Peer review history of the paper *Not toeing the number line for simple arithmetic: Two large-n conceptual replications of Mathieu et al. (Cognition, 2016, Experiment 1)* by Jamie I. D. Campbell, Yalin Chen & Maham Azhar published in Special issue *Direct and Conceptual Replications in Numerical Cognition* in *Journal of Numerical Cognition* (vol 7, 3), https://doi.org/10.5964/jnc.6051

Table of Contents

Initial decision letter

Dear Dr. Campbell,
I now have received the reviews of two expert reviewers on this topic. While reviewer 2 already recommended publication by minor revision, reviewer 1 has some major concerns. The most important issue is a methodological doubt, which actually comes from the ambiguity of the explained timing in the original paper of Mathieu and colleagues (2016). Namely, there seems to be a difference between the presented timing in Figure 1 (which is similar to the present study) and the text in the original paper. Therefore, the current null finding might be the rejection of the alternative finding of 300 ms SOA of the original study (if the fixation point coming after operator was the stimulus for SOA), or the confirmation of 450 ms SOA of the original study (if the operator was the stimulus). Please note that in both cases, the current study is considered as a replication, but with two different messages. I agree with reviewer 1 to ask the authors of the original study to ensure how the task was displayed to solve this issue. Moreover, reviewer 1 asked for further elaboration on theoretical discussions.
According to the reviewers' comment and my read, I am willing to consider a revision of the manuscript, addressing the issues of the reviewers. I will only send it to reviewer 1. I would like you to include a response letter that details how each of the issues raised by the reviewers was addressed. Looking forward to your revision.

Best wishes,
Mojtaba Soltanlou

Reviewer 1:

In this manuscript, the authors attempted to replicate the Experiment 1 of the study of Mathieu and collaborators (2016) that initially showed facilitation for additions when the second operand is displayed on the right of the screen and for subtractions when the second operand is displayed on the left of the screen. This effect, for both operations, was only significant at a 300 ms SOA. The authors of the present manuscript questioned the reliability of the effect on subtractions based on the calculation of Bayes factor. They,

therefore, decided to test this effect with a larger sample size. The authors also tested in two experiments whether these spatial-numerical associations for operations would be solely observed when additions and subtractions are mixed by including two pure blocks of additions and subtractions. The results seem to show that the second operand position has no effect on subtraction, and depends on the composition of the blocks.
The manuscript is well written and easy to follow. The authors report that this project is conceptual replication because they selected a subsample of trials based on the SOA that showed the effect in the previous study. However, there might be a huge issue concerning the time course of the presentation of the stimuli (see below). If my understanding of the text of the original study is correct, I am afraid that this study does not replicate the methodology of Mathieu and collaborators. I see no other solution than to ask the authors of the first study how the stimuli were displayed. I think that the theoretical background should be better explained, even in the abstract (rapid solving procedure vs. retrieval, single-digit vs multi-digit problems). Aside from this, I have several minor comments that I hope will help at clarifying the paper.

Major issue:
Comment 1: The biggest potential issue here is that I am not convinced that the authors really replicated the study of Mathieu. When reading carefully the published manuscript, it is written “Therefore, we varied the delay between the arithmetic sign and O2 in the present experiment, such that the SOA (stimulus onset asynchrony) was 150, 300, or 450 ms” (p232). My understanding of this sentence is that the operator is the stimulus and therefore that the time that lapses between the onset of the operator and the onset of the O2 is of 150, 300 or 450ms (150 ms with the operator on screen + 0, 150 or 300 ms delay). In the present manuscript, the 300 ms duration was applied to the time that passes after the offset of the stimuli/operator. So, the SOA is in fact of 450 ms. This is a time period in which Mathieu did not find an effect of O2 position. If so, the authors missed the critical time period and confirm the absence of reliable effect at 450ms SOA that Mathieu showed in the previous study. As stated above, I am however not sure that this is what was done in the original study because of a potential contradiction between Figure 1 and what is written in the text. Figure 1 seems to suggest that the 150, 300 and 450 ms are the delays between the offset of the operator and the onset of the O2. I have no better solution than to ask the authors of the first manuscript how the task was displayed to clarify this issue.

