

Commentary on “Do Musicians Have Better Short-Term Memory Than Nonmusicians? A Multilab Study”: Where Does This Study Guide Future Research Endeavors?



Tilo Strobach^{1,2} , Emma G. Cunningham³, Shanti Astra³ , and C. Shawn Green³ 

¹Department of Psychology, MSH Medical School Hamburg, Hamburg, Germany; ²ICAN Institute of Cognitive and Affective Neuroscience, MSH Medical School Hamburg, Hamburg, Germany; and

³Department of Psychology, University of Wisconsin-Madison, Madison, Wisconsin

Received 2/4/26; revision accepted 4/30/26

The study by Grassi et al. (2025) represents a strong contribution to the literature on musical expertise, brain plasticity, and research on cognitive training and transfer. By assembling an impressive multilab collaboration across 33 research units and 15 countries that together allowed for a sample size of more than 600 musicians and 600 nonmusicians, the authors collected data via a comprehensive set of span tasks to examine the extent to which musicians differ from nonmusicians in short-term memory performance. Their results convincingly established that such differences do exist, finding effect sizes that vary systematically across stimulus domains: large for musical materials, small for visuospatial materials, and very small for verbal materials.

Importantly, however, Grassi et al. (2025) also explicitly acknowledged that establishing the existence of a group difference between musicians and nonmusicians is only a first step. The deeper theoretical challenge lies in elucidating the cognitive mechanisms that give rise to this difference between groups. Addressing this challenge is nontrivial because it requires moving beyond simple between-groups comparisons toward methods and analyses that can discriminate among competing theoretical accounts of cognitive training and transfer.

Although issues related to causation are always at the fore when results use purely correlational methods, the types of approaches (e.g., long-term intervention studies) necessary to fully address such issues are not always feasible (for additional discussion, see Román-Caballero et al., 2025). Indeed, the criteria for being a member of the musician group in Grassi et al. (2025) included at

least 10 years of direct music instruction, which is infeasible to match experimentally. However, clever methodology and analysis techniques can often nonetheless permit deeper inferences about existing theories (Strobach & Karbach, 2021). One can easily imagine follow-up correlational work that could, with slight shifts in methodology, speak to any number of major theories in the field. Such theories include but are not limited to (a) the capacity-efficiency model of cognitive training and transfer (von Bastian et al., 2022), which attributes plasticity to either increased processing capacity or more efficient use of existing capacity; (b) the model by Lövdén et al. (2010), which conceptualizes plasticity as arising from sustained mismatches between organismic resources and environmental demands, driven by changes in processing efficiency and representations; (c) Taatgen’s (2013) primitive-elements theory, which distinguishes task-specific elements from task-general control skills and posits that transfer occurs when task-general skills overlap across tasks; and (d) the cognitive-routine framework (Gathercole et al., 2019), which proposes that training builds routines that transfer to only structurally similar tasks.

We argue that one particularly powerful and still underused approach to this problem is to focus on the temporal dynamics of performance on transfer tasks (Cochrane et al., 2024; Kattner et al., 2017). That is,

Corresponding Author:

Tilo Strobach, Department of Psychology, MSH Medical School Hamburg, Hamburg, Germany
Email: tilo.strobach@medicalschooll-hamburg.de



Creative Commons NonCommercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits noncommercial use, reproduction, and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Table 1. Performance Scores in the Musical-, Verbal-, and Visuospatial-Span Tasks Across Musicians and Nonmusicians as Averages and Separated Into Blocks 1 and 2

Task	Musicians			Nonmusicians		
	Block 1	Block 2	Average	Block 1	Block 2	Average
Musical span	27.55	28.68	28.12	22.48	20.97	21.72
Verbal span	10.01	10.49	10.25	9.77	10.06	9.91
Visuospatial span	8.87	9.25	9.062	8.43	8.52	8.48

rather than treating performance as a static outcome measured once per task, researchers can instead examine how performance unfolds over time both within a given task and (if participants have completed multiple tasks) across tasks. Analyses that explicitly consider time are uniquely positioned to reveal how groups differ, not merely that they differ, and, in turn, allow for richer inferences about the underlying nature of the group differences (Cochrane et al., 2023).

