Supplementary Analyses

to Kuhlmann, B.G., & Undorf, M. Is All Metamemory Monitoring Spared? A Dual-Process Examination.

Experiment 1: Old/New Recognition Tests

Table S1 shows the older participants' descriptive statistics for the old/new recognition tests in Experiment 1 (for the younger adults' statistics, see Appendix C of Undorf et al., 2016). As in the main analyses, we focus on read words, but Table 2 includes means from heard words for comparison. An ANOVA with age group (younger adults, older adults) and old/new recognition cycle (1, 2) as between-subjects factors on recognition hits (% old responses) to read words yielded main effects of age group (younger > older), $F(1, 121) = 14.19$, $p < .001$, $\eta_p^2 = .11$, and cycle (Cycle 1 < Cycle 2), $F(1, 121) = 16.99$, $p < .001$, $\eta_p^2 = .12$. The interaction was not significant, $F<1$. The same ANOVA on recognition false alarms (% old to new words) yielded only an effect of cycle (Cycle 1 > Cycle 2), $F(1, 121) = 9.22$, $p = .003$, $\eta_p^2 = .071$, both other $F < 1$. Thus, older adults had poorer recognition than younger adults and each age group's recognition performance increased from Cycle 1 to 2. Turning to metamemory monitoring, the same ANOVA on gammas for read words yielded no effects, all $F \leq 1.65$, $p \geq .202$. That is, in contrast to inclusion gammas presented in the main text, monitoring on the old/new recognition test did not increase from Cycle 1 to 2. This potentially reflects reactive effects of the prior inclusion/exclusion testing for those participants who completed old/new recognition in the second cycle. It is again notable that older adults' gammas were descriptively lower than younger adults' (see main text for a discussion). This was also true for gammas for heard words (but again not significant, $p = .123$).

Experiment 2: Comparison of Strategy Users versus Nonusers

In order to examine effects of mediator-based strategy use on *R* monitoring in Experiment 2, we pre-decided (and preregistered) an inclusion criterion of reported mediator-based (interactive imagery or verbal sentences) strategy use on at least 50% of the word-position pairs. Somewhat to our surprise, this led to a rather large number of exclusions of 21 younger (mediator-based strategies used on M = 26.91% trials, SD = 12.70) and 16 older adults (M = 19.38%, SD = 13.02). In the following, we compare excluded strategy *nonusers* to the included strategy *users* (30 per age group) to determine any potential bias that may stem from these exclusions.

Table S2 presents comparisons of participant characteristics. Among the younger adults, strategy nonusers did not differ from users on demographic characteristics (age, education, subjective health), all $t(49) \leq 1.30$, $p \geq .200$. Younger adult nonusers were also comparable to users on the Digit-Symbol Substitution Task, both $t < 1$, but performed marginally poorer on the vocabulary test, $t(49) = 1.79$, $p = .080$. Among the older adults, strategy nonusers also did not significantly differ from users on demographic characteristics, all $t(44) \leq 1.49$, $p \geq .145$, although it is notable that, numerically, nonusers tended to be older and less well educated. Further, older adult nonusers were comparable to users on Digit-Symbol completion performance, $t < 1$, but performed marginally poorer on the incidental associative-memory test of the digit-symbol associations, $t(44) = 1.92$, $p = .061$; as in younger adults, the older nonusers' performance on the vocabulary test was descriptively lower but this was not significant, $t(44) = 1.49$, $p = .144$. That is, although there were no large differences between strategy users and nonusers, strategy nonuse may have partially resulted from difficulty with implementing the instructed strategy due to somewhat poorer cognitive abilities.

