Comment

What Do These Findings Tell Us? Comment on Tinella et al. Cognitive Efficiency and Fitness-to-Drive along the Lifespan: The Mediation Effect of Visuospatial Transformations. *Brain Sci.* 2021, *11*, 1028

Robert E. Kelly, Jr. [1,*] , Anthony O. Ahmed [1] and Matthew J. Hoptman [2,3]

1   Department of Psychiatry, Weill Cornell Medicine, White Plains, NY 10605, USA; aoa9001@med.cornell.edu
2   Clinical Research Division, Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY 10962, USA; matthew.hoptman@nki.rfmh.org or matthew.hoptman@nyumc.org
3   Department of Psychiatry, New York University Grossman School of Medicine, New York, NY 10016, USA
*   Correspondence: rek2005@med.cornell.edu

Tinella et al.'s recent article [1] seems like a natural extension of their previous work [2], wherein the authors provided evidence to "suggest the specific contribution of spatial mental transformation skills in the execution of complex behaviors connected to the fitness to drive". Their more recent paper extends this work by performing a path analysis to explore mediating pathways that can, in part, explain some of the variance in the prediction of driving skill scores. However, additional information would facilitate our understanding of this paper's contribution to science, specifically by addressing the questions "Do these results seem replicable?" and "How do these results advance our understanding of brain function and/or human behavior?" The purpose of this comment is to call attention to these important questions, which often remain unanswered in published scientific papers, and to provide an opportunity for the authors to answer these questions more fully.

In recent years, growing concerns have been raised in the scientific community about the high rates of non-replication of research results across a wide variety of scientific endeavors—much higher than the conventional $\alpha = 0.05$ used for the statistical testing of research hypotheses [3–9]. These high rates of non-replication may, in part, be expected from Bayesian statistics [10], but other major contributing factors include methodological errors [11], poorly powered studies [8,12], and the reporting bias that results from academic pressures to publish in combination with a publication bias favoring positive, "statistically significant" findings [9,13]. This reporting bias inflates effective *p*-values because no correction is made for unreported alternative hypotheses that were tested. The problem is further exacerbated by practices such as hypothesizing after the results are known (HARKing) [14], adjusting parameters of the data-processing pipeline and statistical analyses until the sought-after, statistically significant results are "found", also known as *p*-hacking [10,15–18], and other related practices [19]. Even the data analyses for uncomplicated studies can thus be manipulated, for example, by adjusting the inclusion/exclusion criteria for participants in a study. To minimize these effects, authors should include sufficient methodological detail in their papers to allow readers to properly evaluate the relevance of the chosen methods [17,18].

For example, the two articles by Tinella et al. lack important details concerning the selection of study participants, leading to puzzling discrepancies in the participant selection criteria for the two studies, which were by the same authors, studying the same variables, in the same location. The earlier study included 120 males and 63 females, ages 18–64, whereas the more recent study included only 117 males, ages 18–64, and added 58 males, ages 65–91. So, why were 3 of the males and all 63 females excluded for the recent study? Why did the authors choose to add males only in the 65–91 age group? Why would these

studies' inclusion/exclusion criteria differ at all? Data collection is costly, so it would be of interest for readers to know what prompted the authors to discard much of the data for the recent study.

Moreover, regression-based analytic methods such as mediational models, path analysis, factor analysis, and structural equation models potentially capitalize on chance. The regression weights or path coefficients that depict the magnitude of associations of arrows in path diagrams are maximized for the study sample, in keeping with the least-squares criterion. This is a long-recognized challenge in regression and a source of criticism for path-analytic models [20,21]. A common solution is to designate and fit the original hypothesized model in "training" data and then attempt a replication in an independent cross-validation dataset [22,23]. The degree of replicability could then be inferred by examining the precision efficacy, proportional shrinkage, and/or prior predictive *p*-values [24].