Comment 2: The relevance of the study to the field and the theoretical necessity to replicate is not clear enough. I think that several theoretical aspects should be better introduced and discussed. When going back to the original paper, it is still not very clear to me what this paradigm is actually measuring. Mathieu and colleagues suggested that it is because the position of O2 induced a forced shift of attention to a congruent hemifield that the performance was facilitated by accelerating the solving procedure. However, they acknowledge in the Discussion section that “ …more is right” and ‘‘less is left”. In other words, participants might shift their attention to the left or right side of space not because they move along the MNL, but simply because the appearance of the addition or subtraction sign makes them anticipate that the result of the problem will be smaller or larger than the first operand”. The operator would trigger a shift to the right (+) or to the

left (-) and this will accelerate the detection of the O2 in congruent trials because the digit appears in the locus of attention of the subject. This fast detection would result in faster reaction times than when the O2 is displayed on the opposite side of the screen. In this scenario, the solving procedure time would not be modified. It is simply the time to detect the O2 that will be modified. I would suggest the author of this manuscript to comment on the limits of Mathieu's paradigm and not to focus all the paper on potential statistical flaws.

Comment 3: This article aims at disentangling whether small problems are retrieved from memory (e.g., Campbell & Xue, 2001) or would involve fast calculation procedures (e.g., Fayol & Thevenot, 2012). This fast strategy, according to Mathieu, would be to shift attention along a mental number line. The authors of the present paper should consider that most of the work that was published on spatial associations with mental arithmetic concerns problems with multidigit numbers that cannot be retrieved from memory. The distinction should be carefully made all along the manuscript.

Comment 4: The authors should also consider that, as far as I know, all the studies suggesting that spatial attentional shifts are functionally related to the solving procedure of additions and subtractions by investigating the effect of forced spatial attentional shifts concern multi-digit arithmetic problems. This was observed with optokinetic stimulation (Blini et al., 2019; Masson et al., 2016), interference paradigm (Masson & Pesenti, 2015), (in)congruent movements of the body (Wiemers et al., 2014) and with left and right neglect patients (Dormal et al., 2014; Masson et al., 2018). All the studies focusing on small arithmetical problems merely showed that these simple problems could trigger spatial biases (e.g., Marghetis et al., 2014; Masson & Pesenti, 2014; Pinhas & Fischer, 2008). No studies actually showed that these shifts were important to solve this kind of problem or to accelerate the solving procedure. This is true, except for the paper of Mathieu and colleagues, which uses a paradigm that is difficult to interpret (see above). I suggest the authors to clearly state whether their take-home message (i.e., unreliable SNAs for subtraction with Mathieu's paradigm) applies to arithmetic facts or all types of additions and subtractions.

Comment 5: I wonder why the authors tested some participants that are not native English speakers. The language in which arithmetic was learned is more critical than the first language of the participant for arithmetic performance (e.g., van Rinsveld et al., 2015). It is likely that the non-native participants were impaired by the use of English. And this could have added some noise in the data and masked some small but real effects. I noticed that the authors run complementary analysis without those participants but I wonder whether the authors should simply not include them in the sample.

Minor Issues:
Page 3: "Since then, many experiments using a wide range of (other? Different than the SNARC ?) paradigms have confirmed a link between magnitude and space". Most articles that are cited here are actually the same SNARC paradigm. The authors should perhaps mention other behavioral measures of SNAs (Stoianov et al., 2008 in Cognition) or methods such as eye tracker (e.g., Myachykov et al., 2015 in Acta Psychologica;

Salvaggio et al., 2018 in JEP:LMC) and neuropsychology of neglect patients (e.g., Vuilleumier et al., 2004 in Cortex).

Page 3: “spatial effects have also been observed in connection with performance of addition and subtraction”. Some of the papers that are cited here do not concern directly spatial effects. Mathieu et al., 2018 (which is cited twice at the same place) showed an activation of a region related to spatial attention when showing a “+” sign. However, they did not measure spatial attentional shifts. McCrink et al., (2007) reported an over-estimation bias for additions and an under-estimation bias for subtractions (Operational Momentum effect) which has nothing to do, per se, with a spatial shift. There are plenty of other studies that really measured spatial shifts of attention or spatial biases that could be cited here instead of those two.

