Supplementary materials

Additional information about sample characteristics

We used an ID filter implemented in Prolific to ensure that participants who were recruited for Experiment 1 would not be allowed to take part in Experiment 2.

Sample characteristics: Experiment 1

The number of studies participants had previously successfully completed on Prolific, the number of studies from which they were rejected, their approval rate, and the time they took to complete the experiment are shown in Table S1. The nationality of participants is shown on Figure S1.

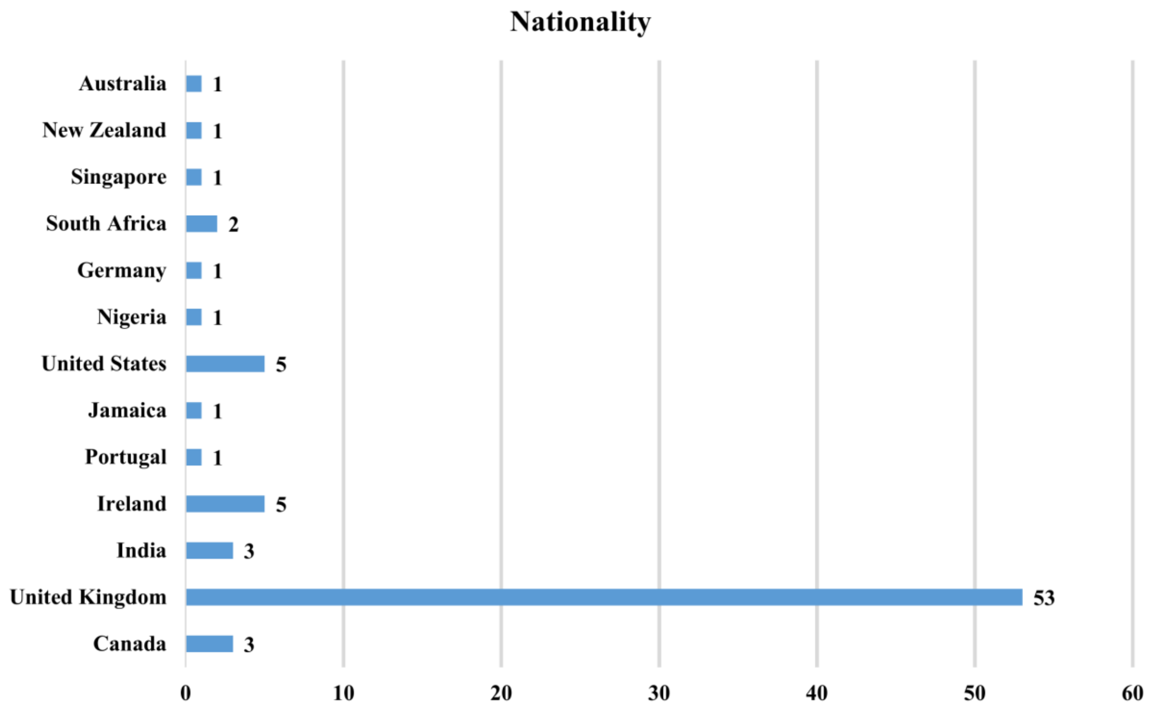
Table S1

Supplementary information about participants included in Experiment 1 (n = 78)

	Number of approvals	Number of rejections	Prolific scores	Time taken (minutes)
Min	21	0	96	39.18
Q1	137.25	0	100	58
Median	291	1	100	67.63
Q3	568.25	2.75	100	79.71
Max	2079	13	100	115

Figure S1

Nationality of participants included in Experiment 1



Sample characteristics: Experiment 2

The number of studies participants had previously successfully completed on Prolific, the number of studies from which they were rejected, their approval rate, and the time they took to complete the experiment are shown in Table S2. The nationality of participants is shown on Figure S2.

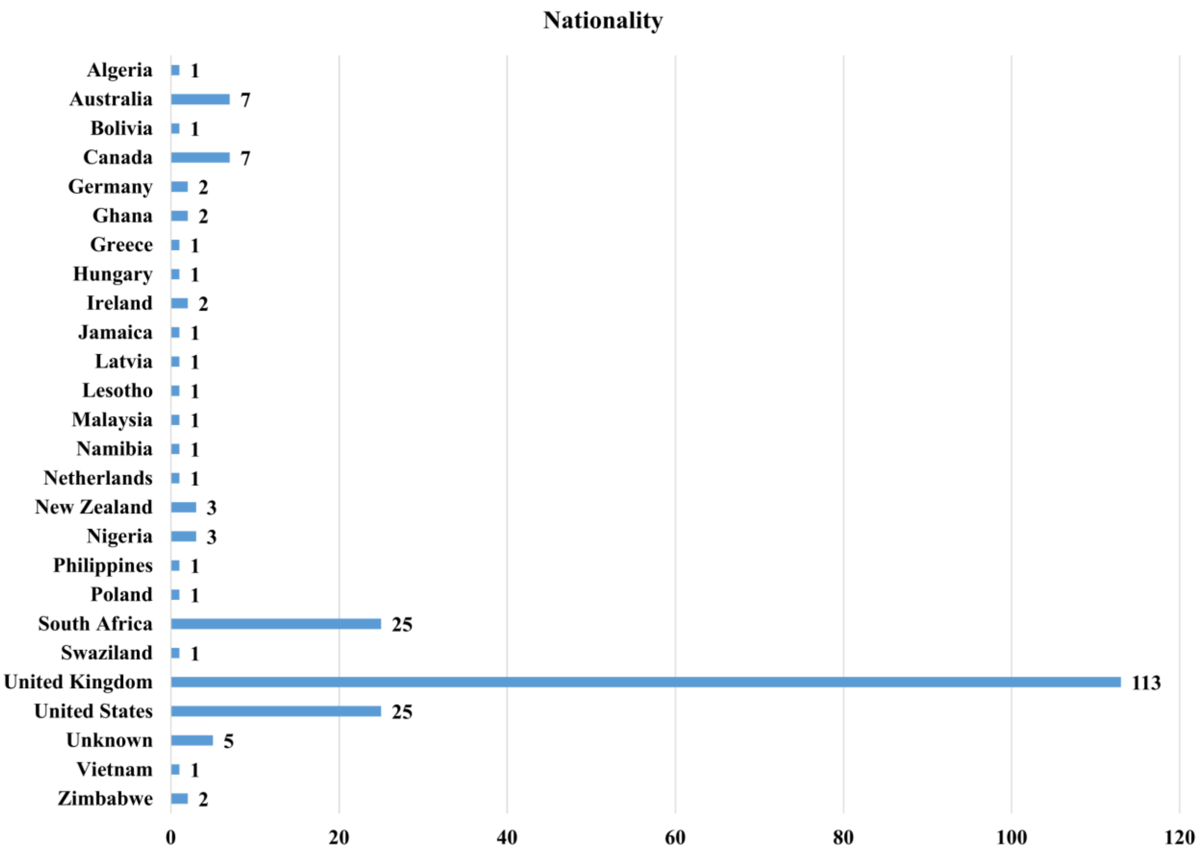
Table S2

Supplementary information about participants included in Experiment 2 (n = 210)

	Number of approvals	*Number of rejections*	*Prolific scores*	*Time taken (minutes)*
Min	18	0	87	43.17
Q1	103	0	100	56.68
Median	208.5	1	100	66.65
Q3	447.75	2	100	83
Max	3526	20	100	145

Figure S2

Nationality of participants included in Experiment 2



Supplementary information about video stimuli

Stimuli used in Experiments 1 and 2 were edited based on the same eight videos showing people engaged in daily life activities (see Figure S3). These had previously been used in studies on event segmentation (e.g., Eisenberg & Zacks, 2016; Kurby & Zacks, 2011; Sargent et al., 2013). Data from these previous studies enabled us to create 40-s videos depicting events with either a high or a low number of normative event boundaries (EBs; see Table 3).

