.e., an automatic,
13 long-term space-number mapping) and a conditional route (i.e., an arbitrary task-specific
14 mapping) are always simultaneously activated (Gevers et al., 2006), the SNARC effect should
15 not be affected by the amount of semantic processing required by the task. However, the
16 interference of the unconditional route increases along with response latency, predicting a
17 stronger SNARC effect for longer RTs. Our results provide evidence for this model, showing
18 that the strength of the SNARC effect was influenced by response latency, whereas the amount
19 of semantic processing required by a task had negligible influence. The significant negative
20 correlations between slopes and both overall RTs (**Figure 5**) and rank ordered RTs (**Figure 4**)
21 also support the idea that the SNARC effect is mainly influenced by response latency. Our
22 findings are not consistent with the conclusions by Wood and colleagues (2008). In fact, the
23 authors found that “*in tasks involving semantic number processing such as parity decision*
24 *task, the SNARC effect tended to be larger than in non-semantic tasks*” (Wood et al., 2008, p.
25 500). This inconsistency might be due to the fact that we compared the SNARC effect of
26 different tasks in a within-participant design, whereas Wood and colleagues compared different
27 samples. To disentangle the influence of semantic processing and response latency on the
28 SNARC effect, future studies could manipulate the difficulty (and thus RTs) of the task and
29 evaluate whether the SNARC effect increases proportionally to RTs.

30 According to the MNL account, the SNARC effect originates from a common long-term
31 memory representation regardless of task requirements. Given this common origin, one would
32 expect to observe a strong correlation between the SNARC effects of the different tasks.
33 However, a correlation emerged only between the slopes of the magnitude and phonology
34 tasks (**Figure 3**). Moreover, the two tasks associated with deep semantic processing
35 (magnitude classification and parity judgement tasks) did not correlate at all ($\rho = 0.06$). The
36 lack of strong correlations is however consistent with the dual-route model and the WM
37 account. The dual-route model assumes that a parallel information processing, which may

1 result in competition or cooperation, occurs between the unconditional route and a task-
2 specific conditional route. Therefore, the interaction between the two routes may vary with
3 different tasks and thus produce uncorrelated SNARC effects. The WM account assumes that
4 the SNARC effect is the result of a temporary binding of numbers to a spatial working memory
5 template. Since the spatial orientation of the number sequence is context- and task-specific,
6 this account is also consistent with uncorrelated SNARC effects. However, it is also possible
7 that the lack of correlation depends on the current sample size and thus this result calls for
8 caution. Future studies with a larger sample size should confirm or reject this result.

9 The absence of a SNARC effect in the color judgement task might be explained assuming
10 that the decision was made before the automatic activation of the MNL could interfere with the
11 response selection. It is certainly true that this task presents a confound between response
12 latency and amount of semantic processing, inasmuch very fast RTs were combined with a
13 decision process that could be performed without processing the Arabic numerals. Two
14 objections can be made against this argument. First, it is well-established that it is virtually
15 impossible to avoid processing written words or symbols. The color-stroop task is the most
16 prominent example amongst many others. Hence, we do not think that it is valid to assume
17 that the number symbol was not processed at all in the color task. Second, ~~However,~~ this
18 confound is absent in the other tasks. In fact, the phoneme detection task presented the
19 longest response latencies but required no semantic processing; whereas the magnitude
20 classification and parity judgement task had halfway response latencies but required deep
21 semantic processing. Therefore, the MNL account would predict a stronger SNARC effect in
22 the phoneme detection task compared to the color judgement task, because the longer RTs of
23 the former would provide the spatial orientation of the MNL with enough time to interfere with
24 the response selection. However, the highest interference (strongest SNARC effect) is
25 expected in the magnitude task, given that the semantic representation at the origin of the
26 SNARC effect is also the reference frame for the response selection. In fact, the activation of
27 the number on the semantic representation is required to classify the stimulus magnitude. Given
28 the absence of a confound between amount of semantic processing and response latency in
29 the magnitude classification and phoneme detection tasks, the overall results support the idea
30 that the SNARC effect is modulated by response latency rather than semantic processing.
31 However, future studies should further investigate the confound ~~present-introduced by~~ in the
32 color judgement task by ~~for example~~ manipulating the processing of the target number. For
33 example, in a go/no-go color judgement task, where participants are asked to discriminate the
34 color only for a specific subset of the number range (e.g., go trials: 1, 4, 6, 9).