Page 4: Mathieu was not the first to show evidence of a spatial shift acting on arithmetic performance. Wiemers, Lindemann & Bekkering (2014), in QJEP, already showed that when participants were moving their arms vertically or horizontally, it affected the performance of addition a subtraction solving. Moreover, as stated above, it is not yet clear whether it’s the attentional shift induced by O2 position that facilitates the performance or the attentional shift induced by the operator that facilitates the detection of O2.

Page 16: It is not surprising that the case study of a right neglect patient (Masson, Pesenti, Coyette, Andres, & Dormal, 2017) did not reveal a deficit for subtraction. Indeed, the right neglect patient with left-brain lesion showed more difficulty for addition solving than for subtraction solving, which is coherent with is incapacity to orient attention towards the right side of space. It is when left neglect patients were tested that a deficit for subtraction was observed (Dormal et al., 2014).

Page 17: “… a generalized leftward association for subtraction, assuming it exists”. I find that the “assuming it exists” part of the sentence unnecessary. There is now substantial evidence from different research teams that subtraction can cause leftward shifts (e.g., Liu et al., 2017; Masson & Pesenti, 2014; Marghetis et al., 2014) and that these shifts are functionally related to subtraction solving, at least of multidigit numbers (e.g., Dormal et al., 2014). This sentence is misleading and cast doubts on many studies that actually have nothing to do with this paradigm and this range of arithmetic problems. Please rephrase.

References

Dormal, V., Schuller, A. M., Nihoul, J., Pesenti, M., & Andres, M. (2014). Causal role of spatial attention in arithmetic problem solving: Evidence from left unilateral neglect. Neuropsychologia, 60, 1–9. http://dx.doi.org/10.1016/j.neuropsychologia.2014.05.007

Fayol, M., & Thevenot, C. (2012). The use of procedural knowledge in simple addition and subtraction problems. Cognition, 123(3), 392-403. https://doi.org/10.1016/j.cognition.2012.02.008

Liu, D., Cai, D., Verguts, T., & Chen, Q. (2017). The time course of spatial attention shifts in elementary arithmetic. Scientific Reports, 7(1), 921. doi:10.1038/s41598-017-01037-3

Marghetis, T., Núñez, R., & Bergen, B. K. (2014). Doing arithmetic by hand: Hand movements during exact arithmetic reveal systematic, dynamic spatial processing. The Quarterly Journal of Experimental Psychology, 67, 1579–1596. http://dx.doi.org/10.1080/17470218.2014.897359
Masson, N., & Pesenti, M. (2014). Attentional bias induced by solving simple and complex addition and subtraction problems. The Quarterly Journal of Experimental Psychology, 67, 1514–1526. http://dx.doi.org/10.1080/17470218.2014.903985
Masson, N., & Pesenti, M. (2016). Interference of lateralized distractors on arithmetic problem solving: A functional role for attention shifts in mental calculation. Psychological Research, 80, 640–651. http://dx.doi.org/10.1007/s00426-015-0668-7
Masson, N., Pesenti, M., Coyette, F., Andres, M., & Dormal, V. (2017). Shifts of spatial attention underlie numerical comparison and mental arithmetic: Evidence from a patient with right unilateral neglect. Neuropsychology, 31(7), 822-833. http://dx.doi.org/10.1037/neu0000361
Masson, N., Pesenti, M., & Dormal, V. (2017). Impact of optokinetic stimulation on mental arithmetic. Psychological Research, 81(4), 840-849. http://dx.doi.org/10.1007/s00426-016-0784-z
Myachykov, A., Cangelosi, A., Ellis, R., & Fischer, M. H. (2015). The oculomotor resonance effect in spatial-numerical mapping. Acta Psychologica, 161, 162–169. https://doi.org/10.1016/j.actpsy.2015.09.006
Pinhas, M., & Fischer, M. (2008). Mental movements without magnitude? A study of spatial biases in symbolic arithmetic. Cognition, 109(3), 408-415. doi:10.1016/j.cognition.2008.09.003
Salvaggio, S., Masson, N., & Andres, M. (2018). Eye position reflects the spatial coding of numbers during magnitude comparison. Journal of Experimental Psychology: Learning, Memory, and Cognition. 45(10), 1910-1921. http://dx.doi.org/10.1037/xlm0000681
Stoianov, I., Kramer, P., Umiltà, C., & Zorzi, M. (2008). Visuospatial priming of the mental number line. Cognition, 106, 770–779. http://dx.doi.org/10.1016/j.cognition.2007.04.013
Vuilleumier, P., Ortigue, S., & Brugger, P. (2004). The number space and neglect. Cortex, 40, 399–410. http://dx.doi.org/10.1016/S0010-9452(08)70134-5
Wiemers, M., Bekkering, H., & Lindemann, O. (2014). Spatial interferences in mental arithmetic: Evidence from the motion–arithmetic compatibility effect. The Quarterly Journal of Experimental Psychology, 67(8), 1557-1570. doi:10.1080/17470218.2014.889180