Descriptive statistics for nonusers are displayed in Table S3 (statistics for users can be found in Table 5 in the main text). Nonusers had poorer recollection than users in both younger,

$t(49) = 2.09$, $p = .042$, $d = 0.59$, and older adults, $t(44) = 2.64$, $p = .011$, $d = 0.85$. In contrast, familiarity was comparable across strategy users and nonusers, both $t < 1$. That is, in line with our hypotheses, use of mediator-based strategies selectively improved recollection but not familiarity. This does not only replicate the main finding of better recollection in the strategy conditions (users only) compared to the standard condition but further suggests that it is using the mediator-based strategy *per se* that improves recollection, beyond any improvements due to the explicit instructions to strategy-condition participants that position would be relevant for the upcoming memory test. Although there were no significant differences between strategy users and nonusers concerning familiarity, it is notable that the strategy users' F estimates were numerically lower than nonusers', further corroborating that the high levels of R in the strategy users may undermine F estimation (but see main text for further discussion). Overall metamemory monitoring, as indexed by inclusion gammas (see main text for discussion), did not differ between users and nonusers in either age group, both $t < 1$.

Estimates of R and F by the six JOL levels for strategy nonusers are displayed in Figure S1; for visual comparison, the estimates for users are repeated from Figure 2 in the main text. We performed the same model comparisons as for the within-age-group comparisons presented in the main text but focused on the strategy condition only and instead compared strategy users and nonusers as conditions. The model descriptions and their AIC values are presented in Table S4. For younger adults, the best fitting model was Model W7, which assumed no F increases (i.e., zero slopes) but linear R increases (i.e., slopes above 0). The data were not supportive of models restricting R slopes to 0 (Models W6 & W8), ΔAICs = 34.36 and 30.79, indicating that both younger adult strategy nonusers' and users' JOLs were predictive of R. Crucially, the data did not support models equating slopes between strategy users and nonusers, neither of R, ΔAIC = 3.61

(Model W3), nor of *F*, ΔAIC = 18.70 (Model W4).[1] As evident in Figure S1, younger adult nonusers' JOLs were slightly less predictive (i.e., smaller slopes across JOL levels) of *R* than users'. In contrast, younger adult nonusers' JOLs were more predictive (i.e., steeper increases) of *F* than users'. As presented in the main text, strategy users' JOLs were *not* predictive of *F* (i.e., zero slopes). The best fitting Model W7 also suggests nonpredictiveness of *F* over strategy users and nonusers combined. As in the main analyses, we evaluated four additional models that tested *R* and *F* slopes against 0 for younger adult strategy users and nonusers separately. Among these, the model equating *F* slopes to zero in the strategy users was the best fitting model, AIC = 3,106.66. In comparison, the model equating younger adult nonusers' *F* slopes to zero did not fare much worse, ΔAIC = 0.59. This suggests that although there were differences in *F* monitoring between younger-adult strategy users and nonusers, neither group's JOLs achieved a good prediction of *F*. Note that nonusers' *R* was still high and above Yonelinas (2002) suggested cut-off for ceiling (.60) such that biased estimates of *F* may be an issue in both younger strategy-condition subgroups. In contrast, and in line with the results on Model W6 presented earlier, all younger participants' JOLs predicted *R* above zero, ΔAIC = 25.18 (users) and ΔAIC = 9.77 (nonusers).

For older adults, the best fitting model was Model W3, which assumed linear *R* slopes that did not differ between strategy users and nonusers. Although there was also good support for the parsimonious Model W5 equating slopes of both *R* and *F* across strategy users and nonusers, ΔAIC = 0.68, Model W4 focusing on comparing F increases, specifically, revealed monitoring differences by strategy use, ΔAIC = 5.90. As evident in Figure S1, *F* increased with JOL in older adult strategy nonusers but not in users. This is also evident by the poor support of model W7

[1] These models additionally assumed linear increases in *R* and *F*. Although Model W2, which makes this linear assumption alone was not well supported by the data, ΔAIC = 3.43, the linear assumption appears to be compatible with the present data as several well-fitting models also make this assumption (e.g., Model W7). Rather, the poor fit of W2 is attributable to its lack of parsimony.

testing *F* slopes against 0 across users and nonusers combined, ΔAIC = 3.06. As reported in the main text, this model was supported by the data of older strategy users. Likewise, in the four additional comparisons testing *R* and *F* slopes against zero separately by strategy use, the model setting *F* slopes to 0 in older-adult strategy users fit best, AIC = 4,043.40. In contrast, older-adult nonusers' *F* slopes were above zero, ΔAIC = 2.04. JOLs were predictive of *R* for both users, ΔAIC = 8.00, and nonusers, ΔAIC = 2.04 (see also Model W7 testing this combined across all participants, ΔAIC = 7.95).