35 We interpreted our results as evidence in favor of the dual-route model. However, is there
36 an alternative interpretation that would allow to explain these results within the MNL
37 framework? In principle the answer is yes. In fact, along with the idea that the deeper the

1 semantic processing required by a task, the stronger the in interference from the spatial
2 orientation of the MNL is, one could also assume that this interference increases proportionally
3 to response latency. Similar to what follows from proposed for the dual-route model, the longer
4 the decision process takes, the higher the chance for the spatial orientation of the MNL to
5 interfere with the response selection is. Although this seems a plausible explanation, we
6 believe that our results do not support this interpretation. The MNL account and the dual-route
7 model make divergent predictions for the relationship between the SNARC effect and the rank
8 ordered RTs in the bin analysis. If one assumes that the interference of the semantic
9 representation increases with the duration of the decision process, the MNL account predicts
10 that the strength of the SNARC effect increases proportionally to the RTs in the parity
11 judgement, phoneme detection and color judgement tasks. This is indeed consistent to what
12 we observed in the bin analysis (see **Figure 4** and **Table 4**). On the contrary, for the magnitude
13 classification task, the strength of the SNARC effect should remain constant and not be
14 affected by response latency. In fact, in this task, the semantic representation must be explicitly
15 activated, since it is strictly required to classify the magnitude of the number, and thus its
16 activation level should not be modulated by response latency. However, this prediction is
17 inconsistent with the results of the magnitude classification task: no SNARC effect in bin 1, a
18 tendency in bin 2, and a SNARC effect emerging only in bin 3 and 4 (see **Figure 4** and **Table**
19 **4**). At this point, it is worth drawing attention to a fundamental difference between the two
20 accounts. Contrary to the MNL account, increasing strength of the SNARC effect with longer
21 response latencies in the magnitude task is an important feature of the dual-route model and
22 explains the shape of the SNARC effect in this task (see Introduction and next paragraph). To
23 sum up, even an alternative interpretation of the MNL account cannot explain the observed
24 relationship between the SNARC effect and the rank ordered RTs.

25 Further evidence in favor of the dual-route model is also provided by the analysis of the
26 shape of the SNARC effect. Consistent with the prediction of this model (Gevers et al., 2006),
27 we found a categorical shape for the SNARC effect in the magnitude classification task and a
28 linear shape for the other tasks. This result is consistent with the meta-analysis of Wood and
29 colleagues (2008), who found that in 60% of the studies included in their review a categorical
30 predictor alone could account for the SNARC effect in the magnitude comparison task. The
31 categorical shape of the SNARC effect in the magnitude classification task also provides
32 further evidence for the idea that response latency has a major impact on this effect. In fact,
33 the categorical shape depends on the fact that the processing of the numbers closer to the
34 fixed standard (i.e., 5) is more time consuming due to the distance effect (Gevers et al., 2006;
35 see also Moyer & Landauer, 1967; Verguts, Fias, & Stevens, 2005). This longer processing
36 provides more time to the unconditional route to interfere with the decision process, and thus
37 generates a stronger SNARC effect for the numbers closer to the fixed standard.

