Supplementary Materials to accompany:

The Role of Rehearsal and Reminding in the Recall of Categorized Word Lists

Geoff Ward[1] and Lydia Tan[2]

[1]University of Essex

[2]University of East London

Correspondence to:

Geoff Ward,   Lydia Tan,
Department of Psychology,   School of Psychology,
University of Essex,   University of East London,
Wivenhoe Park,   Water Lane,
Colchester,   London,
CO4 3SQ,   E15 4LZ,
UK   UK

Telephone: +44 1206 873799
Fax: +44 1206 873590
E-mail: gdward@essex.ac.uk   l.tan@uel.ac.uk

Running header: Rehearsal and reminding in categorized recall

Supplementary Materials

In these Supplementary Materials, we provide more detailed figures and additional analyses for Experiments 2 and 3 of Ward, G. and Tan, L. "The Role of Rehearsal and Reminding in the Recall of Categorized Word Lists". These include: (a) more detailed figures and analyses summarising the probability of rehearsal within a RS for each list type, and (b) the probability of final cued recall following immediate tests of free recall.

SM1. Experiment 2: Blocked Presentation

SM1.1 Detailed analyses of the Probability of Rehearsal for each list type

Figure S1 shows the probability of rehearsing a word within a RS for each task and list type in Experiment 2. The top two panels show the rehearsal data from cued recall, the bottom two panels show the rehearsal data from free recall. The left-hand panels show the effects of varying the number of categories, the right-hand panels show the effects of varying the number of exemplars.

--Figure S1 about here--

A 2 (task: cued recall and free recall) x 3 (list type) x 2 (semantic category: same or different to just-presented word) x 4 Number of Intervening RSs (1, 2-3, 4-6, 7+) mixed ANOVA was performed, comparing the effects of increasing the number of categories (reported in Appendix S1), and a 2 (task: cued recall and free recall) x 3 (list type) x 2 (semantic category: same or different to just-presented word) x 2 Number of Intervening RSs (1, 2-3) mixed ANOVA was performed, comparing the effects of increasing the number of exemplars (reported in Appendix S2). In both analyses, there were no significant main effects of task,

nor were there any significant interactions involving task, showing that the probability of rehearsing an item was similar in free and cued recall with blocked presentations.

In both analyses, the probability of rehearsing a word within a RS increased significantly if it was from the same semantic category as the just-presented word, increased significantly if it had been rehearsed more recently (with fewer intervening RSs since it was last rehearsed), and decreased significantly with increasing list length, whether that was due to increasing the numbers of categories (Appendix S1.3) or increasing the numbers of exemplars per category (Appendix S1.4) in a list. It is also clear from Figure S1 that the effect of increasing the number of exemplars per category (right-hand panels) is more important than the effect of increasing the number of categories (left-hand panels).

Although the probability of rehearsing an item within a RS was strongly affected by the number of intervening RSs since the word had last been rehearsed, the more detailed analyses shown in Figures S2 and S3 for each list type reveals that the probability of rehearsing a word was also affected by the number of prior rehearsals, and the effect of recency was reduced somewhat for words that had been previously rehearsed many times. These analyses for each list type replicate the findings of Experiment 1 (see Figure 4, main manuscript).

--Figures S2 and S3 about here--

SM1.2 Analysis of the Final Cued Recall data

A test of cued recall had originally followed the free recall tests in Tulving and Pearlstone's (1966) study, and had originally helped confirm that items that were not accessible through a

test of immediate free recall could nevertheless be later retrieved in the immediately

subsequent test of cued recall. The test in the original study was highly diagnostic because

each participant received one list and the cued recall test followed immediately after the free

recall test. Our test of final cued recall was, unfortunately, rather less diagnostic, because our

participants were presented and tested with 15 free recall trials prior to the final cued recall.

Specifically, at the end of the experiment, participants in the free recall conditions of

Experiment 2 undertook a final test of cued recall in which they received 30 category cue

labels and had to recall as many of the corresponding category exemplars as possible. For

each participant, two of the category labels were randomly sampled from the categories used

in each of the 15 different lists, such that the five different list types were each sampled six

times.

The final cued recall data (larger grey circles) are shown in Figure S4. The Figure shows that

cued recall decreased non-significantly with increasing number of categories, $F(2,46) = 3.13$,

$MSE = .007$, $\eta_p^2 = .120$, $p = .053$, and decreased significantly with increasing number of

exemplars per category, $F(2,46) = 3.77$, $MSE = .015$, $\eta_p^2 = .141$, $p = .030$.