All in all, these comparisons suggest that the included strategy users were somewhat selected in terms of cognitive abilities (incidental associative memory [older adults], vocabulary [younger adults]). They expectedly performed poorer on recollection. Crucially, they also performed differently in terms of metamemory monitoring: In younger adults, nonusers' JOLs were less predictive of *R* but more predictive of *F*. In older adults, nonusers' JOLs were comparably predictive of *R* but more predictive of *F*. The differences in *R* monitoring between younger adult strategy users and nonusers are interesting as they possibly suggest a strategy-benefit to *R* monitoring. However, as this was not supported in the main analysis comparing younger strategy users to those with standard instructions, the nonusers' poorer monitoring might rather be due to the cognitive-ability differences between users and nonusers. Alternatively, the poorer performance in strategy nonusers may also reflect generally lower motivation of these participants. The nonusers' better monitoring of *F* in both age groups further suggests that *F* estimates in the strategy users might have been biased due to their high levels of *R* (particularly in younger users) and/or that the mediator-based strategy focused participants on recollection-related cues during study, as discussed in detail in the main text.

Table S1

Descriptive Statistics for Old/New Recognition Tests in Experiment 1

	Cycle					
	1			2		
	Read	Heard	New	Read	Heard	New
% old	74.19 (16.09)	71.77 (11.34)	12.18 (8.82)	84.07 (13.98)	82.08 (19.50)	6.22 (5.77)
JOL	40.93 (14.35)	42.14 (13.45)	–	54.06 (21.79)	52.35 (20.11)	–
G	.24*** (.27)	.35*** (.29)	–	.32*** (.37)	.24* (.50)	–

Note. Values in parentheses are standard deviations. JOL = judgment of learning; % old = percentage of old responses; *G* = mean within-subjects gamma correlation between JOLs and recognition performance (asterisks refer to one-sample *t* tests against 0).
* $p < .05$. *** $p < .001$.

Table S2

Participant Characteristics for Strategy Users versus Nonusers from the Strategy Conditions in Experiment 2

Variable	Younger Adults		Older Adults	
	Users	Nonusers	Users	Nonusers
Age	22.33 (2.80)	21.38 (2.24)	68.90 (5.20)	71.50 (7.10)
Years of formal education	15.73 (2.27)	15.05 (1.67)	15.12 (3.36)	13.59 (3.16)
Subjective health rating*	3.30 (0.53)	3.24 (0.63)	3.07 (0.58)	3.00 (0.52)
Digit-Symbol completion*	63.10 (10.74)	62.33 (10.04)	41.77 (9.56)	39.89 (16.37)
Digit-Symbol memory*	7.87 (1.83)	7.90 (1.22)	5.63 (2.74)	4.06 (2.41)
Vocabulary (% correct)*	75.33 (11.89)	69.13 (13.35)	83.00 (8.16)	77.81 (15.60)

Note. Displayed values are mean (standard deviation). Subjective health rating was made on a 4-point scale (1 = *bad*; 4 = *excellent*). For the Digit-Symbol task (Wechsler, 1981), completion is the number of correctly copied symbols in the allotted 90s (only valid if filled out in order as instructed) and memory the number of digit-symbol associations freely recalled afterwards (maximum of 9). Vocabulary is the percentage correct on the German SASKA test (Riegel, 1967).