1 The WM account assumes that the SNARC effect emerges by the positioning of numbers
2 within a temporary spatial template loaded in working memory (Abrahamse, van Dijck, & Fias,
3 2016; Fias & van Dijck, 2016; van Dijck & Fias, 2011; van Dijck, Gevers, & Fias, 2009; van
4 Dijck et al., 2012, 2014). The space-number association results from the fact that this template
5 is spatially oriented from left to right (small to large numbers). Since the space-number
6 association is not an intrinsic property of the number representation but it is instead **temporary**
7 **temporarily** created in working memory (Abrahamse, van Dijck, & Fias, 2016), this account
8 also does not predict that deeper semantic processing generates a stronger SNARC effect.
9 Although this account does not specify whether and how response latency affects the strength
10 of the SNARC effect, it is not incompatible with the results reported in the current study.
11 However, it should be specified which mechanism is responsible for the influence of response
12 latency on the strength of the SNARC effect. The WM account can also explain the categorical
13 shape of the SNARC effect by assuming that the spatial template temporarily loaded in working
14 memory is subject to the distance effect. However, it is worth mentioning that, according to the
15 WM account, the distance effect is generated in working memory, whereas for the dual-route
16 model this effect results from a feature of the long-term memory representation of the numbers
17 (Abrahamse, van Dijck, & Fias, 2016).

18 Despite the very low error rate, the RTs of the phoneme detection task were significantly
19 longer than those of the other tasks. This task required to judge the presence or absence of a
20 group of three phonemes in the number words and thus to transcode the stimulus between
21 two symbolic notations: Arabic to phonological (Nuerk, Moeller, & Willmes, 2015; Silke Göbel
22 et al., 2014). Namely, the visually presented Arabic number had to be transcoded into the
23 internal phonological representation, on which then the search for a specific phoneme could
24 be performed. The phoneme detection task was therefore more demanding compared to the
25 other three tasks and this could explain why this task had the longest RTs and thus the
26 strongest SNARC effect.

27 The tasks investigated in the current study required explicit spatial-directional processing
28 (i.e., response keys were lateralized) and explicit (magnitude classification task) or implicit
29 (parity judgement task, color judgement task, and phoneme detection task) magnitude
30 processing (for a more detailed description of this task classification see Shaki & Fischer,
31 2018). In a recent study, Shaki and Fischer (2018) provided evidence that to some extent the
32 SNARC effect might be an artifact of the procedure used to measure it. In fact, using magnitude
33 classification and parity judgement tasks with go/no-go responses, Shaki and Fischer showed
34 that ~~the~~ horizontal **SNARC-SNAs** requires explicit activation of spatial representation (in
35 stimulus or response code) or explicit magnitude processing. On the other hand, ~~a~~ vertical
36 **SNARCs-effect** emerged also with implicit spatial and magnitude processing. Consistent with
37 our results, Shaki and Fischer's findings are also difficult to reconcile with the assumption of a

1 left-to-right horizontally organized long-term memory representation of numbers. However, the
2 results reported in the current manuscript might be limited to tasks with left/right horizontally
3 distributed response keys. Future studies should investigate whether our results are limited
4 only to tasks that require an explicit spatial-directional processing or whether they are also
5 valid for tasks with implicit spatial processing (e.g., go/no-go responses).

6 In sum, the results reported in the present study suggest that the strength of the SNARC
7 effect is mainly modulated by overall response latencies rather than by the amount of semantic
8 processing required by the tasks. These results are in line with the prediction of the dual-route
9 model and can also be explained within the framework of the WM account. On the other hand,
10 our data provide evidence against the idea that deeper semantic processing generates a
11 stronger SNARC effect, as it would be expected according to the MNL account.

12
13
14
15

1 **Acknowledgement**^[DD3]

2 This work was supported by a grant (DI 2361/1-1) from Deutsche Forschungsgemeinschaft
3 (DFG, German Research Council) to Daniele Didino.