A dynamic perspective makes clear that group differences in cognitive performance can arise in multiple, theoretically distinct ways. At least four such pathways are particularly relevant in the context of musicianship and memory. Crucially, identifying these pathways necessitates examining performance through time (i.e., all of these patterns would give rise to a simple “global group advantage” when using a standard analysis approach in which a single performance value is calculated for each task and time/order is not considered). The first pathway has been referred to as “direct transfer” or “immediate transfer”—in which some type of previous experience (e.g., long-term musical training) allows individuals to excel from the very first trial of a new task (e.g., a music-based short-term memory task). The second pathway one might refer to as simple “learning to learn” (Bavelier et al., 2012; Zhang et al., 2021). In short, the musician and nonmusician groups may differ in how quickly they improve with repeated exposure to the same task because musicianship may improve the capability to learn more rapidly when approaching new tasks. Here, the critical signature of learning to learn is a steeper performance gain across blocks or trials within a task. The third pathway would then be referred to as “hierarchical learning to learn,” wherein groups differ in their ability to extract higher-order regularities from earlier tasks and apply that learning to entirely new tasks encountered later in an experiment (Kattner et al., 2017). Under this account, expert musicians might show enhanced performance specifically on tasks presented later (potentially irrespective of task content), reflecting superior metalearning or task-general adaptation. Finally, in the fourth pathway, groups may differ not in learning per se but in their susceptibility to performance decline

over time (Pessiglione et al., 2025). In this case, “superior performance” may emerge because one group maintains stable performance while another exhibits fatigue-related decrements (or failures of long-term sustained attention). In the case of group differences between musicians and nonmusicians, this effect may emerge as a result of musicians’ unique capacity to continue practicing and improving their musical capabilities consistently through fatigue or attentional strain.

To demonstrate how dynamic analyses can inform theory, we conducted a series of post hoc analyses on the three span tasks used in Grassi et al. (2025): musical span, verbal span, and visuospatial span. Each participant of the Grassi et al. study performed this set of span tasks. The order of the tasks was counterbalanced across participants in a Latin-square way, and each task comprised two blocks executed back-to-back. Although the tasks themselves are not fully suited for truly time-continuous analyses (because the tasks were adaptive in nature and thus time/trial number is confounded with task difficulty), the scale and overarching design do allow novel analyses that can speak more deeply to theory than those using a static approach (see preregistration at <https://osf.io/kjm9d/overview>). The relevant descriptives are presented in Table 1.

Learning Across Tasks: Testing Hierarchical Learning to Learn

The order of the three span tasks (musical, verbal, visuospatial) was randomized across participants in a Latin-square way. This allowed us to examine whether task order influenced performance and whether this effect differed between musicians and nonmusicians as an empirical test of hierarchical learning to learn across tasks. In its most elementary form, this would predict that the size of the musician advantage should grow from task to task independent of the tasks themselves (e.g., there should be a larger verbal-span advantage for musicians when it was the third of the three tasks than when it was the first of the three tasks because musicians were better able to apply learning from earlier tasks than nonmusicians). We conducted separate hierarchical

Table 2. Effects of Task Order for Musicians and Nonmusicians in the Musical-, Verbal-, and Visuospatial-Span Tasks

Task	Variable	Regression coefficient	<i>df</i>	<i>t</i> statistic	<i>p</i> value
Musical span	Order	−0.26	1,354.00	−0.98	.326
	Group × Order	0.50	1,354.00	1.35	.179
Verbal span	Order	0.19	1,354.00	1.839	.066
	Group × Order	−0.02	1,354.00	−0.148	.882
Visuospatial span	Order	−0.00	1,354.00	−0.02	.982
	Group × Order	0.07	1,354.00	0.62	.538

linear regression models for each task for the interaction of task order and group (musicians and nonmusicians, with musicians as the reference group), controlling for within-tasks time point (or block) and any interaction of time point and group. We found no significant order effects or interactions, indicating that neither musicians nor nonmusicians improved on tasks encountered later in the study (Table 2). This pattern indicates that neither musicians nor nonmusicians showed systematic improvement on tasks encountered later in the study, providing little evidence that musicians disproportionately benefit from cross-task learning or task-general adaptation in this context.