Table S3

Descriptive Statistics for Nonusers from the Strategy Conditions in Experiment 2

	Test trial		
Condition/Measure	Inclusion	Exclusion	New
Strategy instructions			
% yes	OA: 66.67 (15.52) YA: 82.67 (11.14)	OA: 41.32 (16.13) YA: 21.30 (15.89)	OA: 13.02 (10.03) YA: 5.42 (4.69)
JOL	OA: 36.21 (16.20) YA: 44.78 (8.00)	OA: 35.71 (16.23) YA: 45.30 (6.86)	–
G	OA: .19 ** (.22) YA: .14 $^{+}$ (.35)	OA. -.01 (.22) YA: .17* (.35)	–

	F	*R*
Strategy instructions	OA: .56 (.16) YA: .53 (.18)	OA: .25 (.19) YA: .61 (.24)

Note. Values in parentheses are standard deviations. OA = older adults; YA = younger adults; % yes = percentage of “yes” responses; JOL = judgment of learning; *G* = mean within-subjects gamma correlation between JOLs and recognition performance (asterisks refer to one-sample *t* tests against 0); *F* = probability of automatic influences of memory (familiarity); *R* = probability of recollection.
$^{+}$ $p < .10$. * $p < .05$. *** $p < .001$

Table S4

Regression Model Descriptions and Akaike's Information Criterion (AIC) Values for the Within-Age-Group Analyses of by Strategy Users Versus Nonusers from the Strategy Conditions in Experiment 2

Model	Parameter constraints	Tested hypothesis	YA	OA
W1	none	None: Estimates free to vary across JOL levels and conditions (x_{Rkn}, x_{Fkn}).	3,119.93	4,064.33
W2	$x_{Rk\bullet}, x_{Fk\bullet}$	Linear slopes (i.e., equal across the six JOL levels) of *R* and *F*?	3,108.66	4,045.41
W3	$x^{a}_{Rk\bullet}, x_{Fk\bullet}$	Linear + comparable across conditions slopes of *R* and linear slopes of *F*?	3,108.85	4,043.48
W4	$x_{Rk\bullet}, x^{a}_{Fk\bullet}$	Linear slopes of *R* and linear + comparable across condition slopes of *F*?	3,123.93	4,049.38
W5	$x^{a}_{Rk\bullet}, x^{a}_{Fk\bullet}$	Linear + comparable slopes of *R* and *F*?	3,107.22	4,044.16
W6	$x^{b}_{Rk\bullet}, x_{Fk\bullet}$	Linear slopes of *F* and zero slopes (i.e., no change) of *R* in both conditions?	3,139.60	4,051.43
W7	$x_{Rk\bullet}, x^{b}_{Fk\bullet}$	Linear increases of *R* and zero slopes (i.e., no change) of *F* in both conditions?	3,105.24	4,046.54
W8	$x^{b}_{Rk\bullet}, x^{b}_{Fk\bullet}$	Zero slopes (i.e., no change) of *R* and *F* in both conditions?	3,136.03	4,052.38

Note. x_{Rkn} = slope parameter capturing change (≥ 0) of recollection (*R*) from Judgment of Learning (JOL) level *n* to *n*+1 in condition *k*; x_{Fkn} = slope parameter capturing change (≥ 0) of familiarity (*F*) from JOL level n to *n*+1 in condition *k*; $x_{Rk\bullet}$ = linear slope of *R* (i.e., x_{Rkn} equated across the six JOL levels) in condition *k*; $x_{Fk\bullet}$ = linear slope of *F*(i.e., x_{Fkn} equated across the six JOL levels) in condition *k*. . Slopes reflect the predictive value of JOLs for R and F, respectively (i.e., relative monitoring accuracy).

[a] Parameters were equated across subgroups (strategy users and nonusers).

[b] Parameters were constrained to zero.