4

5

6

1 REFERENCES

- 2
- 3 Abrahamse, E., van Dijck, J.-P., & Fias, W. (2016). How Does Working Memory Enable
4 Number-Induced Spatial Biases? *Frontiers in Psychology*, 7, 977.
- 5 Bächtold, D., Baumüller, M., & Brugger, P. (1998). Stimulus–response compatibility in
6 representational space. *Neuropsychologia*, 36(8), 731–735
- 7 Bakdash, J.Z., & Marusich, L.R. (2017). Repeated Measures Correlation. *Frontiers in*
8 *Psychology*, 8, 456. doi: 10.3389/fpsyg.2017.00456
- 9 Bakdash, J.Z., & Marusich, L.R. (2018). rmcrr: Repeated Measures Correlation. R package
10 version 0.3.0. <https://CRAN.R-project.org/package=rmcrr>
- 11 Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial vision*, 10(4), 443–446.
- 12 Cipora, K., & Nuerk, H. C. (2013). Is the SNARC effect related to the level of mathematics? No
13 systematic relationship observed despite more power, more repetitions, and more direct
14 assessment of arithmetic skill. *The Quarterly Journal of Experimental Psychology*, 66(10),
15 1974-1991. doi: 10.1080/17470218.2013.772215
- 16 Cipora, K., Soltanlou, M., Reips, U.-D., Nuerk, H.-C. (2019). The SNARC and MARC effects
17 measured online: Large-scale assessment methods in flexible cognitive effects. *Behavior*
18 *Research Methods*. doi: 10.3758/s13428-019-01213-5
- 19 Cipora, K., & Wood, G. (2017). Finding the SNARC instead of hunting it: A 20*20 monte carlo
20 investigation. *Frontiers in Psychology*, 8, 1194.
- 21 Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number
22 magnitude. *Journal of Experimental Psychology: General*, 122(3), 371–396.
- 23 Dehaene, S., Naccache, L., Le Clec'H, G., Koechlin, E., Mueller, M., Dehaene-Lambertz, G.,
24 van de Moortele, P.-F., & Le Bihan, D. (1998). Imaging unconscious semantic priming.
25 *Nature*, 395(6702), 597–600. doi: 10.1038/26967
- 26 Fias, W. (2001). Two routes for the processing of verbal numbers: Evidence from the SNARC
27 effect. *Psychological Research*, 65, 250–259.
- 28 Fias, W., Brysbaert, M., Geypens, F., & D'Ydewalle, G. (1996). The importance of magnitude
29 information in numerical processing: Evidence from the SNARC effect. *Mathematical*
30 *Cognition*, 2(1), 95–110.
- 31 Fias, W., & van Dijck, J.-P. (2016). The temporary nature of number—space interactions.
32 *Canadian Journal of Experimental Psychology*, 70(1), 33–40.
- 33 Fischer, M. H., & Shaki, S. (2014). Spatial associations in numerical cognition – From single
34 digits to arithmetic. *Quarterly Journal of Experimental Psychology*, 67(8), 1461–1483.
- 35 Georges, C., Hoffmann, D., & Schiltz, C. (2017). How and why do number-space associations
36 co-vary in implicit and explicit magnitude processing tasks? *Journal of Numerical Cognition*,
37 3(2), 182–211.