Figure S3

Daily life activities depicted in the eight basic videos

Note. From left to right: a woman washing her car (432s), a woman preparing breakfast (329s), a man gardening (353 s), a man photocopying the pages of a book (348s), a man sweeping a room (329s), a lady setting up a tent (379s), a man changing the tire of his car (342s), a man preparing a livingroom for a birthday party (378s).

Table S3

Supplementary information about the normative segmentation data of the selected 40-s videos

Experiment 1

	Washing car	Breakfast	Gardening	Photocopying	Sweeping	Tent	Tire	Party
Number of participants in the norm	40	42	42	28	28	67	28	42
Number of normative EBs	1.65	4.62	8.17	5	5.8	2.18	3.79	5.24
Density of EBs	Low	Low	High	High	High	Low	High	Low

Experiment 2

	Washing car	Breakfast	Gardening	Photocopying	Sweeping	Tent	Tire	Party
Number of participants in the norm	40	42	42	28	28	67	28	42
Number of normative EBs	1	8.5	4.26	5	5.8	6	0.7	8.45
Density of EBs	Low	High	Low	High	High	High	Low	Low

Note. The number of normative EBs was computed by dividing the total number of EBs identified in the videos during previous studies by the number of participants who performed the segmentation task. In Experiment 2, despite the high number of keypresses, the party video was selected as a low EBs stimulus because the keypresses refer to the repetition of the same action (i.e., a man putting a plate on the table).

Additional information about the complex span tasks

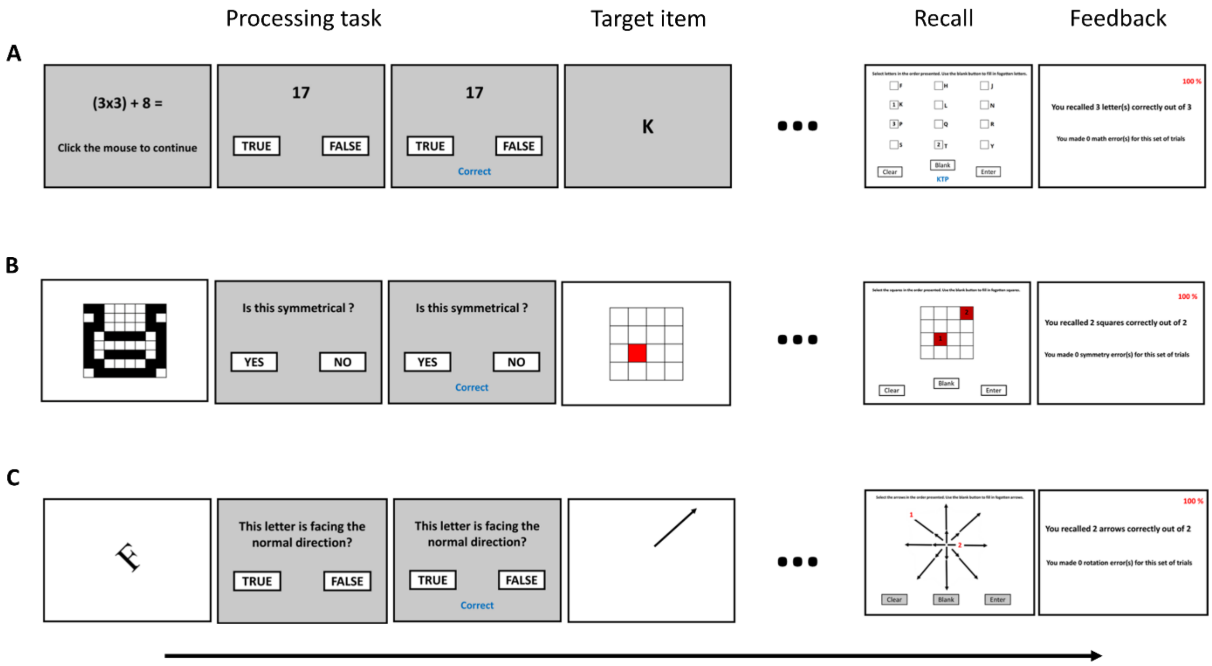
The unfolding of the trials was similar across the three tasks (OSpan, SymSpan, RotSpan). First, participants were presented with a problem from the processing task and were instructed to solve it as quickly as possible (and to click on the mouse as soon as it was done; this step must be completed within a limited time window; see Methods). Then, they gave their answer and immediately received a feedback. Immediately after, one of the items they had to memorize was briefly displayed. A new problem then appeared on the screen and so on. At recall, participants were asked to select the items they were presented with, in the order they were presented. They were provided with a "Blank" button to indicate that they remembered seeing an item but did not know which one and a "Clear" button for modifying their answers. After recall, they received a feedback on how many items they correctly recalled and how many errors they made during the processing problems. Immediately afterwards, the next trial began. The specificities of each task are shown in Figure S4.

Before each task, participants received built-in standardized instructions. Then, they received a practice session consisting of three parts. First, they were trained on the "storage" part alone, then on the "processing" part alone. During these first two parts, they received feedback after each answer they gave. Finally, they had to perform a series of trials including both the "storage" and the "processing" parts (as in the main task).

During the "processing only" part of the training, participants' response times were recorded. During the processing part of the main task, participants had to systematically give their answers within a defined time window (to minimize opportunities to refresh the to-be-remembered items). The upper limit of this time window was individualized for each participant and corresponded to the participant's average response time during the last part of the training + 2.5 *SD*s.

Figure S4

Unfolding of a trial in the three complex span tasks



Note. ***Panel A***: *Operation span. Participants first see a math problem (parenthetical multiplication or division problem, followed by a number to add or subtract from the product or dividend ; e.g., (3x3) + 8 = ?; all terms and signs are randomly selected). Then, a digit (e.g., 3) is presented and the participants are required to click either a "true" or "false" box to indicate whether the number presented is the correct solution to the problem they saw just before. After receiving feedback on their response, they are presented with a letter for 1 s (consonant; written in Arial font, size 28). Then, another math problem is displayed and so on. At recall, a 4 x 3 matrix of letters (F, H, J, K, L, N, P, Q, R, S, T, and Y; i.e., the 12 letters used in the task) is displayed. Participants must select the letters they have seen, in the correct order, by checking the boxes next to them.* ***Panel B***: *Symmetry span. During the processing task, an 8 × 8 matrix is presented with some squares filled with black. Then, participants have to judge whether the black-square design is symmetric along its vertical axis by clicking on either a "yes" or "no" box. Next, a 4 x 4 array with one cell filled with red is displayed for 650 ms. At recall, participants have to click on the cells of an empty matrix to reproduce the sequence of red square locations in the order they appeared during the previous presentations.* ***Panel C***: *Rotation Span. The processing task starts with the presentation of a normal or mirror-reversed G, F, R, J or L rotated at 0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°. Then, participants must indicate whether the letter was in the normal orientation or mirror reversed (by clicking either on a "yes" or a "no" box). After that, a short or long arrow pointing in one of eight directions is displayed for 650 ms. At recall, participants are presented with eight short and eight long arrows radiating from the center of the screen. Participants have to recall all of the arrows from the preceding displays, in the order they appeared, by clicking on their head.*