Reviewer 2:

The current study includes two conceptual replication experiments of Experiment 1 in Mathieu et al. (2016, DOI: 10.1016/j.cognition.2015.10.002).
In the original paper, Mathieu et al. found evidence that shift of attention on a mental number line has an important role in mental calculation (addition and subtraction). Addition was associated with rightward attention shift (i.e., moving rightward on a horizontally oriented mental number line) and subtraction with a leftward shift (moving leftward on a MNL).
The current conceptual replication aims to investigate to what extent the effects reported in Mathieu et al study are due to the experimental procedure (presentation sequence). In both replication experiments, additions and subtractions were presented both in pure blocks (only one operation was presented) and mixed blocks (both addition and subtraction were randomly presented in the same block).
Experiment 1 showed an association between addition and rightward attention shift only in mixed blocks, and no effect emerged for subtraction. Experiment 2 showed an association between both operations and rightward attention shift.
The authors conclude that these results do not provide evidence for a relationship between mental arithmetic and movement of attention on a MNL. The authors propose that the effects found in Mathieu et al study are generated by heuristics and strategies that depend on the specific experimental setup.

Evaluation.
The manuscript is well written and the experiments are well designed and executed. The study addresses a relevant question for the field of numerical cognition (are visuospatial attention and mental arithmetic functionally related?). Both experiments included a large sample (74 participants each).
My recommendation is to accept the manuscript while addressing the following minor points. (I agree with an Open Peer Review.)

1. The authors suggest that the effects are due to the use of the more-or-less heuristic (e.g., p. 6). This heuristic can explain the results for addition problems (i.e., RTs advantage when the second operand is presented on the right), but does not explain the results with subtraction (i.e., no effect in experiment 1, and RTs advantage when the second operand is presented on the right in experiment 2). I think that neither the attentional shift on the MNL nor the more-or-less heuristic can fully explain the results. This should be mentioned in the manuscript.
2. T-test analyses. When describing the results reported in Mathieu et al. (2016), I think that the authors of the current manuscript reported two-tailed p-values for the t-test (e.g., page 4). Mathieu et al (2016) reported one-tailed p-values. This difference should be described in the manuscript.
3. Eta-squared for t-tests. Cohen's dz is a more common effect size measure for the t-test. Why did the authors choose to report the eta-squared for the t-tests?
4. Rounding p-value. The number of decimal values used for the p-values is not consistent. For example, "p = .002" and "p = .01" (page 4, lines 13 and 15) use 3 and 2