- 1 Georges, C., Hoffmann, D., & Schiltz, C. (2018). Implicit and explicit number-space
2 associations differentially relate to interference control in young adults with ADHD. *Frontiers*
3 *in Psychology*, 9, 775.
- 4 Gevers, W., Santens, S., Dhooge, E., Chen, Q., Van den Bossche, L., Fias, W., & Verguts, T.
5 (2010). Verbal-spatial and visuospatial coding of number-space interactions. *Journal of*
6 *Experimental Psychology: General*, 139(1), 180–90.
- 7 Gevers, W., Verguts, T., Reynvoet, B., Caessens, B., & Fias, W. (2006). Numbers and space:
8 A computational model of the SNARC effect. *Journal of Experimental Psychology: Human*
9 *Perception and Performance*, 32(1), 32–44.
- 10 Ginsburg, V., and Gevers, W. (2015). Spatial coding of ordinal information in short-and long-
11 term memory. *Frontiers in Human Neuroscience*. 9:8. doi: 10.3389/fnhum.2015.00008
- 12 Göbel, S. M. (2015). Up or down? Reading direction influences vertical counting direction in
13 the horizontal plane—a cross-cultural comparison. *Frontiers in psychology*, 6, 228.
- 14 Göbel, S. M., Maier, C. A., & Shaki, S. (2015). Which numbers do you have in mind? Number
15 generation is influenced by reading direction. *Cognitive processing*, 16(1), 241–244.
- 16 Göbel, S. M., Shaki, S., & Fischer, M. H. (2011). The cultural number line: a review of cultural
17 and linguistic influences on the development of number processing. *Journal of Cross-*
18 *Cultural Psychology*, 42(4), 543–565.
- 19 Göbel, S. M., Watson, S. E., Lervåg, A., & Hulme, C. (2014). Children’s arithmetic
20 development: It is number knowledge, not the approximate number sense, that counts.
21 *Psychological science*, 25(3), 789–798.
- 22 Greenhouse, S. W., and Geisser, S. (1959). On methods in the analysis of profile data.
23 *Psychometrika*, 24, 95–112.
- 24 Huber, S., Klein, E., Moeller, K., & Willmes, K. (2016). Spatial-Numerical and ordinal positional
25 associations coexist in parallel. *Frontiers in Psychology*, 7, 438. doi:
26 10.3389/fpsyg.2016.00438
- 27 Nuerk, H.-C., Moeller, K., & Willmes, K. (2015). Multi-digit Number Processing: Overview,
28 Conceptual Clarifications, and Language Influences. In Cohen Kadosh, R., & Dowker, A.
29 (Eds.), *Oxford library of psychology. The Oxford handbook of numerical cognition*, New
30 York, NY, US: Oxford University Press, pp. 106–139.
- 31 Hubbard, E. M., Piazza, M., Pinel, P., and Dehaene, S. (2005). Interactions between number
32 and space in parietal cortex. *Nature Reviews Neuroscience*, 6(6), 435–448.
- 33 Koechlin, E., Nacchache, L., Block, E., & Dehaene, S. (1999). Primed numbers: exploring the
34 modularity of numerical representations with masked and unmasked semantic priming.
35 *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1882–
36 1905. doi: 10.1037/0096-1523.25.6.1882

1 Kunde, W., Kiesel, A., & Hoffmann, J. (2003). Conscious control over the content of
2 unconscious cognition. *Cognition*, 88(2), 223-242. doi:10.1016/S0010-0277(03)00023-4

3 Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a
4 practical primer for t-tests and ANOVAs. *Frontiers in psychology*, 4, 863.

5 Maxfield, L. (1997). Attention and semantic priming: A review of prime task effects.
6 *Consciousness and cognition*, 6(2-3), 204-218. doi: 10.1006/ccog.1997.0311

7 Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality.
8 *Nature*, 215(5109), 1519–1520.

9 Naccache, L., & Dehaene, S. (2001). Unconscious semantic priming extends to novel unseen
10 stimuli. *Cognition*, 80(3), 223–237. doi: 10.1016/S0010-0277(00)00139-6

11 Nuerk, H.-C., Wood, G., & Willmes, K. (2005). The universal SNARC effect: The association
12 between number magnitude and space is amodal. *Experimental Psychology*, 52(3), 187–
13 194.

14 Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming
15 numbers into movies. *Spatial vision*, 10(4), 437–442.

16 Pinhas, M., Tzelgov, J., & Ganor-Stern, D. (2012). Estimating linear effects in ANOVA designs:
17 the easy way. *Behavior research methods*, 44(3), 788–794.

18 Proctor, R. W., & Cho, Y. S. (2006). Polarity correspondence: A general principle for
19 performance of speeded binary classification tasks. *Psychological Bulletin*, 132(3), 416–
20 442.

21 R Core Team (2015). *R: A language and environment for statistical computing*. Vienna: R
22 Foundation for Statistical Computing.