Learning and Stability Within Tasks: Simple Learning to Learn and Fatigue Resistance

Within each task, participants performed two blocks of trials back to back. This allowed us to examine whether the musician advantage stayed the same or grew across blocks and thus empirically test both simple learning to learn and fatigue resistance. To test for time-dependent effects on task-performance differences between groups across the two time points (Block 1 and Block 2), we conducted separate hierarchical linear regression models for each span task on the interaction between the group (musicians vs. nonmusicians) and the time point, controlling for the order the participant experienced the three span tasks, and any interaction between group and

order (Table 3). We found that musicians outperformed nonmusicians in the musical-span task, and this advantage was stronger at the second time point than the first, $b = -1.86$, $t(1,356) = -5.48$, $p < .001$. Importantly, the interactive effect between time point and group in this case was driven by a decrease in performance in the nonmusician group in Block 2 (1.902-unit decrease), $t(695) = 8.32$, $p < 0.001$, with relatively no change in the musician group across time points (0.027-unit increase), $t(668) = 0.108$, $p = .914$. In the verbal-span task, we found that both groups performed significantly better in Block 2; however, the interaction between time point and group was not significant. Finally, in the visuospatial-span task, both musicians and nonmusicians significantly improved their performance from Block 1 to Block 2, but the musicians improved more than the nonmusicians. The data are thus consistent with enhanced simple learning to learn in some contexts (visuospatial span) and greater resistance to performance decline in others (musical span).

Early Performance Levels: A Window Into Direct Transfer

To further clarify the nature of the short-term memory advantage in musicians, we examined group differences at the earliest possible time point (Block 1). We note that this is not an ideal test of direct (or immediate) transfer because Block 1 performance is already an aggregate measure of performance across multiple trials

Table 3. Effects of Within-Tasks Time Point for Musicians and Nonmusicians in the Musical-, Verbal-, and Visuospatial-Span Tasks

Task	Variable	Regression coefficient	<i>df</i>	<i>t</i> statistic	<i>p</i> value
Musical span	Time point	−0.02	1,356.00	−0.09	.926
	Group × Time Point	−1.86	1,356.00	−5.48	<.001
Verbal span	Time point	0.47	1,356.00	5.67	<.001
	Group × Time Point	−0.15	1,356.00	−1.34	.179
Visuospatial span	Time point	0.37	1,356.00	4.32	<.001
	Group × Time Point	−0.27	1,356.00	−2.21	.027

(i.e., over a number of trials over which we know that learning can occur; Cochrane & Green, 2021b). However, examining Block 1 performance alone allows us a preliminary picture into the potential for direct transfer that could be tested further in future work using more suitable methods.

We conducted post hoc *t* tests of the group difference in Block 1 to detect how much of a group difference existed even in our earliest time point (Table 1). We found that musicians performed significantly better than nonmusicians on Block 1 of the musical-span task, with a score (represented by the sum of correct responses over the total number of trials completed) that was 5 units higher, $b = -5.07$, $t(1,430) = -16.69$, $p < .001$. Musicians also perform significantly better on the visuospatial task, with a score 0.44 units higher than nonmusicians, $b = -0.44$, $t(1,431) = -3.90$, $p < .001$. Finally, musicians were not significantly better than nonmusicians on the verbal task, but their scores tended to be slightly higher as well, $b = -0.25$, $t(1,431) = -1.91$, $p = .056$. With the caveats noted above in place (with respect to Block 1 not being an ideal measure of immediate performance), these early differences indicate that dynamic effects may unfold on top of immediate performance differences.