decimal values. However, I think that the second p-value (p = .01) is “p = .006” if rounded to 3 decimal values. Please, use the same number of decimal values for all the p-values.
5. Page 5, line 11. In the analysis “[-12 ms, t(21) = 0.79, , SE = 15.19, η^2 = .03, BF01 = 3.40]” the p-value (p = 0.438) is not reported.
6. Page 6, line 5. “After presentation of O1 for 1 second the operation […]” should be “After presentation of O1 for 500 ms the operation […]”.
7. Power. I did not understand how power and sample size were calculated. In the sentence “The power is .9 for a sample of 72 or more to detect an operation-specific O2 position effect about half the size observed by Mathieu et al. Experiment 1 (eta_p^2 = .24)”, I did not understand how the partial eta-squared (.24) was calculated.
8. Page 7. I think that the mean age is 20.8 rather than 21.1. Please also report the standard deviation for age.
9. List of problems. A table showing the list of problems (page 7, last paragraph) will make easier for the reader to quickly find the problems used in the experiments.
10. RTs log-transformed. I did not find a supplementary JASP file with the log-transformed RT for experiment 1. There is only a file for experiment 2 (“E2 log RT 4-way n=74.1.jasp”).
11. Outliers. The authors excluded from the analysis RTs that “were more than 4 SD from a participant’s mean RT in each Operation × O2 Position × Mixture cell”. In the original paper, Mathieu et al excluded “RT greater than 2 SDs from the mean for each participant”. This difference should be reported in the manuscript. Also, are the results different if the Mathieu et al outlier procedure is used?
12. Page 10. The 2 t-tests for the comparisons between “slower subtraction in mixed-operation blocks relative to pure-operation blocks (854 vs 822 ms) whereas mean addition RT did not differ between single-operation (987 ms) and mixed-operations blocks (989 ms)” are not reported.
13. The authors should choose the term “pure-operation” or “single-operation” (e.g., page 10) and use it consistently.
14. Page 12 and 15. After reporting the results for the RTs analyses, the authors should briefly mention that the results do not change if dRT is used as dependent variable (as in Mathieu et al).
15. I would suggest the authors add a figure that compares the results of the current replication experiments with those in the original paper. It will help the reader to compare the two studies.
16. In the file that describe the JASP files (“JASP file variables explained.docx”), the authors should add a brief description of what each JASP file contains. For example, “E2 RT 3-way n=62 no R2L.1.jasp” is the file for experiment 2, with only English speaking participants, with not log-transformed RTs, etc.

Author's Response

Dear Dr. Soltanlou,

Thank you for the opportunity to revise and resubmit our potential contribution to the replication issue of JNC. The reviewers provided valuable feedback and advice and we thank them for sharing their knowledge and experience with us. In the following, we explain our response to each of the reviewers' comments.

Sincerely,
Jamie Campbell

Reviewer A:

Major issues
Comment 1: The biggest potential issue here is that I am not convinced that the authors really replicated the study of Mathieu. When reading carefully the published manuscript, it is written "Therefore, we varied the delay between the arithmetic sign and O2 in the present experiment, such that the SOA (stimulus onset asynchrony) was 150, 300, or 450 ms" (p232). I have no better solution than to ask the authors of the first manuscript how the task was displayed to clarify this issue.

Response: Jérôme Prado confirmed in a communication to Jamie Campbell that the delay between the operation sign and O2 was as depicted in their Figure 1, with ISIs of 150, 300 and 450. Therefore, the SOAs between operator onset and O2 in their Experiment 1 were 300, 450 or 600 ms. We explain this on p . 4 (see also Footnote 1).

Comment 2: The relevance of the study to the field and the theoretical necessity to replicate is not clear enough. I think that several theoretical aspects should be better introduced and discussed. When going back to the original paper, it is still not very clear to me what this paradigm is actually measuring. Mathieu and colleagues suggested that it is because the position of O2 induced a forced shift of attention to a congruent hemifield that the performance was facilitated by accelerating the solving procedure. However, they acknowledge in the Discussion section that " …more is right" and "less is left". In other words, participants might shift their attention to the left or right side of space not because they move along the MNL, but simply because the appearance of the addition or subtraction sign makes them anticipate that the result of the problem will be smaller or larger than the first operand". The operator would trigger a shift to the right (+) or to the left (-) and this will accelerate the detection of the O2 in congruent trials because the digit appears in the locus of attention of the subject. This fast detection would result in faster reaction times than when the O2 is displayed on the opposite side of the screen. In this scenario, the solving procedure time would not be modified. It is simply the time to detect the O2 that will be modified. I would suggest the author of this manuscript to comment on the limits of Mathieu's paradigm and not to focus all the paper on potential statistical flaws.

Response: The reviewer articulates this argument very clearly. We now state (p. 18): "As mentioned by Mathieu et al. (p. 234), the plus sign (+) cues that the answer will be more than O1 whereas the minus sign (-) cues that it will be less than O1. We suggest that this spatial encoding of operation (i.e., go up or down relative to O1) may be particularly likely if the operator sign is presented very briefly in the context of mixed-operation trials. With this operation encoding strategy, the addition operator shifts attention rightward, facilitating encoding of O2 when it appears on the right relative to the left side. In this view, the effect is not intrinsic to the addition process per se, but instead resides in speeding up O2 encoding time. In theory, this spatial encoding effect would be reversed for subtraction, although our experiments did not find an advantage for subtraction with O2 displaced leftward."