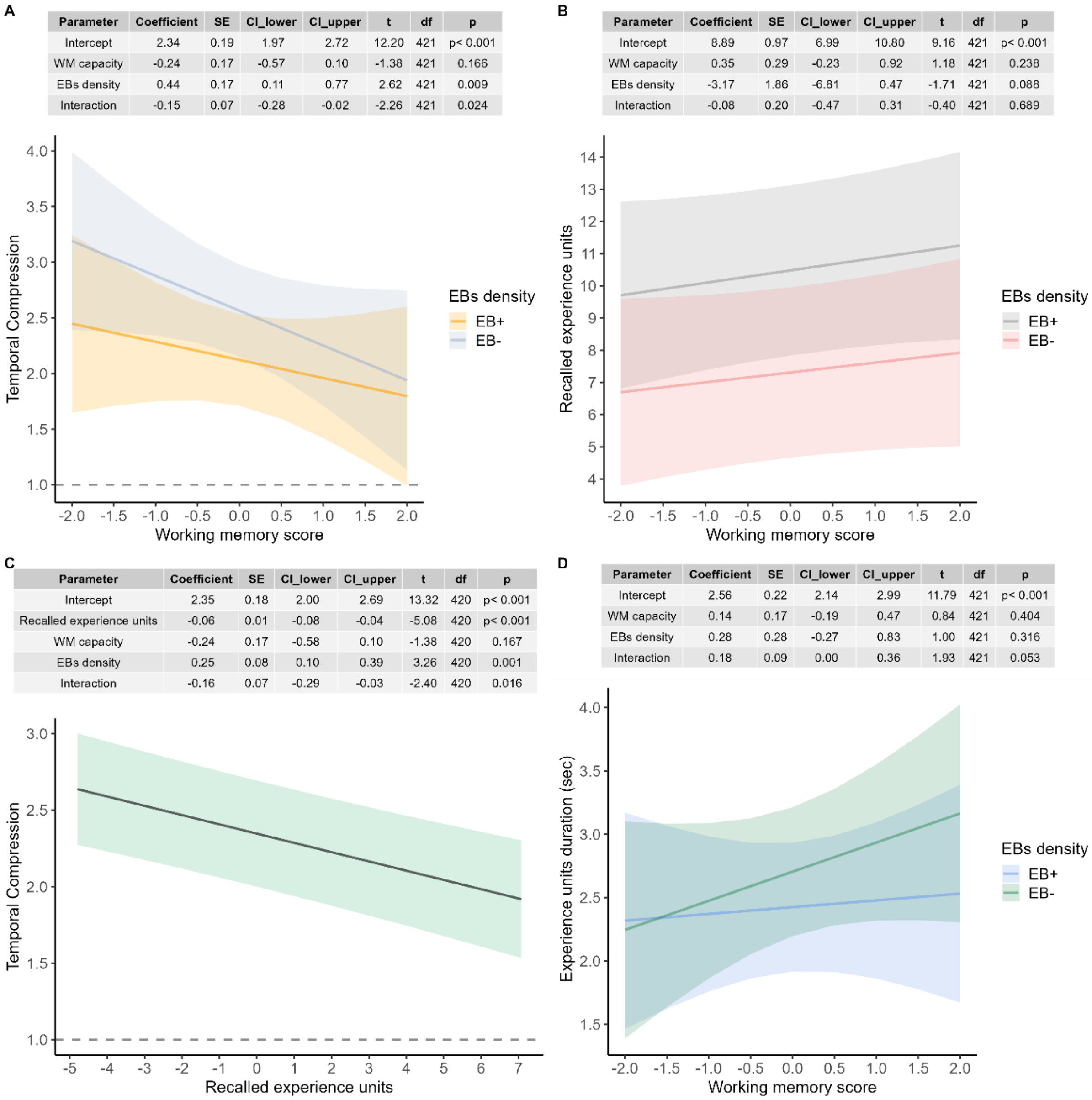
Processing accuracy criteria in the WM tasks

When using complex span tasks to evaluate WM capacity, it is usually recommended to exclude participants who show poor performance on the processing part of the tasks (to avoid including participants who neglected the processing part of the tasks; Conway et al., 2005). A common approach is to exclude participants who scored below 85% of accuracy (Redick et al., 2012; Unsworth et al., 2005), which is what we had planned to do when pre-registering the study. However, Gonthier et al. (2016) argued that this criterion is too severe and its application to our sample would have led the exclusion of more than 25% of the data. Thus, following the approach of Gonthier et al., (2016), we decided to only exclude participants whose processing accuracy was below the fifth lower percentile of the sample for one of the three tasks (see the *Method* section).

To ensure that the relations we observed between our variables of interest were not dependent on the chosen exclusion criterion, we also performed our main analyses (for the two experiments) with a dataset including only participants who scored above 85% on the processing parts of the three tasks. The results were similar to those reported in the main text (for Experiment 1, see Figure S5; for Experiment 2 see Figure S6).

Figure S5

Results for the main analyses of Experiment 1 when including only participants who scored above 85% for the processing part of each of the three complex span tasks (N participants = 56, N observations = 425)



A

Parameter	Coefficient	SE	CI_lower	CI_upper	t	df	p
Intercept	2.34	0.19	1.97	2.72	12.20	421	p< 0.001
WM capacity	-0.24	0.17	-0.57	0.10	-1.38	421	0.166
EBs density	0.44	0.17	0.11	0.77	2.62	421	0.009
Interaction	-0.15	0.07	-0.28	-0.02	-2.26	421	0.024

B

Parameter	Coefficient	SE	CI_lower	CI_upper	t	df	p
Intercept	8.89	0.97	6.99	10.80	9.16	421	p< 0.001
WM capacity	0.35	0.29	-0.23	0.92	1.18	421	0.238
EBs density	-3.17	1.86	-6.81	0.47	-1.71	421	0.088
Interaction	-0.08	0.20	-0.47	0.31	-0.40	421	0.689

C

Parameter	Coefficient	SE	CI_lower	CI_upper	t	df	p
Intercept	2.35	0.18	2.00	2.69	13.32	420	p< 0.001
Recalled experience units	-0.06	0.01	-0.08	-0.04	-5.08	420	p< 0.001
WM capacity	-0.24	0.17	-0.58	0.10	-1.38	420	0.167
EBs density	0.25	0.08	0.10	0.39	3.26	420	0.001
Interaction	-0.16	0.07	-0.29	-0.03	-2.40	420	0.016

D

Parameter	Coefficient	SE	CI_lower	CI_upper	t	df	p
Intercept	2.56	0.22	2.14	2.99	11.79	421	p< 0.001
WM capacity	0.14	0.17	-0.19	0.47	0.84	421	0.404
EBs density	0.28	0.28	-0.27	0.83	1.00	421	0.316
Interaction	0.18	0.09	0.00	0.36	1.93	421	0.053