The meaning of the derived study findings also merits explanation. For example, how does the finding that "mental rotation" (MRT) partially "mediates" the relationship between "global cognitive functioning" (Montreal Cognitive Assessment, MoCA) and "resilience of attention" (Determination Test, DT) contribute to our understanding of brain function or human behavior? This finding itself is not surprising: the MoCA includes some visuospatial testing and all of the studied variables correlate with each other to some degree. Sometimes there is a value in confirming expectations, but, in this case, what does the finding tell us in terms of latent constructs reflecting brain function? In particular, what is meant by the construct "global cognitive functioning", which the authors describe as corresponding to the observable MoCA scores? The MoCA was developed as a measure of cognitive impairment in elderly persons suspected of developing dementia or mild cognitive impairment and was validated on a sample of three groups of elderly adults: Alzheimer's disease, mild cognitive impairment, or no cognitive impairment [25]. Although the MoCA serves well to predict cognitive impairment and to identify persons having Alzheimer's dementia or mild cognitive impairment [26], its meaning becomes unclear when applied to a sample of cognitively intact individuals.

To illustrate the point, consider the following analogy. Suppose we were to use student height to predict student age among students in secondary school, grades 1–12. If we were to limit our study to include equal numbers of students from grades 1, 5, and 12, we would surely find that height is an excellent predictor of age and, further, that height is also an excellent predictor of age for virtually any large sample of secondary school students randomly drawn from grades 1–12. However, height would be a terrible predictor of age for samples randomly drawn from grade 12 only (Figure 1). In the same way, we should not anticipate that MoCA score will reflect "cognitive impairment" in healthy controls from ages 18–91 as well as it did in studies validating the MoCA as a useful measure of cognitive impairment. Compared with cognitively impaired people, we can expect that differences in MoCA scores among healthy controls would reflect confounding variables in the prediction of cognitive impairment, such as education level [27] and intelligence [28], to a greater degree. Many other variables could be considered, such as "physical exercise" or "mental exercise". It is not clear what is being measured when the MoCA is applied to healthy adults, so what does MoCA score tell us about brain function in the current study?

Finally, the question concerning what this study tells us about human behavior also merits consideration. The MoCA is a natural subject of study, given its use in real-world applications concerning fitness-to-drive among individuals thought to be cognitively impaired due to dementia. Having moderate to severe dementia is considered evidence of unfitness-to-drive [29], so the MoCA score, weighed together with other relevant clinical information, can help to determine the diagnosis of dementia and fitness to drive. However, the relevant question here is "What does a study of cognitively intact men tell us about fitness to drive among cognitively impaired elderly adults?" Why not directly study cognitively impaired adults?
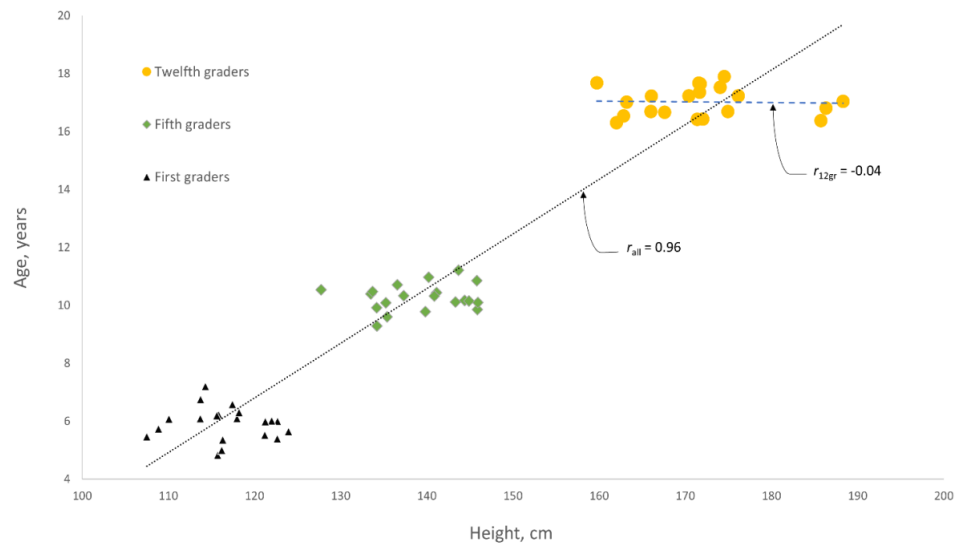
Figure 1. Scatterplot of student age vs. height, showing that height is an excellent predictor of age, by Pearson's *r,* for all students combined (r_{all}), but not for students in grade 12 only (r_{12gr}). This illustration depicts secondary school students from grades 1, 5, and 12 using simulated data. The dotted and dashed lines show the lines of best fit, using least squares linear regression, for all students and for students in grade 12 only, respectively.

