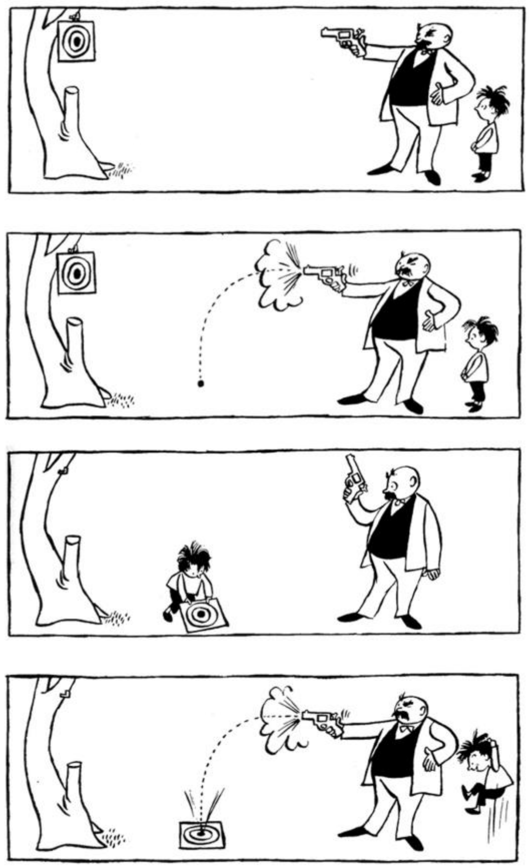
Appendix A. Example comic strip



Artwork by cartoon artist e.o. plauen (Erich Ohser). Works by e.o. plauen are in the public domain.

Appendix B. Example narrative script

1.1 There are a man and a child/a father and a son.
1.2 A bull's eye is hanging in the tree.
1.3 The father has a gun.
1.4 They stand at a short distance from the tree.
1.5 The father is aiming at the bull's eye.
1.6 The boy is watching.

2.1 The father shoots.
2.2 The bullet doesn't go straight.
2.3 The bullet hits the ground.
2.4 It lands in between the tree and the father.
2.5 The boy is watching.

3.1 The boy has an idea.
3.2 He takes the bull's eye out of the tree.
3.3 He puts it on the ground, where the bullet had landed earlier.
3.4 The father looks confused.

4.1 The father shoots another time.
4.2 The bullet is not going straight again.
4.3 This time it hits the bull's eye exactly in the middle.
4.4 The boy jumps in the air.

Appendix C. Gesture frequency and gesture rate per narrative event

In addition to the analyses of gesture rate per 100 words and of the proportion of multimodal events reported in the main paper, here we also report analyses of simple *gesture frequency* (i.e., the number of gestures produced per narration) and of *gesture rate per narrative event* (i.e., dividing the number of gestures a given participant produced by the number of associated narrative events, see Galati & Brennan, 2014), for each condition within each trial separately.

Means and standard deviations per age group and condition for these additional gesture-based measures are reported in Table C1.

Table C1. Means (and *SD*) for gesture frequency and gesture rate per narrative event for each age group and condition. CG = common ground condition, no-CG = no common ground condition.

	Younger		*Older*	
	CG	*No-CG*	*CG*	*No-CG*
Gesture frequency	2.02 (1.39)	4.78 (2.39)	3.01 (2.09)	2.74 (2.18)
Gestures/narr. event	.56 (.42)	.81 (.34)	.60 (.43)	.51 (.43)

To investigate the influence of age and the common ground manipulation on gesture frequency and gesture rate per narrative event, we fitted linear mixed-effect models in R as described in the methods section of the main paper.

Table C2 summarized the results for the models predicting the two dependent measures based on age and common ground manipulation.

Table C2. Linear mixed-effects models for the effects of age and common ground manipulation on gesture frequency and gesture rate per narrative event. Age group = young and Condition = CG[a] are on the intercept. N = 32.[b]

	Gesture frequency				*Gestures/narrative event*			
	β	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*
Intercept	2.02	.50	4.07	< .001	.48	.12	3.94	< .001
Age group$_{old}$	.99	.5	1.62	.12	.13	.14	.94	.35
Condition$_{no\text{-}CG}$[a]	2.76	.36	7.65	< .0001	.33	.06	5.1	< .0001
Age group$_{old}$: Condition$_{no\text{-}CG}$	-3.03	.51	-5.94	< .0001	-.42	.09	-4.61	< .0001

[a] CG = common ground; no-CG = no common ground.