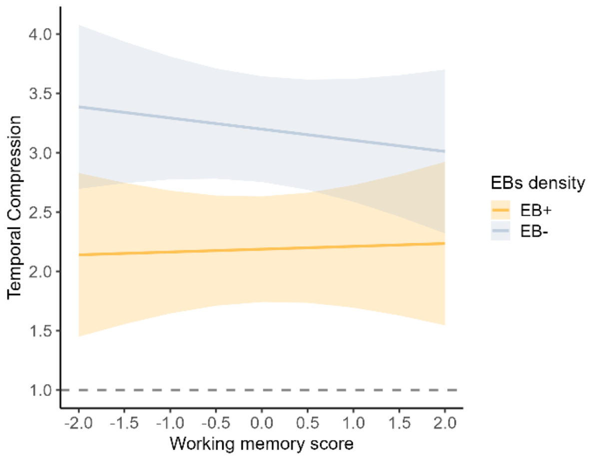
Note. (A) Prediction of temporal compression by WM capacity for high and low EBs density videos. (B) Prediction of the number of recalled experience units by WM capacity for high and low EBs density videos. (C) Prediction of temporal compression by the number of recalled experience units. (D) Prediction of remembered experience units duration by WM capacity, for high and low EBs density videos.

Figure S6

Results for the main analyses of Experiment 2 when including only participants who scored above 85% for the processing part of each of the three complex span tasks (N participants = 166, N observations = 1220)

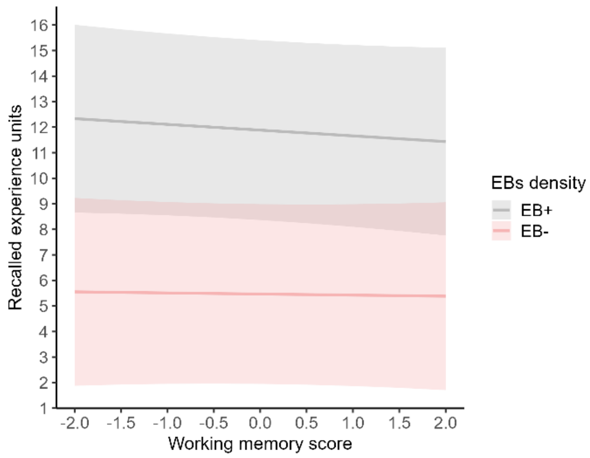
A

Parameter	Coefficient	SE	CI_lower	CI_upper	t	df	p
Intercept	2.69	0.18	2.33	3.05	14.67	1216	p< 0.001
WM capacity	-0.04	0.13	-0.29	0.22	-0.27	1216	0.788
EBs density	1.01	0.27	0.49	1.53	3.80	1216	p< 0.001
Interaction	-0.12	0.06	-0.24	0.01	-1.83	1216	0.067



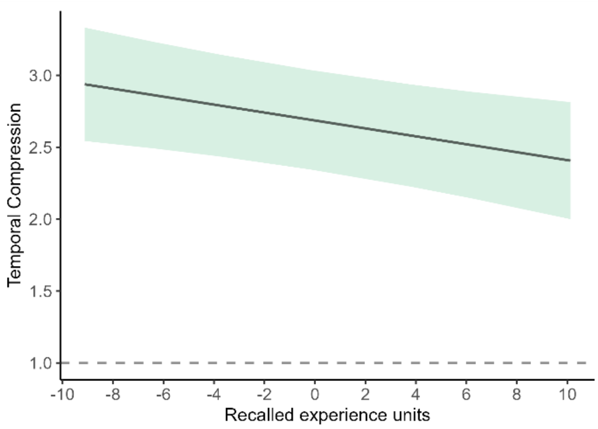
B

Parameter	Coefficient	SE	CI_lower	CI_upper	t	df	p
Intercept	8.68	1.28	6.17	11.18	6.78	1216	p< 0.001
WM capacity	-0.13	0.26	-0.63	0.37	-0.52	1216	0.601
EBs density	-6.42	2.51	-11.34	-1.49	-2.55	1216	0.011
Interaction	0.18	0.18	-0.17	0.53	1.02	1216	0.307



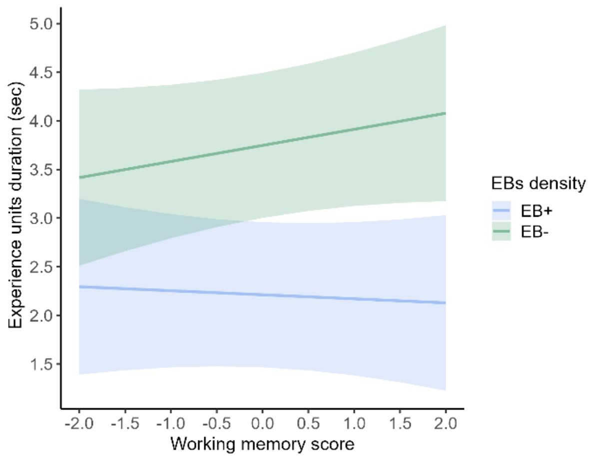
C

Parameter	Coefficient	SE	CI_lower	CI_upper	t	df	p
Intercept	2.69	0.18	2.35	3.04	15.29	1215	p< 0.001
Recalled experience units	-0.03	0.01	-0.05	-0.01	-2.58	1215	0.01
WM capacity	-0.03	0.13	-0.29	0.22	-0.24	1215	0.809
EBs density	0.82	0.26	0.32	1.32	3.21	1215	0.001
Interaction	-0.11	0.06	-0.23	0.02	-1.67	1215	0.096



D

Parameter	Coefficient	SE	CI_lower	CI_upper	t	df	p
Intercept	2.98	0.28	2.43	3.53	10.61	1216	p< 0.001
WM capacity	0.06	0.12	-0.18	0.30	0.51	1216	0.613
EBs density	1.54	0.51	0.53	2.54	3.00	1216	0.003
Interaction	0.21	0.09	0.04	0.37	2.42	1216	0.016



Note. (A) Prediction of temporal compression by WM capacity for high and low EBs density videos. (B) Prediction of the number of recalled experience units by WM capacity for high and low EBs density videos. (C) Prediction of temporal compression by the number of recalled experience units. (D) Prediction of remembered experience units duration by WM capacity, for high and low EBs density videos.

Inter-rater agreement

Experiment 1

Following guidelines proposed by Hallgren (2012) and Koo & Li (2016), inter-rater reliability was assessed with a *two-way agreement ICC* (single rater, *average-measures*; (McGraw & Wong, 1996; Shrout & Fleiss, 1979). The resulting *ICC* was 0.93 (95%*CI* [0.90, 0.95]). Nevertheless, the number of recalled experience units was not normally distributed. We thus computed a non-parametric agreement index (binomial proportions test; Bland & Altman, 1999), which indicated that the difference between the two raters was equal or inferior to one experience unit in 86.67% of the cases (95%*CI* [79.34, 91.67]). We further inspected the agreement between the two raters through a Bland-Altman analysis (Giavarina, 2015). We looked for the presence of potential fixed bias, or, in other words, systematic variations in the differences between the number of experience units identified by the first and the second rater as a function of the number of recalled experience units (computed by averaging estimations of the two raters). There was no notable bias.

Experiment 2

Again we had good inter-rater agreement for the scoring of the number of recalled experience units (*ICC* = 0.96, 95%*CI* [0.94, 0.97], robust agreement = 0.80, 95%*CI* [0.76, 0.84]). The Bland-Altman analysis did not reveal any substantial bias.