Implications for Future Research: Why Considering Time Matters

The central theoretical lesson of these analyses is that considering performance dynamics dramatically deepens inference. Collapsing performance across time (i.e., via calculating a single dependent variable for tasks aggregating over a large number of trials) risks obscuring whether group differences reflect faster learning, greater endurance, or broader task-general adaptation. Dynamic analyses, by contrast, allow researchers to falsify specific mechanistic predictions. Given that the methods were not designed with time-continuous analyses in mind, our analyses here cannot adjudicate definitively among all theoretical models of cognitive transfer. However, even with methods that are less than ideal with respect to considering time, the results nonetheless illustrate how temporal information can constrain interpretation.

Looking forward, future studies could extend this approach by more purposefully designing experiments that explicitly manipulate opportunities for various time-continuous effects to manifest (e.g., Cochrane & Green, 2021a; Dale et al., 2021). In addition, incorporating “tests of no interest” (Green et al., 2014) would allow researchers to further disentangle stimulus-specific from domain-general explanations of the musicians’ memory advantage.

Conclusions

Grassi et al. (2025) clearly established that musicians differ from nonmusicians in short-term-memory performance. The next major advance lies in understanding how those differences arise. In the present commentary, we illustrate that treating performance as a dynamic process reveals a richer and more nuanced picture of the differences that exist between groups, one in which musicians appear to benefit from enhanced within-tasks learning and resistance to performance decline over time as well as possibly some degree of immediate benefits to task performance. We hope this emphasis on temporal dynamics inspires future work that leverages time not as a nuisance variable but as a theoretical lens.

Transparency

Action Editor: David A. Sbarra

Editor: Felix J. Thoemmes

Author Contributions

Tilo Strobach: Conceptualization; Project administration; Resources; Supervision; Writing – original draft; Writing – review & editing.

Emma G. Cunningham: Data curation; Formal analysis; Methodology; Writing – review & editing.

Shanti Astra: Data curation; Formal analysis; Methodology; Writing – review & editing.

C. Shawn Green: Data curation; Formal analysis; Methodology; Writing – review & editing.

Declaration of Conflicting Interests

The authors declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Open Practices

This article has received the badge for Preregistration. More information about the Open Practices badges can be found at <http://www.psychologicalscience.org/publications/badges>.



ORCID iDs

Tilo Strobach  <https://orcid.org/0000-0001-7840-0651>

Shanti Astra  <https://orcid.org/0009-0003-5742-7347>

C. Shawn Green  <https://orcid.org/0000-0002-9290-0262>

References

- Bavelier, D., Green, C. S., Pouget, A., & Schrater, P. (2012). Brain plasticity through the life span: Learning to learn and action video games. *Annual Review of Neuroscience*, 35, 391–416. <https://doi.org/10.1146/annurev-neuro-060909-152832>
- Cochrane, A., & Green, C. S. (2021a). Assessing the functions underlying learning using by-trial and by-participant models: Evidence from two visual perceptual learning paradigms. *Journal of Vision*, 21(13), Article 5. <https://doi.org/10.1167/jov.21.13.5>