Comment 3: This article aims at disentangling whether small problems are retrieved from memory (e.g., Campbell & Xue, 2001) or would involve fast calculation procedures (e.g., Fayol & Thevenot, 2012). This fast strategy, according to Mathieu, would be to shift attention along a mental number line. The authors of the present paper should consider that most of the work that was published on spatial associations with mental arithmetic concerns problems with multidigit numbers that cannot be retrieved from memory. The distinction should be carefully made all along the manuscript.

Response: We added the following on p. 4: "...most, studies of an association between space and arithmetic have used multi-digit problems (78 + 6, 78 - 6), whereas Mathieu et al. tested single-digit items (e.g., 8 + 6, 8 - 6). Solving multi-digit problems, however, includes solving one or more single-digit component problems; therefore, the mechanism of spatial associations for single-digit problems would also contribute to corresponding effects in multi-digit problems. This accentuates the importance of pursing spatial effects for single-digit problems."

Comment 4: The authors should also consider that, as far as I know, all the studies suggesting that spatial attentional shifts are functionally related to the solving procedure of additions and subtractions by investigating the effect of forced spatial attentional shifts concern multi-digit arithmetic problems. This was observed with optokinetic stimulation (Blini et al., 2019; Masson et al., 2016), interference paradigm (Masson & Pesenti, 2015), (in)congruent movements of the body (Wiemers et al., 2014) and with left and right neglect patients (Dormal et al., 2014; Masson et al., 2018). All the studies focusing on small arithmetical problems merely showed that these simple problems could trigger spatial biases (e.g., Marghetis et al., 2014; Masson & Pesenti, 2014; Pinhas & Fischer, 2008). No studies actually showed that these shifts were important to solve this kind of problem or to accelerate the solving procedure. This is true, except for the paper of Mathieu and colleagues, which uses a paradigm that is difficult to interpret (see above). I suggest the authors to clearly state whether their take-home message (i.e., unreliable SNAs for subtraction with Mathieu's paradigm) applies to arithmetic facts or all types of additions and subtractions.

Response: As we mentioned previously, solving a multi-digit addition or subtraction involves solving one or more single-digit component problems. Therefore, spatial effects

for single-digit problems should also occur with multi-digit problems perhaps superimposed on other spatial effects caused by the operator itself. Nonetheless, we substantially revised the last paragraph of the General Discussion with Comment 4 in mind.

Comment 5: I wonder why the authors tested some participants that are not native English speakers. The language in which arithmetic was learned is more critical than the first language of the participant for arithmetic performance (e.g., van Rinsveld et al., 2015). It is likely that the non-native participants were impaired by the use of English. And this could have added some noise in the data and masked some small but real effects. I noticed that the authors run complementary analysis without those participants but I wonder whether the authors should simply not include them in the sample.

Response: As there was no evidence in the literature that we are aware of that participants' native language affects spatial processing for arithmetic, there was no theoretically valid reason to exclude them. But, for Experiment 2, given the relatively large number of right-to-left readers, we provided a supplementary analysis that excluded right-to-left readers that showed no differences in the results with respect to the spatial manipulation compared to the full sample.

Minor Issues:
Page 3: "Since then, many experiments using a wide range of (other? Different than the SNARC ?) paradigms have confirmed a link between magnitude and space". Most articles that are cited here are actually the same SNARC paradigm. The authors should perhaps mention other behavioral measures of SNAs (Stoianov et al., 2008 in Cognition) or methods such as eye tracker (e.g., Myachykov et al., 2015 in Acta Psychologica; Salvaggio et al., 2018 in JEP:LMC) and neuropsychology of neglect patients (e.g., Vuilleumier et al., 2004 in Cortex).

Response: We have added additional references, but we hasten to note that this is not intended to be a review article.