R packages

The R project dependencies were managed with renv version 0.17.3 (Ushey & Wickham, 2023). Data formatting and pre-processing were performed with the help of the dplyr package (v1.0.5; Wickham et al., 2021).

Internal consistency/reliability measures were computed with the psych package (v2.3.6; Revelle, 2023).

Regarding inter-rater agreement, *ICCs* were computed with the package irr (v0.84.1; Gamer et al., 2019) and the non-parametric agreement indices with the package SimplyAgree (v0.1.2; Caldwell, 2022).

Robust linear mixed-effects models were fitted with the robustlmm package (v2.4.4, Koller, 2016). *CIs* and *p-values* associated with fixed effects coefficients, as well as models' *R2s,* were computed with functions from the packages parameters (v0.13.0; Lüdecke et al., 2020) and performance (v0.7.1; Lüdecke et al., 2021). For each model, we extracted estimated marginal means, estimated marginal slopes and their standard errors with the package emmeans (v1.6.3; Lenth, 2016). Extraction of models' estimates across different values of manipulated variables (and their asymptotic 95% *CI*s) was done with the effects package (v4.2.0; Fox & Weisberg, 2019).

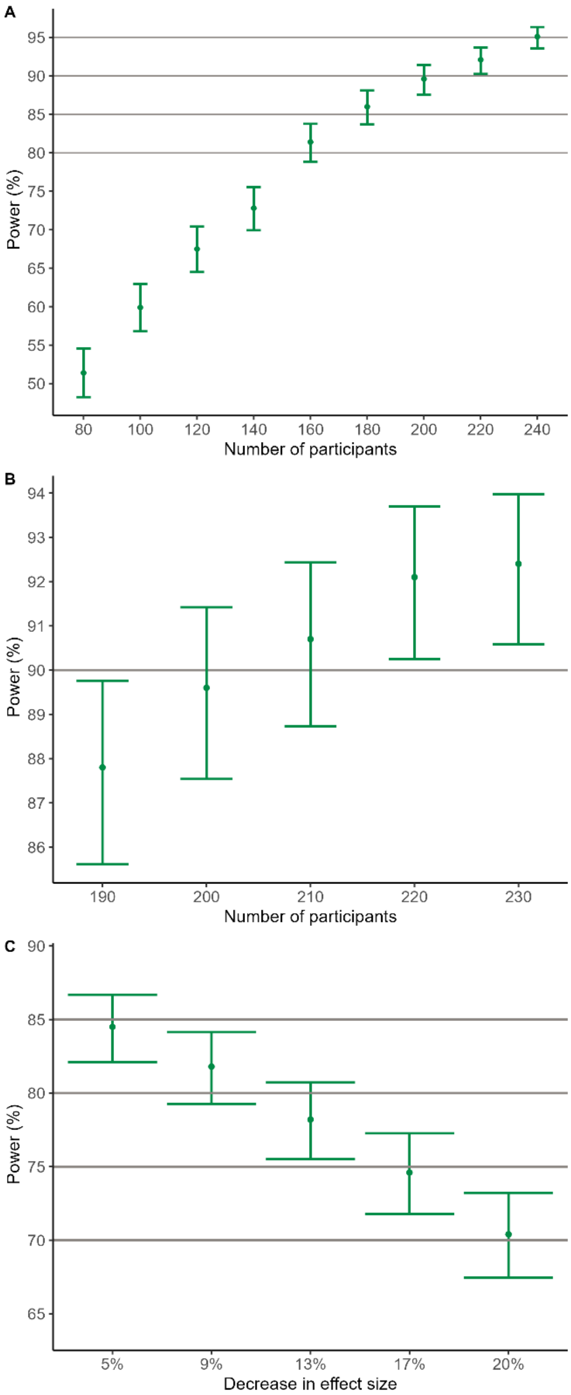
Tables were made with the packages insight (v0.14.15; Lüdecke et al., 2019), flextable (v0.9.1; Gohel & Skintzos, 2023) and rempsyc (v0.1.1; Thériault, 2023). Figures were made with the packages ggplot2 (v3.3.5; Wickham, 2016), gridExtra (v2.3; Auguie, 2017) and ggpubr (v0.4.0; Kassambara, 2020).

Power analyses for Experiment 2

The sample size of Experiment 2 was estimated a priori using the R package simr (v1.0.5; Green & MacLeod, 2016; Kumle et al., 2021). We used data from Experiment 1 to run Monte-Carlo simulations (Brysbaert & Stevens, 2018; DeBruine & Barr, 2021) in order to estimate the power we would have for the detection of the interaction between WM capacity and EBs density in the prediction of temporal compression across a range of sample sizes (between 80 and 240 participants). The estimated power (and its 95% *CI*) for each sample size is reported on Figure S7 (panels A and B). We also conducted a sensitivity analysis. With a sample size fixed at 210 participants, we estimated the power we would have for the detection of an interaction, with an effect size from 5% to 20% smaller than the one observed in Experiment 1 (Figure S7C).

Figure S7

Power curves for the detection of the interaction between WM capacity and EBs density in the prediction of temporal compression (based on 1000 simulations)



Note. (A) estimates of statistical power (and its 95% CI) across a range of sample sizes (ranging from 80 to 240 participants). (B) estimates for sample sizes ranging from 190 to 230 participants (zoom on the panel A). (C) estimates of the statistical power achievable with a sample size of 210 participants and an effect size from 5% to 20% smaller than the one obtained in Experiment 1.

Descriptive statistics

Descriptive statistics for each WM task and outcome variables are shown in Tables S4 and S5 (for Experiment 1) and Tables S7 and S8 (for Experiment 2). A detailed distributional plot for each outcome variable is displayed in Figure S8 for Experiment 1 and S10 for Experiment 2. In Tables S6 (for Experiment 1) and S9 (for Experiment 2), we report correlations between all variables involved in the statistical analyses. Finally, Figures S9 (for Experiment 1) and S11 (for Experiment 2) show the distribution of the differences of temporal compression rates between EB+ and EB- stimuli as a function of participants' WM score.

Table S4

Descriptive statistics regarding WM tasks and correlations between them (Experiment 1)

Span Tasks	Descriptive Statistics					Correlations (Pearson's *r*)		
	N	Mean	SD	Skew	Kurtosis	OSpan	RotSpan	SymSpan
Ospan	78	19.63	5.21	-1.60	2.71			0.34
SymSpan	78	10.37	2.67	-0.37	-0.48		0.50	
RotSpan	78	9.24	3.25	-0.33	-0.95	0.17		

Table S5

Descriptive statistics for outcome variables (Experiment 1)

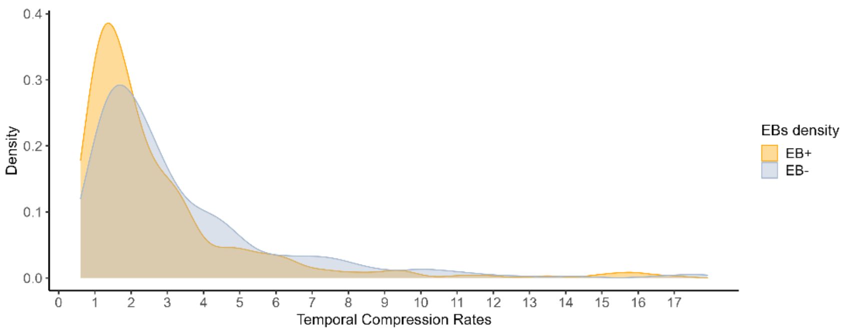
Outcome	N	Mean	SD	Skew	Kurtosis	Distribution
TCR	78	3.19	2.60	2.44	7.07	
Recalled experience units	78	8.47	2.24	-0.06	-0.74	
Recalled experience units duration	78	2.84	1.74	1.72	3.83	

