



Trying to discredit the Duplo task with a partial replication: Reply to Paulus and Kammermeier (2018)



Paula Rubio-Fernández^{a,b}

^a Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, USA

^b Department of Philosophy, University of Oslo, Norway

ARTICLE INFO

Keywords:

Failed vs. partial replication
Theoretical background
Conceptual replication
Task manipulations
Double standards
Replication crisis

ABSTRACT

Kammermeier and Paulus (2018) report a partial replication of the results of Rubio-Fernández and Geurts (2013) but present their study as a failed replication. Paulus and Kammermeier (2018) insist on a negative interpretation of their findings, discrediting the Duplo task against their own empirical evidence. Here I argue that Paulus and Kammermeier may try to make an impactful contribution to the field by adding to the growing skepticism towards early Theory of Mind studies, but fail to make any significant contribution to our understanding of young children's Theory of Mind abilities.

Rubio-Fernández and Geurts (2013) reported two significant patterns of results with the Duplo task (a modified version of the Sally-Anne task): 3-year-olds performed above chance in the Duplo task and significantly better than in a standard false-belief task. Kammermeier and Paulus (2018) failed to replicate the first pattern of results with 3-year-olds (who were at chance in the Duplo task) and observed above-chance performance only with 4- and 5-year-olds. However, they successfully replicated the second pattern of results and extended it in two ways: unlike Rubio-Fernández and Geurts (2013), who used a single standard false-belief task as a control (the Smarties task), Kammermeier and Paulus (2018) used two standard false-belief tasks as controls (an unexpected-contents task and a change-of-location task), and both their 3-year-olds and their 4-year-olds performed significantly better in the Duplo task than in either of the two control tasks (where they performed below-chance and at-chance, respectively).

In view of these results, Rubio-Fernández (2018) argued that Kammermeier and Paulus (2018) was a partial replication of Rubio-Fernández and Geurts (2013), rather than a failed replication, as the authors had argued. Far from acknowledging their mixed findings, Kammermeier and Paulus did not even mention in their Abstract the pattern of results that they had successfully replicated and extended (in contrast to other authors of failed replications of Theory of Mind studies; see, e.g., Kulke, Von Duhn, Schneider, & Rakoczy, 2018; Powell, Hobbs, Bardis, Carey, & Saxe, 2018). In their response, entitled 'How to deal with a failed replication of the Duplo task?', Paulus and Kammermeier (2018) continue to fail to acknowledge their partial replication of the results of Rubio-Fernández and Geurts (2013), discrediting the Duplo task against their own empirical evidence.

In this rebuttal, I will briefly respond to Paulus and Kammermeier (2018) before closing with some reflections on the value of failed replications and the devaluation of the original studies.

Reply to Paulus and Kammermeier (2018)

Kammermeier and Paulus (2018) did not explain the theoretical motivation behind the Duplo task designed by Rubio-Fernández and Geurts (2013), despite their attempt to directly replicate their results. Paulus and Kammermeier (2018) explain this choice on the grounds that "a detailed theoretical argumentation would result in a much longer piece". First of all, summarizing Rubio-Fernández and Geurts's theoretical position can be done in less than a page (as my Letter to the Editor illustrates; see also Rubio-Fernández, 2017), and since the special issue where Kammermeier and Paulus (2018) published their study did not have a word limit, their

<https://doi.org/10.1016/j.cogdev.2018.07.006>

Received 29 June 2018; Received in revised form 17 July 2018; Accepted 18 July 2018

Available online 12 November 2018

0885-2014/ © 2018 Elsevier Inc. All rights reserved.

explanation for failing to give any theoretical context to the Duplo task is unsatisfactory.

More importantly, Kammermeier and Paulus (2018) mischaracterized Rubio-Fernández and Geurts's interpretation of their own findings, falsely claiming that they had used them as strong evidence for early false-belief understanding. As I explained in my Letter to the Editor, Bart Geurts and I have never interpreted our results as unequivocal evidence of false-belief understanding in 3-year-olds and have always been open about possible low-level explanations of our own findings. However, Kammermeier and Paulus may have attributed us a much stronger position than we have ever defended because that makes their failed replication more impactful: after all, if the authors of the original study already suggest low-level interpretations of their own findings, what are the implications of a failed replication? Be that as it may, Kammermeier and Paulus (2018) should have ensured a proper characterization of our theoretical position before arguing against it.

Paulus and Kammermeier (2018) deny that there might be an expectancy bias in their work and argue that they genuinely tried to replicate Rubio-Fernández and Geurts's (2013) findings. If that was the case, it is remarkable that these authors did not contact us to validate their protocol (as other authors who have attempted to replicate our findings did). There may of course be differences across developmental labs, but it seems strange that a group may observe unexpected results with 191 children without going back to the drawing board to discuss (with the original authors) if there may be something wrong with their protocol.