Page 3: "spatial effects have also been observed in connection with performance of addition and subtraction". Some of the papers that are cited here do not concern directly spatial effects. Mathieu et al., 2018 (which is cited twice at the same place) showed an activation of a region related to spatial attention when showing a "+" sign. However, they did not measure spatial attentional shifts. McCrink et al., (2007) reported an over-estimation bias for additions and an under-estimation bias for subtractions (Operational Momentum effect) which has nothing to do, per se, with a spatial shift. There are plenty of other studies that really measured spatial shifts of attention or spatial biases that could be cited here instead of those two.

Response: We replaced the word "spatial" with "directional".

Page 4: Mathieu was not the first to show evidence of a spatial shift acting on arithmetic performance. Wiemers, Lindemann & Bekkering (2014), in QJEP, already showed that

when participants were moving their arms vertically or horizontally, it affected the performance of addition a subtraction solving. Moreover, as stated above, it is not yet clear whether it's the attentional shift induced by O2 position that facilitates the performance or the attentional shift induced by the operator that facilitates the detection of O2.

Response: We added this reference and corrected the text on p. 4.

Page 16: It is not surprising that the case study of a right neglect patient (Masson, Pesenti, Coyette, Andres, & Dormal, 2017) did not reveal a deficit for subtraction. Indeed, the right neglect patient with left-brain lesion showed more difficulty for addition solving than for subtraction solving, which is coherent with is incapacity to orient attention towards the right side of space. It is when left neglect patients were tested that a deficit for subtraction was observed (Dormal et al., 2014).

Response: We removed the reference to Masson et al. (2017).

Page 17: "… a generalized leftward association for subtraction, assuming it exists". I find that the "assuming it exists" part of the sentence unnecessary. There is now substantial evidence from different research teams that subtraction can cause leftward shifts (e.g., Liu et al., 2017; Masson & Pesenti, 2014; Marghetis et al., 2014) and that these shifts are functionally related to subtraction solving, at least of multidigit numbers (e.g., Dormal et al., 2014). This sentence is misleading and cast doubts on many studies that actually have nothing to do with this paradigm and this range of arithmetic problems. Please rephrase.

Response: We deleted “assuming it exists".

Reviewer B:

1. The authors suggest that the effects are due to the use of the more-or-less heuristic (e.g., p. 6). This heuristic can explain the results for addition problems (i.e., RTs advantage when the second operand is presented on the right), but does not explain the results with subtraction (i.e., no effect in experiment 1, and RTs advantage when the second operand is presented on the right in experiment 2). I think that neither the attentional shift on the MNL nor the more-or-less heuristic can fully explain the results. This should be mentioned in the manuscript.

Response: We revised the last paragraph accordingly.

2. T-test analyses. When describing the results reported in Mathieu et al. (2016), I think that the authors of the current manuscript reported two-tailed p-values for the t-test (e.g., page 4). Mathieu et al (2016) reported one-tailed p-values. This difference should be described in the manuscript.

Response: We now note (p. 11) that Matheiu et al. reported one-tailed t-tests for spatial effects, which is questionable though given that this was the first test of their new paradigm. Nonetheless, in their Experiment 1 these tests were nominally significant using a two-tailed criterion. We report p-values for two-tailed tests throughout.

3. Eta-squared for t-tests. Cohen's dz is a more common effect size measure for the t-test. Why did the authors choose to report the eta-squared for the t-tests?

Response: We used eta squared throughout for consistency with the ANOVAs reported.

4. Rounding p-value. The number of decimal values used for the p-values is not consistent. For example, "p = .002" and "p = .01" (page 4, lines 13 and 15) use 3 and 2 decimal values. However, I think that the second p-value (p = .01) is "p = .006" if rounded to 3 decimal values. Please, use the same number of decimal values for all the p-values.

Response. We reported p to two places (p = .01) unless the first two were both 0 then we report 3 (p = .001).

5. Page 5, line 11. In the analysis "[-12 ms, t(21) = 0.79, , SE = 15.19, ?^2 = .03, BF01 = 3.40]" the p-value (p = 0.438) is not reported.

We fixed this.

6. Page 6, line 5. "After presentation of O1 for 1 second the operation […]" should be "After presentation of O1 for 500 ms the operation […]".

Thank you for catching that.