Figure S8

Density plot and descriptive statistics for the three outcome variables (Experiment 1). Values are reported separately for each EBs density

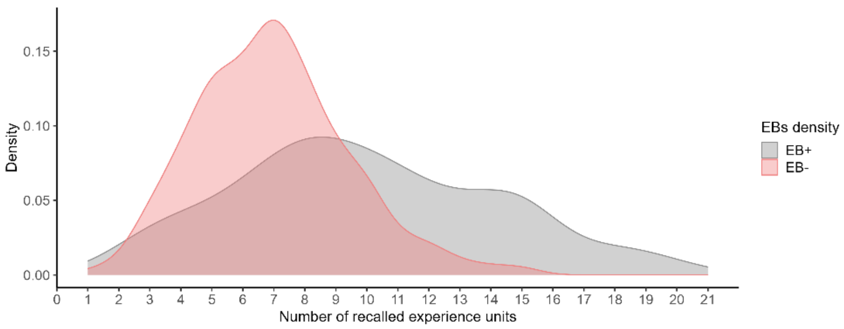
A

EBs_Density	Q1	Median	Q3	Mean	SD	Skewness	Kurtosis
EB+	1.21	1.87	3.22	2.77	2.63	2.95	0.64
EB-	1.50	2.37	4.14	3.30	2.77	2.37	0.70



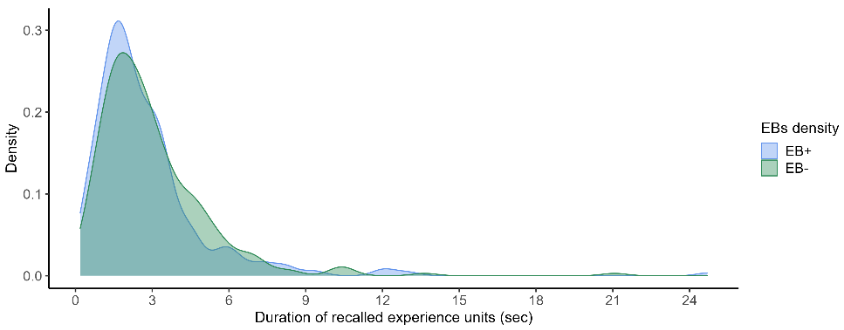
B

EBs_Density	Q1	Median	Q3	Mean	SD	Skewness	Kurtosis
EB+	7	10	13	9.97	4.25	0.26	0.82
EB-	5	7	8	6.96	2.45	0.50	0.77



C

EBs_Density	Q1	Median	Q3	Mean	SD	Skewness	Kurtosis
EB+	1.43	2.2	3.33	2.78	2.41	3.82	0.63
EB-	1.59	2.5	3.81	2.99	2.22	2.96	0.68



*Note. (A) Temporal compression rates. (B) The number of recalled experience units. (C) The duration of recalled experience units. Skewness and Geary's kurtosis measures (*Borroni & De Capitani, 2022; Geary, 1936) *were computed with the package moments (v0.14.1; Komsta & Novomestky, 2022)*

Table S6

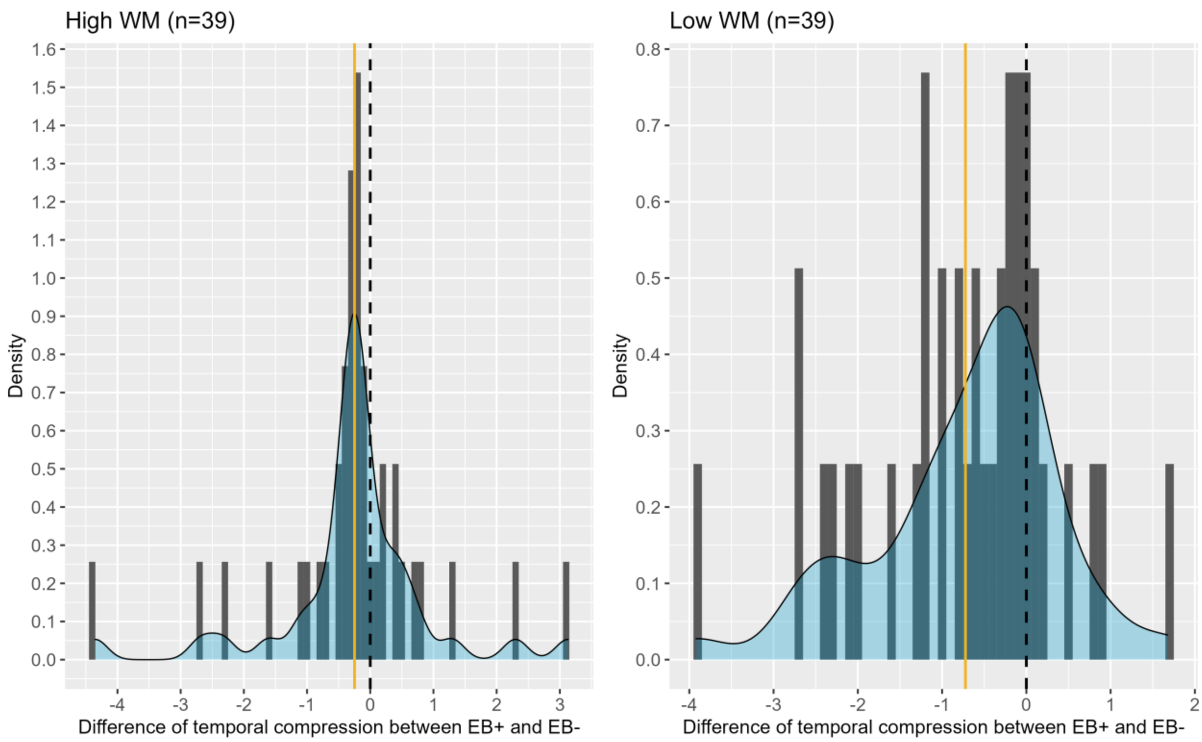
Correlation matrices for Experiment 1

Correlations between WM scores and the three outcome variables for EB- stimuli (Spearman's rho)				
	TCR	Recalled experience units	Recalled experience units duration	WM Score
TCR	1.00	-0.31	-0.87	-0.25
Recalled experience units	-0.31	1.00	-0.09	0.28
Recalled experience units duration	-0.87	-0.09	1.00	0.12
WM Score	-0.25	0.28	0.12	1.00

Correlations between WM scores and the three outcome variables for EB+ stimuli (Spearman's rho).				
	TCR	Recalled experience units	Recalled experience units duration	WM Score
TCR	1.00	-0.28	-0.77	-0.16
Recalled experience units	-0.28	1.00	-0.28	0.25
Recalled experience units duration	-0.77	-0.28	1.00	-0.01
WM Score	-0.16	0.25	-0.01	1.00

Figure S9

Density plots representing the distribution of the differences of temporal compression between EB+ and EB- stimuli (at the subject level; Experiment 1)



Note. The golden line represents the average difference. (A) participants whose WM score was above the sample median. (B) participants whose WM score was below the sample median.