- Cochrane, A., & Green, C. S. (2021b). Trajectories of performance change indicate multiple dissociable links between working memory and fluid intelligence. *npj Science of Learning*, 6, Article 33. <https://doi.org/10.1038/s41539-021-00111-w>
- Cochrane, A., Lu, Z.-L., & Green, C. S. (2024). Modeling perceptual learning as a continuous function of time-on-task increases theoretical specificity and statistical power. *Journal of Cognitive Enhancement*, 8, 294–307. <https://doi.org/10.1007/s41465-024-00303-3>
- Cochrane, A., Sims, C. R., Bejjanki, V. R., Green, C. S., & Bavelier, D. (2023). Multiple timescales of learning indicated by changes in evidence-accumulation processes during perceptual decision-making. *npj Science of Learning*, 8(1), Article 19. <https://doi.org/10.1038/s41539-023-00168-9>
- Dale, G., Cochrane, A., & Green, C. S. (2021). Individual difference predictors of learning and generalization in perceptual learning. *Attention, Perception, & Psychophysics*, 83(5), 2241–2255. <https://doi.org/10.3758/s13414-021-02268-3>
- Gathercole, S. E., Dunning, D. L., Holmes, J., & Norris, D. (2019). Working memory training involves learning new skills. *Journal of Memory and Language*, 105, 19–42. <https://doi.org/10.1016/j.jml.2018.10.003>
- Grassi, M., Talamini, F., Altoè, G., Brattico, E., Caclin, A., Carretti, B., Draai-Zerbib, V., Ferreri, L., Gambarota, F., Grahn, J., Fovino, L. G. N., Rocco, M., Rodriguez-Fornells, A., Swaminathan, S., Tillmann, B., Vuust, P., Wilbiks, J., Zentner, M., Aguilar, K., . . . Zappa, A. (2025). Do musicians have better short-term memory than non-musicians? A multilab study. *Advances in Methods and Practices in Psychological Science*, 8(4). <https://doi.org/10.1177/25152459251379432>
- Green, S., Strobach, T., & Schubert, T. (2014). Methodological standards in training and transfer experiments. *Psychological Research*, 78(6), 756–772. <https://doi.org/10.1007/s00426-013-0535-3>
- Kattner, F., Cochrane, A., Cox, C. R., Gorman, T. E., & Green, C. S. (2017). Perceptual learning generalization from sequential perceptual training as a change in learning rate. *Current Biology*, 27(6), 840–846. <https://doi.org/10.1016/j.cub.2017.01.046>
- Lövdén, M., Bäckman, L., Lindenberger, U., Schaefer, S., & Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychological Bulletin*, 136(4), 659–676. <https://doi.org/10.1037/a0020080>
- Pessiglione, M., Blain, B., Wiehler, A., & Naik, S. (2025). Origins and consequences of cognitive fatigue. *Trends in Cognitive Sciences*, 29, 730–749. <https://doi.org/10.1016/j.tics.2025.02.005>
- Román-Caballero, R., Başkent, D., Caclin, A., Ferreri, L., Fiveash, A., François, C., Grassi, M., Harding, E. E., Lima, C., Overy, K., Rodriguez-Fornells, A., Roncaglia-Denissen, M. P., Schellenberg, E. G., Slevc, L. R., Talamini, F., Tillmann, B., Trainor, L., & Zappa, A. (2025). *Common evidence, multiple interpretations: Commentaries on a multilab study on musicians' short-term memory*. OSF. https://doi.org/10.31234/osf.io/2z7jc_v1
- Strobach, T., & Karbach, J. (2021). *Cognitive training: An overview of features and applications* (2nd ed.). Springer. <https://doi.org/10.1007/978-3-030-39292-5>
- Taatgen, N. A. (2013). The nature and transfer of cognitive skills. *Psychological Review*, 120(3), 439–471. <https://doi.org/10.1037/a0033138>
- von Bastian, C. C., Belleville, S., Udale, R. C., Reinhartz, A., Essouni, M., & Strobach, T. (2022). Mechanisms underlying training-induced cognitive changes. *Nature Reviews Psychology*, 1, 30–41. <https://doi.org/10.1038/s44159-021-00001-3>
- Zhang, R.-Y., Chopin, A., Shibata, K., Lu, Z.-L., Jaeggi, S. M., Buschkuhl, M., Green, C. S., & Bavelier, D. (2021). Action video game play facilitates “learning to learn.” *Communications Biology*, 4(1), Article 1154. <https://doi.org/10.1038/s42003-021-02652-7>