--Figures S4 about here--

Many of the words recalled in the final test of cued recall had earlier been recalled in the

immediate free recall tests that followed the list presentations (plotted in Figure S4 as smaller

black circles); whereas some of words recalled in the final test of cued recall had not

previously been recalled in the immediate free recall test (plotted in Figure S4 as smaller

unfilled circles). Considering the effects of increasing number of categories, a 2 (prior recall:

recalled, unrecalled) x 3 (number of categories: 3x8, 6x8, 8x8) within-subjects ANOVA

revealed a highly significant effect of whether or not the word had previously been recalled,

$F(1,23) = 419.0$, $MSE = .028$, $\eta_p^2 = .948$, $p < .001$, a significant effect of number of categories, $F(2,46) = 6.328$, $MSE = .011$, $\eta_p^2 = .215$, $p = .004$, and a non-significant interaction, $F(2,46) = 0.87$, $MSE = .012$, $\eta_p^2 = .037$, $p = .424$. Considering the effects of increasing number of exemplars, a 2 (prior recall: recalled, unrecalled) x 3 (number of exemplars: 8x3, 8x6, 8x8) within-subjects ANOVA revealed a highly significant effect of whether or not the word had previously been recalled, $F(1,23) = 478.3$, $MSE = .022$, $\eta_p^2 = .954$, $p < .001$, a non-significant effect of number of exemplars, $F(2,46) = 0.69$, $MSE = .026$, $\eta_p^2 = .029$, $p = .506$, and a non-significant interaction, $F(2,46) = 0.18$, $MSE = .030$ $\eta_p^2 = .008$, $p = .835$.

SM 1.3. Detailed analyses of clustering and output order in Experiment 2 using Lag LRS

In the blocked design of Experiment 2, the participants rarely rehearsed the study items to much later list positions and so the CRP-Lag analyses are very similar to the CRP-Lag LRS analyses. For completeness, the CRP-Lag LRS analysis is presented here. Figure S5 shows the Lag LRS-CRP plots for the five free recall list types in Experiment 2. A 3 x 9 within-subjects ANOVA was performed on the CRPs for the free recall data (20 participants contributed to all cells) with the number of categories (3, 6, and 8) and Lag LRS (Lag LRS <-10, -10 to -6, -5 to -3, -2 to -1, 0, +1 to +2, +3 to +5, +6 to +10, and >+10) as the within-subjects factors. There was a highly significant main effect of number of categories, $F(2, 38) = 17.97$, $MSE = .002$, $\eta_p^2 = .486$, $p < .001$, reflecting the higher probabilities of making transitions when there were only 3 categories, there was a significant main effect of Lag LRS, $F(8, 152) = 77.70$, $MSE = .005$, $\eta_p^2 = .804$, $p < .001$, reflecting the increased transitions to words that were last rehearsed at similar list positions, reaching a maximum CRP value at Lag LRS = 0, and a significant interaction between transition category and Lag LRS, $F(16,$

304) = 1.87, *MSE* = .002, η_p^2 = .090, *p* = .022, reflecting particularly high CRP values for the 3x8 category at Lag LRS = 0.

--Figures S5 about here--

A second analysis was performed examining the effects of increasing number of exemplars within a category on the Lag LRS-CRP plots. A 3 x 9 within-subjects ANOVA was conducted on the CRPs for the free recall data (19 participants contributed to all cells), with the number of exemplars (3, 6, and 8) and Lag (Lags <-10, -10 to -6, -5 to -3, -2 to -1, 0, +1 to +2, +3 to +5, +6 to +10, and >+10) as within-subjects factors. This revealed a highly significant main effect of number of exemplars, *F*(2, 36) 35.20, *MSE* = .001, η_p^2 = .662, *p* < .001, reflecting the higher CRPs of making transitions when there were fewer exemplars; there was a significant main effect of Lag LRS, *F*(8, 144) = 176.6, *MSE* = .003, η_p^2 = .908, *p* < .001, reflecting again the higher probabilities of making Lag LRS = 0 transitions. The interaction was again significant, *F*(16, 288) = 8.80, *MSE* = .002, η_p^2 = .328, *p* < .001, reflecting particularly high CRP values for the 8x3 category at Lag LRS = 0.

 In summary, there were large and significant Lag LRS effects when the free recall output orders were analysed irrespective of semantic category.

Figure S6 shows the Lag LRS-CRP plots for the same-category and different-category transitions for each list type in the free recall conditions of Experiment 2. Unfortunately, when the Lag LRS-CRP data were further conditionalized by same- or different-category transitions, few participants contributed to all the cells, even when only a restricted range of Lag LRSs were used. Thus, a 2 (transition type: same – or different-category transition) x 3 (number of categories: 3, 6, and 8) x 5 (Lag LRS: -5 to-3, -2 to -1, 0, +1 to +2, +3 to +5)

within-subjects ANOVA with 8 participants showed a significant main effect of transition type, $F(1, 7) = 133.4$, $MSE = .008$, $\eta^2 = .950$, $p < .001$, reflecting the higher CRPs of making same-category transitions; there was a significant main effect of Lag LRS, $F(4, 28) = 12.28$, $MSE = .008$, $\eta^2 = .637$, $p < .001$, reflecting again the higher probabilities of making Lag LRS $= 0$ transitions, and a significant interaction between transition type and $F(4, 28) = 12.12$, $MSE = .006$, $\eta^2 = .634$, $p < .001$, reflecting far greater effects of Lag LRS with the same-category transitions. All other main effects and interactions were non-significant (all Fs<2.6, ps $> .10$).