23 Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics.
24 *Psychological Bulletin*, 86(3), 446–461.

25 RStudio Team (2015). *RStudio: Integrated development for R*. Boston, MA: RStudio, Inc.

26 Santens, S., & Gevers, W. (2008). The SNARC effect does not imply a mental number line.
27 *Cognition*, 108(1), 263–270.

28 Schroeder, P. A., Nuerk, H. C., & Plewnia, C. (2017a). Prefrontal neuromodulation reverses
29 spatial associations of non-numerical sequences, but not numbers. *Biological Psychology*,
30 128, 39–49. doi: 10.1016/j.biopsycho.2017.07.008

31 Schroeder, P. A., Nuerk, H. C., & Plewnia, C. (2017b). Switching between multiple codes of
32 SNARC-like associations: Two conceptual replication attempts with anodal tDCS in sham-
33 controlled cross-over design. *Frontiers in Neuroscience*, 11, 654. doi:
34 10.3389/fnins.2017.00654

35 Schroeder, P. A., Nuerk, H. C., & Plewnia, C. (2017c). Space in numerical and ordinal
36 information: A common construct? *Journal of Numerical Cognition*, 3(2), 164-181. doi:
37 10.5964/jnc.v3i2.40

- 1 Shaki, S., Fischer, M. H., & Göbel, S. M. (2012). Direction counts: A comparative study of
2 spatially directional counting biases in cultures with different reading directions. *Journal of*
3 *Experimental Child Psychology*, 112(2), 275–281.
- 4 Shaki, S., & Fischer, M. H. (2008). Reading space into numbers: A cross-linguistic comparison
5 of the SNARC effect. *Cognition*, 108(2), 590–599.
- 6 Shaki, S., & Fischer, M. H. (2018). Deconstructing spatial-numerical associations. *Cognition*,
7 175, 109–113.
- 8 Shaki, S., Fischer, M. H., & Petrusic, W. M. (2009). Reading habits for both words and numbers
9 contribute to the SNARC effect. *Psychonomic Bulletin & Review*, 16(2), 328–331.
- 10 Tzelgov, J., Zohar-Shai, B., & Nuerk, H. C. (2013). On defining quantifying and measuring the
11 SNARC effect. *Frontiers in psychology*, 4, 302.
- 12 van Dijck, J.-P., Abrahamse, E. L., Acar, F., Ketels, B., & Fias, W. (2014). A Working Memory
13 Account of the Interaction between Numbers and Spatial Attention. *Quarterly Journal of*
14 *Experimental Psychology*, 67(8), 1500–1513.
- 15 van Dijck, J.-P., & Fias, W. (2011). A working memory account for spatial-numerical
16 associations. *Cognition*, 119(1), 114–119.
- 17 van Dijck, J.-P., Gevers, W., & Fias, W. (2009). Numbers are associated with different types
18 of spatial information depending on the task. *Cognition*, 113(2), 248–253.
- 19 van Dijck, J.-P., Gevers, W., Lafosse, C., & Fias, W. (2012). The Heterogeneous Nature of
20 Number–Space Interactions. *Frontiers in Human Neuroscience*, 5, 182.
- 21 Van Dijck, J.-P., Ginsburg, V., Girelli, L., & Gevers, W. (2015). Linking numbers to space: From
22 the mental number line towards a hybrid account. In Cohen Kadosh, R., & Dowker, A.
23 (Eds.), *Oxford library of psychology. The Oxford handbook of numerical cognition*, New
24 York, NY, US: Oxford University Press, pp. 89–105.
- 25 Verguts, T., Fias, W., & Stevens, M. (2005). A model of exact small-number representation.
26 *Psychonomic Bulletin & Review*, 12(1), 66–80.
- 27 Wood, G., Willmes, K., Nuerk, H.-C., & Fischer, M. H. (2008). On the cognitive link between
28 space and number: a meta-analysis of the SNARC effect. *Psychology Science Quarterly*,
29 50(4), 489–525.

30
31
32