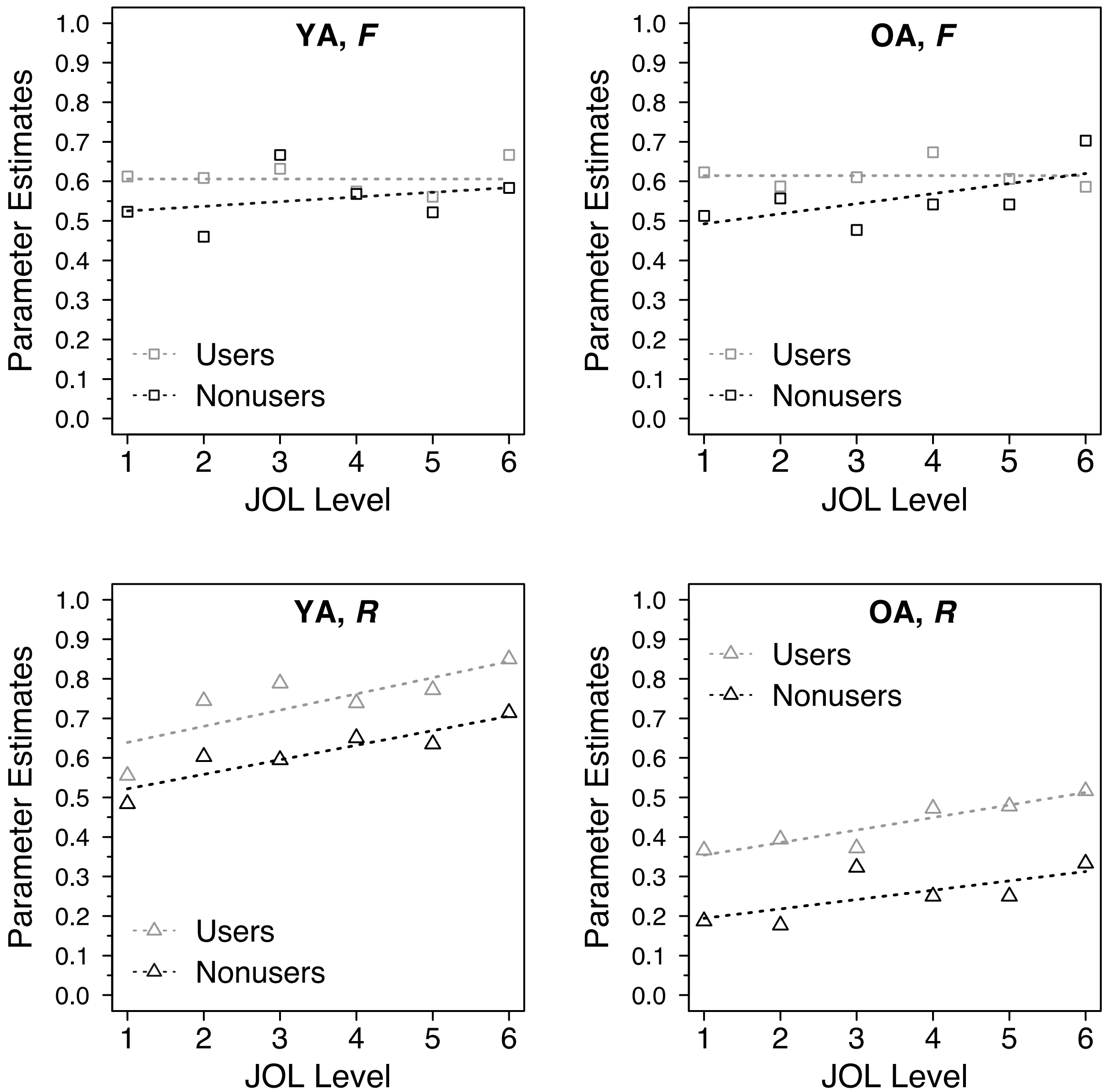

Figure S1. Parameter estimates for familiarity (*F*, top row) and recollection (*R*, bottom row) for younger-adult (left column) and older-adult (right column) strategy users and nonusers in the strategy conditions from Experiment 2 across six judgment of learning (JOL) levels (see main text).