[b] Both models contain random intercepts for participants and items. The model predicting gestures/narrative event includes by-participant random slopes for common ground manipulation.

The absence of a main effect for age group indicates that there was no age-related difference for the two measures in the CG condition. The significant main effect for common ground manipulation indicates that for the younger adults, gesture frequency and gesture rate were higher in the no-CG as opposed to the CG condition. The significant interactions between age group and common ground manipulation indicate that the increase in gesture frequency and rate was only significantly present in the younger but not the older adults.

Individual contrasts confirm this analysis, with younger adults producing significantly more gestures and gesturing at a significantly higher rate in the no-CG as opposed to the CG condition (β = 2.76, *SE* = .37, t = 7.41, p < .0001 and β = .33, SE = .06, t = -5.10, p < .0001, respectively), whereas this difference was not significant for older adults (p > .05). Comparisons further showed that younger and older adults did not differ in the rate at which they gestured in the CG condition for both measures (both p's > .05). However, there was a significant age-related difference in the no-CG condition (for gesture frequency, β = 2.04, *SE* = .81, t = 2.52, p = .02; for gestures/narrative event, β = .29, *SE* = .14, t = 2.09, p = .04), such that younger adults produced significantly more gestures and gestured at a significantly higher rate than older adults.

Appendix D. Cognitive test battery

Here, we provide a more detailed description of the different tasks we used to assess the individual cognitive abilities, including details on task administration and scoring procedure. With the exception of the Operation span task, which was computer-based, all other tasks used to measure cognitive skills were pen-and-paper versions.

Verbal working memory (Verbal WM)

The Operation span task (O-span) is a standard measure of verbal working memory (Turner & Engle, 1989). The Dutch version of the task used here, as well as the scoring procedure, are based on Shao, Roelofs, and Meyer (2012). Participants were required to evaluate the accuracy of 60 simple mathematical operations while remembering unrelated words for later serial recall. The O-span score was calculated as the sum of words that were recalled in the proper order on trials with correct responses to the mathematical problem, the highest possible score being 60. Due to time-out, O-span data could not be collected from one older male participant.

Visual working memory (visual WM)

To assess the visuo-spatial component of visual WM, participants performed the Visual Patterns Test (VPT, Della Sala, Gray, Baddeley, & Wilson, 1997). Participants are briefly presented with visual patterns of increasing complexity, and have to reproduce these patterns. The VPT score is the highest level of complexity at which at least one of three patterns is recalled correctly. Due to time-out, VPT data could not be collected from two older female participants and one older male participant.

To assess the visuo-sequential component of visual WM, participants performed the Corsi Block-Tapping Task (CBT, Corsi, 1972). The task was administered based on the protocol proposed by Kessels, van Zandvoort, Postma, Kappelle, & de Haan (2000). In this test, participants are asked to imitate the experimenter in tapping nine black cubes mounted on a black board in sequences of increasing length, going in steps of two sequences per level. The final score for each participant was calculated as the length of the last correctly repeated sequence multiplied by the number of correctly repeated sequences (i.e. the number of correct trials).

Executive control

Participants performed the Trail Making Test part A and B (TMT A and TMT B) (Parkington & Leiter, 1949) in order to assess their executive control. In part A, participants use a pencil to connect a series of 25 encircled numbers in numerical order. In part B, the subject connects 25 encircled numbers and letters in numerical and alphabetical order, alternating between numbers and letters, requiring the continuous shifting of attention between numbers and letters. The difference between the time to complete part A and part B (TMT B-A) is seen as a measure of inhibition/interference control (isolating the switching component of part B by subtracting the visual search and speed component of part A, see Sanchez-Cubillo, Perianez, Adrover-Roig, Rodriguez-Sanchez, Rios-Lago, Tirapu, & Barcelo, 2009).

Semantic fluency

The animal naming task is a standard measure of semantic fluency (Isaacs & Kennie, 1973). Participants are asked to generate as many unique animal names as possible within 60 seconds. Every unique response is given a point, with repetitions receiving no point.

References

Corsi, P.M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34*, 819B.

Della Sala, S., Gray, C., Baddeley, A., & Wilson, L. (1997). The visual patterns test: A new test of short-term visual recall. Feltham, Suffolk: Thames Valley Test Company.

Isaacs, B., & Kennie, A.T. The set test as an aid to the detection of dementia in old people. *British Journal of Psychiatry, 123* (1973), 467–471.