Table S7

Descriptive statistics regarding WM tasks and correlations between them (Experiment 2)

Span Tasks	Descriptive Statistics					Correlations (Pearson's *r*)		
	N	Mean	SD	Skew	Kurtosis	OSpan	RotSpan	SymSpan
Ospan	210	20.03	4.54	-1.07	1.02			0.31
SymSpan	210	9.93	3.04	-0.63	0.14		0.53	
RotSpan	210	9.60	2.98	-0.38	-0.41	0.33		

Table S8

Descriptive statistics for outcome variables (Experiment 2)

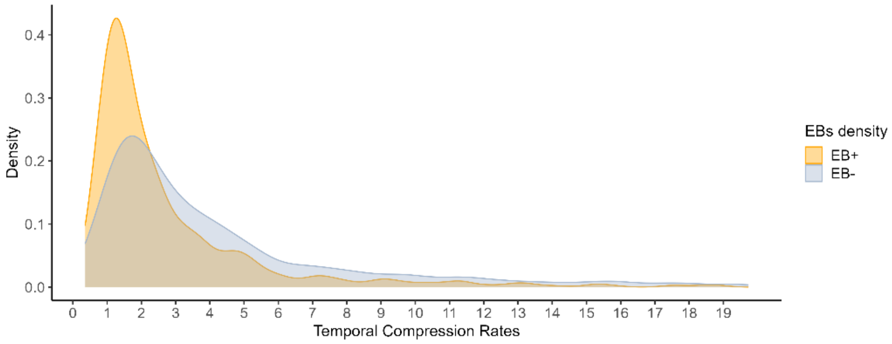
Outcome	N	Mean	SD	Skew	Kurtosis	Distribution
TCR	210	3.61	2.75	1.71	3.26	
Recalled experience units	210	8.76	3.21	0.47	0.52	
Recalled experience units duration	210	3.93	3.26	2.60	8.81	

Figure S10

Density plot and descriptive statistics for the three outcome variables (Experiment 2). Values are reported separately for each EBs density

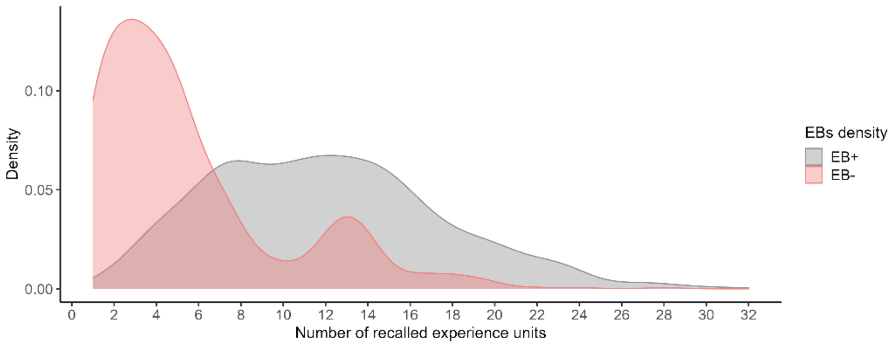
A

EBs_Density	Q1	Median	Q3	Mean	SD	Skewness	Kurtosis
EB+	1.20	1.78	3.16	2.77	2.78	2.81	0.65
EB-	1.66	2.82	5.09	4.18	3.80	1.91	0.72



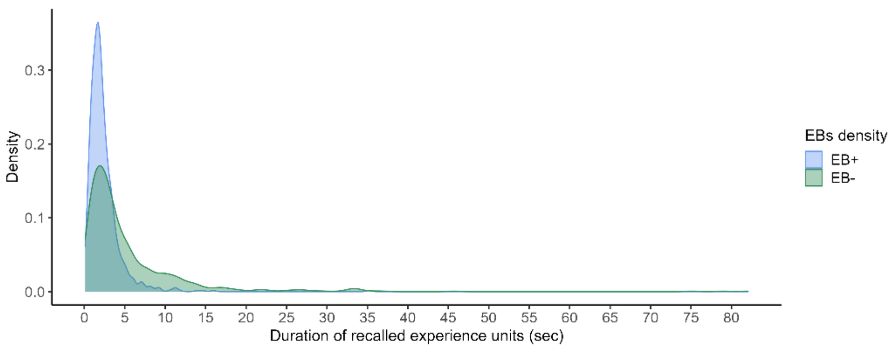
B

EBs_Density	Q1	Median	Q3	Mean	SD	Skewness	Kurtosis
EB+	8	12	15	11.97	5.45	0.49	0.81
EB-	2	4	7	5.58	4.49	1.41	0.76



C

EBs_Density	Q1	Median	Q3	Mean	SD	Skewness	Kurtosis
EB+	1.24	1.93	2.88	2.40	2.17	6.02	0.58
EB-	1.74	3.20	6.16	5.42	7.48	5.09	0.57



Note. (A) Temporal compression rates. (B) The number of recalled experience units. (C) The duration of recalled experience units.