Paulus and Kammermeier also reject that a recent study by Bialecka-Pikul, Kosno, Białek, and Szpak (2018a, 2018b) should be considered as a successful, large-scale replication with the Duplo task. While it is true that Bialecka-Pikul et al. introduced their own manipulations to the original Duplo task, they designed and report their study as a conceptual replication of the studies by Rubio-Fernández and Geurts (2013, 2016). More specifically, Bialecka-Pikul et al. (2018a, 2018b) predicted that a more interactive version of the Duplo task should reveal even better performance than the original task. In their third experiment, Rubio-Fernández and Geurts (2013) observed that asking the standard false-belief question at the end of the Duplo task resulted in below-chance performance (unlike the original design, which used an open test question with above-chance results). Bialecka-Pikul et al. raised to this challenge and engaged their 3.5-year-olds in the deception of their protagonist with above-chance results, despite their using the standard test question. It must be noted that other than for the Experimenter's inviting the child to trick the protagonist, the design of Bialecka-Pikul et al.'s task was comparable to the Duplo task used by Rubio-Fernández and Geurts's (2013; Experiment 2c).

Counting Bialecka-Pikul et al.'s conceptual replication, there have been seven different versions of the Duplo task and 3-year-olds have been able to pass only three of them, while all seven were reported. It must be noted that all versions of the Duplo task were theoretically motivated and included single manipulations to the original paradigm. Because of these individual manipulations, the results (both successes and failures) make for a coherent line of research. Kammermeier and Paulus (2018), however, did not discuss any of the manipulations that have been used with the Duplo task or their theoretical implications. Instead, Paulus and Kammermeier (2018) warn the readership against the "difficulty to evaluate which aspects of procedural adaptations are relevant and (...) the danger of post-hoc speculations based on specific desired result patterns". Given that these authors have not even mentioned in their papers what procedural adaptations have been used with the Duplo task, their criticism may try to discredit our work but adds no new insights to the methodological debate on modified false-belief tasks.

The last two points in Paulus and Kammermeier's response are red herrings: first, they defend why Kammermeier and Paulus (2018) said that they had been in close contact with me (as the first author of the Duplo studies) and try to make a case for how our sparse email exchange could nonetheless count as 'close contact'. However, in my Letter to the Editor, I did not challenge Kammermeier and Paulus (2018) for saying that we were in close email contact (whatever that may mean), but for repeatedly claiming that they had been in close contact with me to "ensure an exact replication protocol". As I said before, Kammermeier and Paulus never shared with us any footage, description or details of their protocol and therefore, their claims that they had ensured an exact replication procedure are simply false (as the authors themselves acknowledge).

Finally, Paulus and Kammermeier (2018) try to explain why Kammermeier and Paulus (2018) may have been 'ambivalent' about their interpretation of the Duplo task. First, if these authors believe that 3-year-olds may not understand the Duplo task and perform randomly (as Paulus and Kammermeier, 2018 argue), then it is unclear why they were trying to replicate the original results to begin with. Second, in my Letter to the Editor, I did not point out a mere 'ambivalence' in their interpretation of the Duplo task but a clear double standard: when Kammermeier and Paulus (2018) motivate their replication study, they argue that if young children understand false beliefs, they should pass the Duplo task; and indeed, they then question 3-year-olds' Theory of Mind abilities in view of their chance performance in that task. However, when they later discuss 3- and 4-year-olds' better performance in the Duplo task relative to two standard false-belief tasks, they dismiss the Duplo task as a reliable test of false-belief understanding. Basically, the Duplo task is reliable only when 3-year-olds fail it: when 3- and 4-year-olds show better performance than in two standard false-belief tasks, the Duplo task becomes uninterpretable. This is not a mere ambivalence but a double standard, which once again discredits the Duplo task but does not further our understanding of early social cognition.

Closing remarks: learning from other literatures

While failed replications are important for the advancement of any science, they can also be detrimental when they lower the research and publication standards of a field. As part of my work on Theory of Mind and bilingualism (Rubio-Fernández, 2017), I have closely followed the replication crisis in the area of bilingualism and Executive Control. After a number of failed replications were published (e.g., Hilchey & Klein, 2011; Paap & Greenberg, 2013), while original studies with positive results continued to pile up (Grundy, Chung-Fat-Yim, Friesen, Mak, & Bialystok, 2017; Poarch & Bialystok, 2015), the field went into a turmoil trying to decide whether there is such a thing as a 'bilingual advantage' in Executive Control (for a taste of the bitter debate, see de Bruin, Treccani, & Della Sala, 2015: *Cognitive advantage in bilingualism: An example of publication bias?*; Bialystok, Kroll, Green, MacWhinney, & Craik,

2015). *Publication bias and the validity of evidence: What's the connection?*; for a review, see Antoniou, 2018).