In summary, important questions remain unanswered in the article in question, concerning participant inclusion/exclusion criteria, replicability of the path analysis, and implications concerning brain function and human behavior. Elucidating these issues might better enable readers to evaluate this study's contribution to advancing scientific knowledge.

Data Availability Statement: Data for the figure were generated using LibreOffice Calc version 5 for Linux to derive random pairs of values for height and age, based on the normal distribution. For first-graders, the mean and standard deviation for height (cm) were [115, 4], and for age (years) [6, 0.5]. The corresponding values for fifth-graders were height [139, 5] and age [10, 0.5]. For twelfth-graders, they were height [170, 9] and age [17, 0.5].

References

1. Tinella, L.; Lopez, A.; Caffò, A.O.; Nardulli, F.; Grattagliano, I.; Bosco, A. Cognitive Efficiency and Fitness-to-Drive along the Lifespan: The Mediation Effect of Visuospatial Transformations. *Brain Sci.* **2021**, *11*, 1028. [CrossRef] [PubMed]
2. Tinella, L.; Lopez, A.; Caffò, A.O.; Grattagliano, I.; Bosco, A. Spatial Mental Transformation Skills Discriminate Fitness to Drive in Young and Old Adults. *Front. Psychol.* **2020**, *11*, 1–17. [CrossRef] [PubMed]
3. Schooler, J.W. Metascience could rescue the "replication crisis". *Nature* **2014**, *515*, 9. [CrossRef] [PubMed]
4. Poldrack, R.A.; Baker, C.I.; Durnez, J.; Gorgolewski, K.J.; Matthews, P.M.; Munafò, M.R.; Nichols, T.E.; Poline, J.B.; Vul, E.; Yarkoni, T. Scanning the horizon: Towards transparent and reproducible neuroimaging research. *Nat. Rev. Neurosci.* **2017**, *18*, 115–126. [CrossRef]
5. Jager, L.R.; Leek, J.T. An estimate of the science-wise false discovery rate and application to the top medical literature. *Biostatistics* **2014**, *15*, 1–12. [CrossRef]
6. Ioannidis, J.P.A. Discussion: Why "An estimate of the science-wise false discovery rate and application to the top medical literature" is false. *Biostatistics* **2014**, *15*, 28–36. [CrossRef]

7. Ioannidis, J.P.A. Contradicted and initially stronger effects in highly cited clinical research. *J. Am. Med. Assoc.* **2005**, *294*, 218–228. [CrossRef]

8. Ioannidis, J.P.A. Why Most Published Research Findings Are False. *PLoS Med.* **2005**, *2*, e124. [CrossRef]

9. Nosek, B.A.; Lakens, D. Registered reports: A method to increase the credibility of published results. *Soc. Psychol.* **2014**, *45*, 137–141. [CrossRef]

10. Nuzzo, R. Scientific method: Statistical errors. *Nature* **2014**, *506*, 150–152. [CrossRef]

11. Vul, E.; Harris, C.; Winkielman, P.; Pashler, H. Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition. *Perspect. Psychol. Sci.* **2009**, *4*, 274–290. [CrossRef]

12. Button, K.S.; Ioannidis, J.P.A.; Mokrysz, C.; Nosek, B.A.; Flint, J.; Robinson, E.S.J.; Munafò, M.R. Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.* **2013**, *14*, 365–376. [CrossRef]

13. Gorgolewski, K.J.; Nichols, T.; Kennedy, D.N.; Poline, J.B.; Poldrack, R.A. Making replication prestigious. *Behav. Brain Sci.* **2018**, *41*, e131. [CrossRef]