7. Power. I did not understand how power and sample size were calculated. In the sentence "The power is .9 for a sample of 72 or more to detect an operation-specific O2 position effect about half the size observed by Mathieu et al. Experiment 1 (eta_p^2 = .24)", I did not understand how the partial eta-squared (.24) was calculated.

Response: The value was derived from the F-test information provided in Mathieu p. 233 for the main effect of operation on dRT in the 300 ms condition. The F(1, 33) = 11.16 converts to partial eta sqr = .2526 and we should have used this instead of the approximate value of .24 to calculate the apriori power. The MorePower calculator indicates that n = 72 yields power of .9 to detect a main effect of eta_p^2 >= .2526 / 2 = .13. We clarified this on p. 7.

8. Page 7. I think that the mean age is 20.8 rather than 21.1. Please also report the standard deviation for age.

Response: We did this.

9. List of problems. A table showing the list of problems (page 7, last paragraph) will make easier for the reader to quickly find the problems used in the experiments.

Response: We think the current presentation is an efficient and accessible explanation of the problem sets (p. 7) rather than add another table.

10. RTs log-transformed. I did not find a supplementary JASP file with the log-transformed RT for experiment 1. There is only a file for experiment 2 ("E2 log RT 4-way n=74.1.jasp").

Response: It is included now.

11. Outliers. The authors excluded from the analysis RTs that "were more than 4 SD from a participant's mean RT in each Operation × O2 Position × Mixture cell". In the original paper, Mathieu et al excluded "RT greater than 2 SDs from the mean for each participant". This difference should be reported in the manuscript. Also, are the results different if the Mathieu et al outlier procedure is used?

Response: In our procedure the experimenter flagged spoiled-RT trials on which the microphone did not pickup the onset of the participants' vocal response. This eliminates many potential outliers. On top of this we used a 4 SD outlier rule that in combination with spoiled RTs trimmed a similar proportion of RTs as Mathieu at et. Experiment 1 (4.7%, p. 232).

12. Page 10. The 2 t-tests for the comparisons between "slower subtraction in mixed-operation blocks relative to pure-operation blocks (854 vs 822 ms) whereas mean addition RT did not differ between single-operation (987 ms) and mixed-operations blocks (989 ms)" are not reported.

Response: We report these in the revision (p. 11)

13. The authors should choose the term "pure-operation" or "single-operation" (e.g., page 10) and use it consistently.

Response: We use pure-operation consistently now.

14. Page 12 and 15. After reporting the results for the RTs analyses, the authors should briefly mention that the results do not change if dRT is used as dependent variable (as in Mathieu et al).

Response: We do not think this is necessary as it is bound to be true. The F-value for the Operation x O2 Position interaction in the analysis of absolute RT is identical to the F-value for the main effect of operation in the analysis of dRT.

15. I would suggest the authors add a figure that compares the results of the current replication experiments with those in the original paper. It will help the reader to compare the two studies.

Response: We think the comparisons with the Mathieu et al. results provided in each RT section provide a straightforward comparison of their and our results.

16. In the file that describe the JASP files ("JASP file variables explained.docx"), the authors should add a brief description of what each JASP file contains. For example, "E2 RT 3-way n=62 no R2L.1.jasp" is the file for experiment 2, with only English speaking participants, with not log-transformed RTs, etc.

Second Decision Letter

Dear Jamie,

Thank you for your careful revision. Your article entitled "Not toeing the number line for simple arithmetic: Two large-n conceptual replications of Mathieu et al. (Cognition, 2016, Experiment 1)" has now been accepted for publication in the Journal of Numerical Cognition (JNC), the special issue on Direct and Conceptual Replication in Numerical Cognition – congratulations!

Kind regards,
Mojtaba

Reviewer A:

Comments for author(s):
I would like to thank the authors for taking into account my comments in this new version of the manuscript. In the context of this special issue on replicability, it was important to clarify the time course of a trial and to clarify the inconsistency between the text and the figure in the study of Mathieu. I have no further comments and thus recommend this manuscript for publication.

Reviewer B:

Comments for author(s):
The authors made an excellent job. The experiments are well designed and the manuscript is well written. I have no other comments. My recommendation is to accept the manuscript. I just found the following typo:
Page 10. In “This accentuates the importance of pursing spatial effects for single-digit problems”, “pursuing” is misspelt.