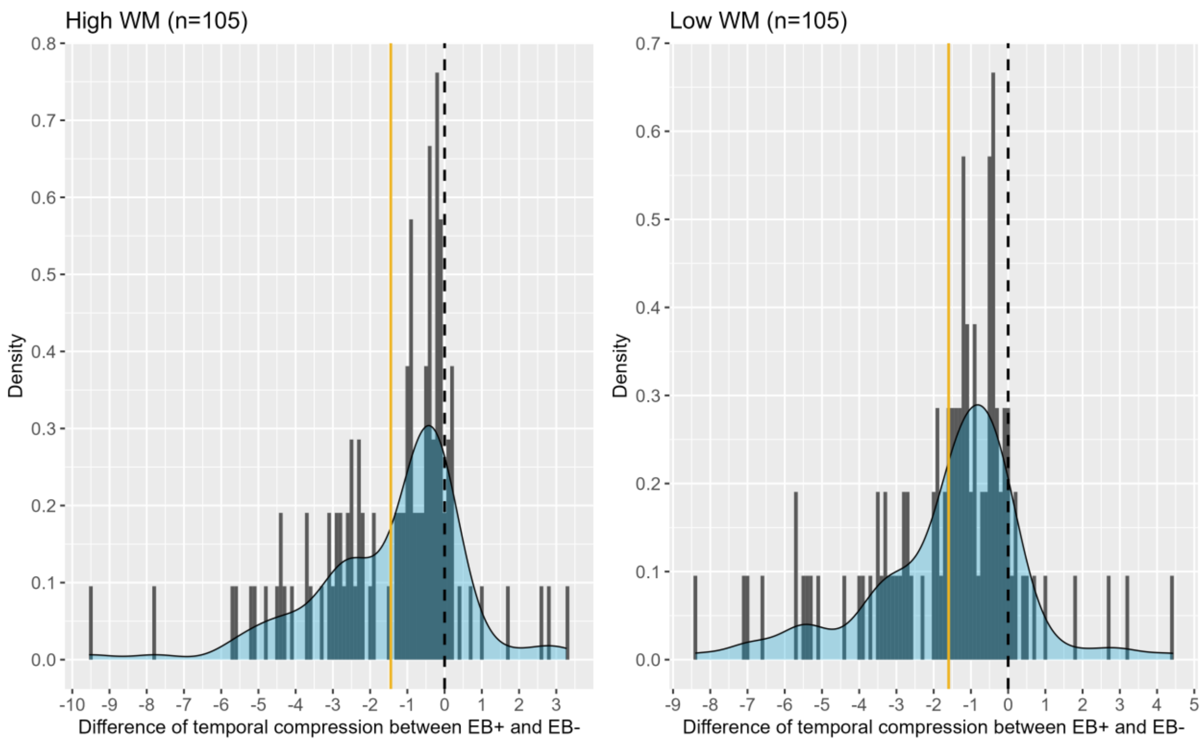
Table S9

Correlation matrices for Experiment 2

Correlations between WM scores and the three outcome variables for EB- stimuli (Spearman's rho)				
	TCR	Recalled experience units	Recalled experience units duration	WM Score
TCR	1.00	-0.27	-0.67	-0.07
Recalled experience units	-0.27	1.00	-0.36	-0.05
Recalled experience units duration	-0.67	-0.36	1.00	0.10
WM Score	-0.07	-0.05	0.10	1.00
Correlations between WM scores and the three outcome variables for EB+ stimuli (Spearman's rho)				
	TCR	Recalled experience units	Recalled experience units duration	WM Score
TCR	1.00	-0.44	-0.70	-0.04
Recalled experience units	-0.44	1.00	-0.18	-0.01
Recalled experience units duration	-0.70	-0.18	1.00	-0.02
WM Score	-0.04	-0.01	-0.02	1.00

Figure S11

Density plots representing the distribution of the differences of temporal compression between EB+ and EB- stimuli (at the subject level; Experiment 2)



Note. The golden line represents the average difference. (A) participants whose WM score was above the sample median. (B) participants whose WM score was below the sample median.

References

Auguie, B. (2017). *gridExtra : Miscellaneous Functions for « Grid » Graphics*. https://CRAN.R-project.org/package=gridExtra

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical Methods in Medical Reasearch*, *8*, 135-160.

Borroni, C. G., & De Capitani, L. (2022). Some measures of kurtosis and their inference on large datasets. *AStA Advances in Statistical Analysis*, *106*(4), 573-607. https://doi.org/10.1007/s10182-022-00442-y

Brysbaert, M., & Stevens, M. (2018). Power Analysis and Effect Size in Mixed Effects Models: A Tutorial. *Journal of Cognition*, *1*(1), 9. https://doi.org/10.5334/joc.10

Caldwell, A. R. (2022). SimplyAgree: An R package and jamovi Module for Simplifying Agreement and Reliability Analyses. *Journal of Open Source Software*, *7*(71), 4148. https://doi.org/10.21105/joss.04148

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769-786. https://doi.org/10.3758/BF03196772

DeBruine, L. M., & Barr, D. J. (2021). Understanding Mixed-Effects Models Through Data Simulation. *Advances in Methods and Practices in Psychological Science*, *4*(1), 251524592096511. https://doi.org/10.1177/2515245920965119

Eisenberg, M. L., & Zacks, J. M. (2016). Ambient and focal visual processing of naturalistic activity. *Journal of Vision*, *16*(2), 5. https://doi.org/10.1167/16.2.5

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd éd.). Sage. https://socialsciences.mcmaster.ca/jfox/Books/Companion/index.html

Gamer, M., Lemon, J., & Singh, I. F. P. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement*. https://www.r-project.org

Geary, R. C. (1936). Moments of the Ratio of the Mean Deviation to the Standard Deviation for Normal Samples. *Biometrika*, *28*, 295-307.

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica*, *25*(2), 141-151. https://doi.org/10.11613/BM.2015.015

Gohel, D., & Skintzos, P. (2023). *flextable: Functions for Tabular Reporting*. https://CRAN.R-project.org/package=flextable

Gonthier, C., Thomassin, N., & Roulin, J.-L. (2016). The composite complex span: French validation of a short working memory task. *Behavior Research Methods*, *48*(1), 233-242. https://doi.org/10.3758/s13428-015-0566-3

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493-498. https://doi.org/10.1111/2041-210X.12504

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, *8*(1), 23-34. https://doi.org/10.20982/tqmp.08.1.p023

Kassambara, A. (2020). *ggpubr: Ggplot2 Based Publication Ready Plots*. https://rpkgs.datanovia.com/ggpubr/

Koller, M. (2016). **robustlmm**: An *R* Package for Robust Estimation of Linear Mixed-Effects Models. *Journal of Statistical Software*, *75*(6). https://doi.org/10.18637/jss.v075.i06

Komsta, L., & Novomestky, F. (2022). *moments: Moments, Cumulants, Skewness, Kurtosis and Related Tests*. https://CRAN.R-project.org/package=moments

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Kumle, L., Võ, M. L.-H., & Draschkow, D. (2021). Estimating power in (generalized) linear mixed models : An open introduction and tutorial in R. *Behavior Research Methods*, *53*(6), 2528-2543. https://doi.org/10.3758/s13428-021-01546-0

Kurby, C. A., & Zacks, J. M. (2011). Age differences in the perception of hierarchical structure in events. *Memory & Cognition*, *39*(1), 75-91. https://doi.org/10.3758/s13421-010-0027-2

Lenth, R. V. (2016). Least-Squares Means: The *R* Package **lsmeans**. *Journal of Statistical Software*, *69*(1). https://doi.org/10.18637/jss.v069.i01

Lüdecke, D., Ben-Shachar, M., Patil, I., & Makowski, D. (2020). Extracting, Computing and Exploring the Parameters of Statistical Models using R. *Journal of Open Source Software*, *5*(53), 2445. https://doi.org/10.21105/joss.02445

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Lüdecke, D., Waggoner, P., & Makowski, D. (2019). insight: A Unified Interface to Access Information from Model Objects in R. *Journal of Open Source Software*, *4*(38), 1412. https://doi.org/10.21105/joss.01412

McGraw, K. O., & Wong, S. P. (1996). Forming Inferences About Some Intraclass Correlation Coefficients. *Psychological Methods*, *1*(1), 30-46.

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring Working Memory Capacity With Automated Complex Span Tasks. *European Journal of Psychological Assessment*, *28*(3), 164-171. https://doi.org/10.1027/1015-5759/a000123

Sargent, J. Q., Zacks, J. M., Hambrick, D. Z., Zacks, R. T., Kurby, C. A., Bailey, H. R., Eisenberg, M. L., & Beck, T. M. (2013). Event segmentation ability uniquely predicts event memory. *Cognition*, *129*(2), 241-255. https://doi.org/10.1016/j.cognition.2013.07.002

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass Correlations: Uses in Assessing Rater Reliability. *Psychological Bulletin*, *86*(2), 420-428.

Thériault, R. (2023). *rempsyc: Convenience Functions for Psychology*. https://rempsyc.remi-theriault.com

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498-505. https://doi.org/10.3758/BF03192720

Ushey, K., & Wickham, H. (2023). *renv: Project environments*. https://CRAN.R-project.org/package=renv

Wickham, H. (2016). *ggplot: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., & Müller, K. (2021). *dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr

William Revelle. (2023). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University. https://CRAN.R-project.org/package=psych