Similarly, a 2 (transition type: same – or different-category transition) x 3 (number of exemplars: 3, 6, and 8) x 5 (Lag LRS: -5 to-3, -2 to -1, 0, +1 to +2, +3 to +5) within-subjects ANOVA with 6 participants showed a significant main effect of transition type, $F(1, 5) = 104.3$, $MSE = .022$, $\eta^2 = .954$, $p < .001$, reflecting the higher CRPs of making same-category transitions; there was a significant main effect of Lag LRS, $F(4, 20) = 6.12$, $MSE = .017$, $\eta^2 = .550$, $p = .002$, reflecting again the higher probabilities of making Lag LRS $= 0$ transitions, and there was a significant main effect of number of exemplars, $F(2, 10) = 20.07$, $MSE = .011$, $\eta^2 = .801$, $p < .001$, reflecting higher CRPs for the list type with the fewest number of exemplars (8x3). There was a significant interaction between transition type and Lag LRS, $F(4, 20) = 7.23$, $MSE = .013$, $\eta^2 = .591$, $p < .001$, reflecting far greater effects of Lag LRS with the same-category transitions. There was also a significant interaction between transition type and number of exemplars, $F(2, 10) = 15.48$, $MSE = .009$, $\eta^2 = .756$, $p = .001$, reflecting far greater effects of number of category for the sae-category transitions. The remaining interactions were non-significant (all Fs < 1.4, all ps $> .2$).

--Figures S6 about here--

SM2. Experiment 3: Randomized Presentation

SM2.1 Detailed analyses of the Probability of Rehearsal for each list type

Figure S7 shows the probability of rehearsing a word within a RS for each list type in the free recall and cued recall trials in Experiment 3. The top two panels show the rehearsal data from cued recall; the bottom two panels show the rehearsal data from free recall. The left-hand panels show the effects of varying the number of categories; the right-hand panels show the effects of varying the number of exemplars.

--Figures S7 about here--

A pair of 2 (task: cued recall and free recall) x 3 (list type) x 2 (semantic category: same or different to just-presented word) x 4 Number of Intervening RSs (1, 2-3, 4-6, 7+) mixed ANOVAs were performed on the probability of rehearsing a word in a RS, comparing the effects of increasing the number of categories (Appendix S2.1) and comparing the effects of increasing the number of exemplars per category (Appendix S2.2). In both analyses, there was no significant main effect of task, nor were there any significant interactions involving task in either analysis, showing that the probability of rehearsing an item was similar in free and cued recall with randomized presentations. The probability of rehearsing a word within a RS increased significantly if it is from the same semantic category as the just-presented word, increased significantly if the word has been rehearsed recently (with fewer intervening RSs since it was last rehearsed), and the effects of number of intervening RSs were greater with

the same-category rehearsals. In Experiment 3, the probability of rehearsal did not differ significantly with increasing the number of categories (left-hand panels) but was greatly reduced by increasing numbers of exemplars per category (right-hand panels). Indeed, the probability of rehearsing an item was much higher in the 8x3 compared with 8x6 and 8x8 list types, this difference was greatest when rehearsing more recent different-category words and when rehearsing more distant same-category words.

The more detailed analyses shown in Figures S8 and S9 shows that, like Experiments 1 and 2, the probability of rehearsing a word for all list types and both tasks, was affected by the number of prior rehearsals, the recency of the most recent rehearsal, and whether or not the just-presented item was from the same category as the rehearsed word.

--Figures S8 and S9 about here--

SM2.2 Analysis of the Final Cued Recall data

Participants in the free recall conditions of Experiment 3 undertook a final test of cued recall in which they received 30 category cue labels and had to recall as many of the category exemplars as possible. As in Experiment 2, two of the labels came from categories in each of the 15 different list types that were randomly selected with the constraint that for each participant, each of the five list types was sampled six times. The final cued recall data are shown in Figure S10. The Figure shows that cued recall decreased significantly with increasing number of categories, $F(2,46) = 9.18$, $MSE = .005$, $\eta_p^2 = .285$, $p < .001$, and decreased non-significantly with increasing number of exemplars per category, $F(2,46) = 3.09$, $MSE = .013$, $\eta_p^2 = .118$, $p = .055$.