14. Kerr, N.L. HARKing: Hypothesizing after the results are known. *Personal. Soc. Psychol. Rev.* **1998**, *2*, 196–217. [CrossRef]

15. Steegen, S.; Tuerlinckx, F.; Gelman, A.; Vanpaemel, W. Increasing Transparency through a Multiverse Analysis. *Perspect. Psychol. Sci.* **2016**, *11*, 702–712. [CrossRef]

16. Botvinik-Nezer, R.; Holzmeister, F.; Camerer, C.; Dreber, A.; Huber, J.; Johannesson, M.; Kirchler, M.; Iwanir, R.; Mumford, J.; Adcock, A.; et al. Variability in the analysis of a single neuroimaging dataset by many teams. *bioRxiv* **2019**. [CrossRef]

17. Head, M.L.; Holman, L.; Lanfear, R.; Kahn, A.T.; Jennions, M.D. The Extent and Consequences of P-Hacking in Science. *PLoS Biol.* **2015**, *13*, e1002106. [CrossRef]

18. Simmons, J.P.; Nelson, L.D.; Simonsohn, U. False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.* **2011**, *22*, 1359–1366. [CrossRef]

19. Andrade, C. HARKing, Cherry-Picking, P-Hacking, Fishing Expeditions, and Data Dredging and Mining as Questionable Research Practices. *J. Clin. Psychiatry* **2021**, *82*, 20f13804. [CrossRef]

20. Cliff, N. Some cautions concerning the application of causal modeling methods. *Multivariate Behav. Res.* **1983**, *18*, 115–126. [CrossRef]

21. Freedman, D.A. As Others See Us: A Case Study in Path Analysis. *J. Educ. Stat.* **1987**, *12*, 101–128. [CrossRef]

22. Ahmed, A.O.; Kirkpatrick, B.; Galderisi, S.; Mucci, A.; Rossi, A.; Bertolino, A.; Rocca, P.; Maj, M.; Kaiser, S.; Bischof, M.; et al. Cross-cultural Validation of the 5-Factor Structure of Negative Symptoms in Schizophrenia. *Schizophr. Bull.* **2019**, *45*, 305–314. [CrossRef]

23. Browne, M.W.; Cudeck, R. Alternative Ways of Assessing Model Fit. *Sociol. Methods Res.* **1992**, *21*, 230–258. [CrossRef]

24. Brooks, G.P.; Gordon, P. Precision Efficacy Analysis for Regression. In Proceedings of the Annual Meeting of the Mid-Western Educational Research Association, Chicago, IL, USA, 14–17 October 1998; pp. 1–67.

25. Nasreddine, Z.S.; Phillips, N.A.; Bédirian, V.; Charbonneau, S.; Whitehead, V.; Collin, I.; Cummings, J.L.; Chertkow, H. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **2005**, *53*, 695–699. [CrossRef]

26. Roalf, D.R.; Moberg, P.J.; Xie, S.X.; Wolk, D.A.; Moelter, S.T.; Arnold, S.E. Comparative accuracies of two common screening instruments for classification of Alzheimer's disease, mild cognitive impairment, and healthy aging. *Alzheimers Dement.* **2013**, *9*, 529–537. [CrossRef]

27. Kim, J.I.; Sunwoo, M.K.; Sohn, Y.H.; Lee, P.H.; Hong, J.Y. The MMSE and MoCA for Screening Cognitive Impairment in Less Educated Patients with Parkinson's Disease. *J. Mov. Disord.* **2016**, *9*, 152–159. [CrossRef]

28. Sugarman, M.A.; Axelrod, B.N. Utility of the Montreal Cognitive Assessment and Mini-Mental State Examination in predicting general intellectual abilities. *Cogn. Behav. Neurol.* **2014**, *27*, 148–154. [CrossRef]

29. Johansson, K.; Lundberg, C. The 1994 International Consensus Conference on Dementia and Driving: A brief report. *Alzheimer Dis. Assoc. Disord.* **1997**, *11*, 62–69. [CrossRef]