Kessels, R., van Zandvoort, M., Postma, A., Kappelle, L.J., & de Haan, E. (2000). The Corsi Block-Tapping Task: Standardization and normative data. *Applied Neuropsychology, 7*(4), 252-258.

Parkington, J.E., & Leiter, R.G. (1949). Partington's Pathway Test. *The Psychological Service Center Bulletin, 1,* 9–20.

Sanchez-Cubillo, I., Perianez, J.A., Adrover-Roig, D., Rodriguez-Sanchez, J.M., Rios-Lago, M., Tirapu, J., & Barcelo, F. (2009). Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society, 15*, 438-450.

Shao, Z., Roelofs, A., & Meyer, A. (2012). Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *Quarterly Journal of Experimental Psychology*, *65*, 1927-1944.

Turner, M.L., & Engle, R.W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127-154.

Appendix E. Full model summaries additional analyses

Table E1. Linear mixed-effects models for the effects of age and common ground manipulation on explicit references to common ground and addressee feedback. Age group = young and Condition = CG[a] are on the intercept. N = 32.[b]

	Explicit reference to common ground				*Addressee feedback*			
	β	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*
Intercept	.74	.07	9.96	< .001	.07	.01	10.93	< .001
Age group$_{old}$	-.41	.10	-3.87	< .001	-.02	.01	-1.99	.06
Condition$_{no\text{-}CG}$[a]	-.67	.07	-9.84	< .001	-.02	.01	-2.96	.004
Age group$_{old}$: Condition$_{no\text{-}CG}$	.51	.10	5.33	< .001	-	-	-	-

[a] CG = common ground; no-CG = no common ground.
[b] Both models contain random intercepts for participants and items, but no by-participant random slopes for common ground manipulation.

Table E2. Linear mixed-effects models for the effects of age, common ground manipulation, and addressee feedback on narrative event count and gesture rate per 100 words. Age group = young and Condition = CG[a] are on the intercept. N = 32.[b]

	Narrative events				*Gesture rate per 100 words*			
	β	*SE*	*t*	*p*	*β*	*SE*	*t*	*p*
Intercept	6.43	.85	7.55	.009	3.11	1.97	1.58	.17
Addressee feedback	-10.63	4.15	-2.56	.01	36.16	9.90	3.65	< .001
Age group$_{old}$	-	-	-	-	1.59	1.48	1.07	.29
Condition$_{no\text{-}CG}$[a]	-	-	-	-	3.01	.87	3.48	< .001
Age group$_{old}$: Condition$_{no\text{-}CG}$	-	-	-	-	-4.42	1.19	-3.73	< .001

[a] CG = common ground; no-CG = no common ground.
[b] Both models contain random intercepts for participants and items. The model predicting narrative event count includes by-participant random slopes for the common ground manipulation.

Appendix F. Correlations between cognitive predictors and dependent measures

Tables F1 and F2 list the correlations between dependent variables and cognitive predictors (z-scored) for younger and older adults respectively. Note that we multiplied the inhibitory control task's scores with -1, so that higher scores would represent better performance. In the younger adults, none of the cognitive measures were significantly correlated with the dependent variables. In the older adults, verbal WM was positively correlated with word and narrative event count, such that the higher the verbal WM, the larger the number of words and narrative events.

Table F1. Spearman's rank correlation rho for the dependent measures and cognitive predictors (z-scored). Younger adults.

	Words	*Narrative events*	*Gestures/ 100 words*	*% Multimodal events*
Verbal WM	-0.01	-0.04	0.12	0.1
Visuo-sequential WM	.01	.1	.26	.18
Visuo-spatial WM	-.27	.02	-.15	-.36
Executive control	0.06	0.16	0.15	0.15
Semantic fluency	0.26	0.38	0.23	0.28

***** = p < .001, ** = p < .01, * = p < .05**

Table F2. Spearman's rank correlation rho for the dependent measures and cognitive predictors (z-scored). Older adults.

	Words	*Narrative events*	*Gestures/ 100 words*	*% Multimodal events*
Verbal WM	0.58*	0.59*	-0.31	-0.08
Visuo-sequential WM	.37	.22	.15	.26
Visuo-spatial WM	.47	.48	-.31	-.27
Executive control	0.25	0.3	-0.33	-0.39
Semantic fluency	0.01	-0.18	0.24	0.33

***** = p < 0.001, ** = p < 0.01, * = p < 0.05**