The lowest point in this debate came when studies started to publish null results with tasks that had not been originally designed to test for a bilingual advantage, and simply required the use of some Executive Function (e.g., visuospatial perspective taking during conversation). This was disheartening because nobody in the field had ever claimed that bilingualism should have a boosting effect on all components of Executive Function, but only on those related to the language switching experience (and even identifying those is challenging; see Bialystok, 2015, 2017). Therefore, the growing expectation that if a bilingual advantage existed at all, it should show on any task that required Executive Control was simply unfounded. However, it was on this false premise that null results were being published and interpreted as a challenge to the mainstream view.

Paulus and Kammermeier's (2018) insistence on interpreting the results of Kammermeier and Paulus (2018) as a failed replication with the Duplo task (rather than a partial replication), their mischaracterization of our own interpretation of 3-year-olds' success in the Duplo task as strong evidence of false-belief understanding, and their falsely claiming that they had ensured an exact replication protocol while questioning the reliability of the task as a Theory of Mind test (only when it showed better performance than standard false-belief tasks) reveals a determination to make an impactful contribution to the recent literature on failed replications and add to the growing skepticism towards early Theory of Mind. One may wonder, however, what kind of contribution that makes to the field. Paulus and Kammermeier say that "it is important to learn about failed replications" and while that may be true, it surely is not a matter of spreading the word as one would spread chaos. A more constructive approach is, in my opinion, to try to learn *from* failed replications by running careful studies with highly controlled protocols. This has been in fact our approach since we first designed the Duplo task and its variations. Of the six versions of the task that we designed, 3-year-olds were only able to pass two of them. However, we found their failures as interesting and informative as their successes precisely because they revealed the limits of 3-year-olds' capacity to pass the Duplo task.

Acknowledgements

This research was supported by a Young Research Talent Grant from the Research Council of Norway (230718).

References

- Antoniou, M. (2018). The advantages of bilingualism debate. *Annual Review of Linguistics* (in press).
- Bialecka-Pikul, M., Kosno, M., Bialek, M., & Szpak, M. (2018a). Let's do it together! The role of interaction in false belief understanding. *Journal of Experimental Child Psychology* (in press).
- Bialecka-Pikul, M., Kosno, M., Bialek, M., & Szpak, M. (2018b). Let's do it together! The role of interaction in false belief understanding. *Poster presented at the 2018 Budapest CEU Conference on Cognitive Development (BCCCD18)*.
- Bialystok, E. (2015). Bilingualism and the development of executive function: The role of attention. *Child Development Perspectives*, 9(2), 117–121.
- Bialystok, E. (2017). The bilingual adaptation: How minds accommodate experience. *Psychological Bulletin*, 143(3), 233–262.
- Bialystok, E., Kroll, J. F., Green, D. W., MacWhinney, B., & Craik, F. I. (2015). Publication bias and the validity of evidence: What's the connection? *Psychological Science*, 26(6), 944–946.
- de Bruin, A., Treccani, B., & Della Sala, S. (2015). Cognitive advantage in bilingualism: An example of publication bias? *Psychological Science*, 26(1), 99–107.
- Grundy, J. G., Chung-Fat-Yim, A., Friesen, D. C., Mak, L., & Bialystok, E. (2017). Sequential congruency effects reveal differences in disengagement of attention for monolingual and bilingual young adults. *Cognition*, 163, 42–55.
- Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on nonlinguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review*, 18(4), 625–658.
- Kammermeier, M., & Paulus, M. (2018). Do action-based tasks evidence false-belief understanding in young children? *Cognitive Development* (in press).
- Kulke, L., Von Duhn, B., Schneider, D., & Rakoczy, H. (2018). Is implicit Theory of Mind a real and robust phenomenon? Results from a systematic replication study. *Psychological Science* (in press).
- Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66(2), 232–258.
- Paulus, M., & Kammermeier, M. (2018). How to deal with a failed replication of the Duplot task? A response to Rubio-Fernández (2018). *Cognitive Development* (in press).
- Poarch, G. J., & Bialystok, E. (2015). Bilingualism as a model for multitasking. *Developmental Review*, 35, 113–124.
- Powell, L. J., Hobbs, K., Bardis, A., Carey, S., & Saxe, R. (2018). Replications of implicit theory of mind tasks with varying representational demands. *Cognitive Development* (in press).
- Rubio-Fernández, P. (2018). What do failed (and successful) replications with the Duplo task show? (in press) *Cognitive Development*.
- Rubio-Fernández, P. (2017). Why are bilinguals better than monolinguals at false-belief tasks? *Psychonomic Bulletin & Review*, 24(3), 987–998.
- Rubio-Fernández, P., & Geurts, B. (2013). How to pass the false-belief task before your fourth birthday. *Psychological Science*, 24(1), 27–33.
- Rubio-Fernández, P., & Geurts, B. (2016). Don't mention the marble! The role of attentional processes in false-belief tasks. *Review of Philosophy and Psychology*, 7(4), 835–850.