--Figures S10 about here--

As for Experiment 2, a 2 (prior recall: recalled, unrecalled) x 3 (number of categories: 3x8, 6x8, 8x8) within-subjects ANOVA revealed a highly significant effect of whether the word had previously been recalled, $F(1,23) = 1101.7$, $MSE = .013$, $\eta_p^2 = .980$, $p < .001$, a significant effect of number of categories, $F(2,46) = 10.16$, $MSE = .008$, $\eta_p^2 = .306$, $p < .001$, and a non-significant interaction, $F(2,46) = 1.15$, $MSE = .011$, $\eta_p^2 = .048$, $p = .325$.

Considering the effects of increasing number of exemplars, a 2 (prior recall: recalled, unrecalled) x 3 (number of exemplars: 8x3, 8x6, 8x8) within-subjects ANOVA revealed a highly significant effect of whether the word had previously been recalled, $F(1,22) = 436.8$, $MSE = .028$, $\eta_p^2 = .952$, $p < .001$, a non-significant effect of number of categories, $F(2,44) = 2.15$, $MSE = .022$, $\eta_p^2 = .089$, $p = .128$, and a non-significant interaction, $F(2,44) = 3.12$, $MSE = .020$ $\eta_p^2 = .124$, $p = .054$.

Figure S1: Data from Experiment 2. The effect of list length on the Probability of Rehearsing an item within a RS in cued recall (top) and free recall (bottom) when list length was increased by increasing the number of categories (Left) and by increasing the number of exemplars per category (Right).
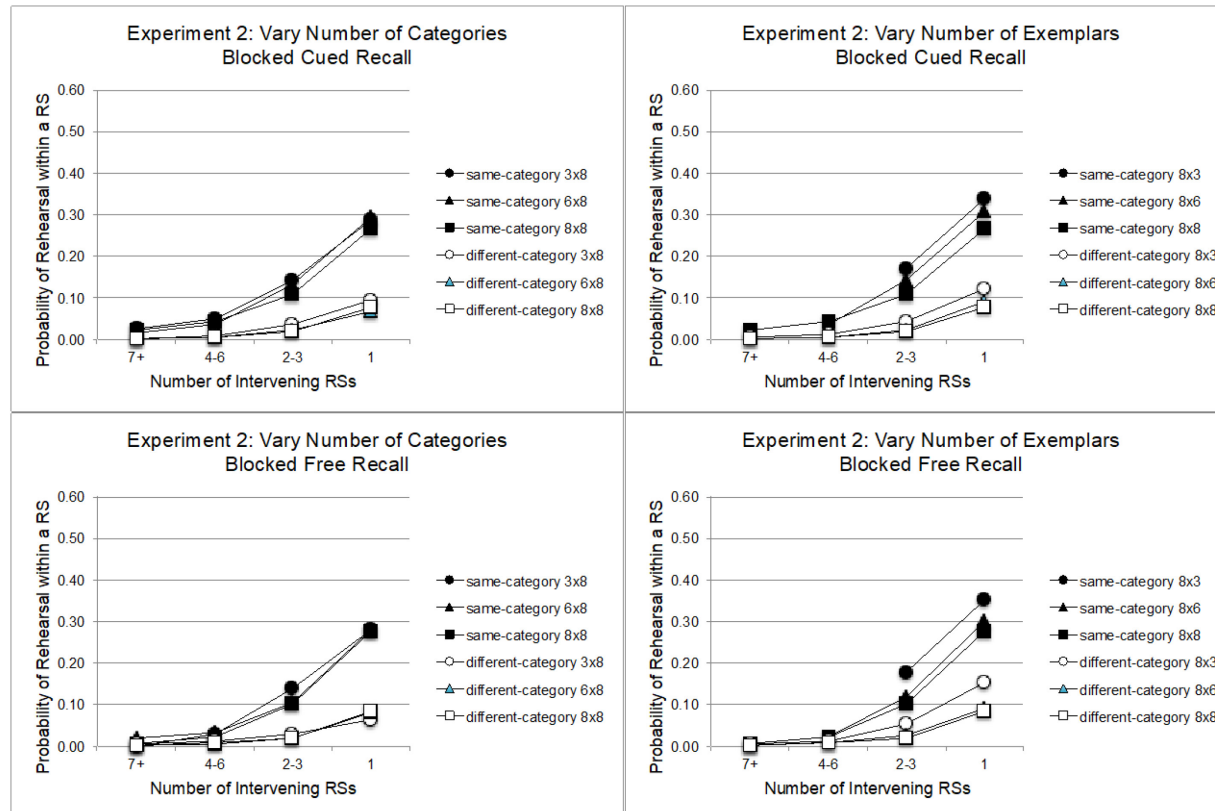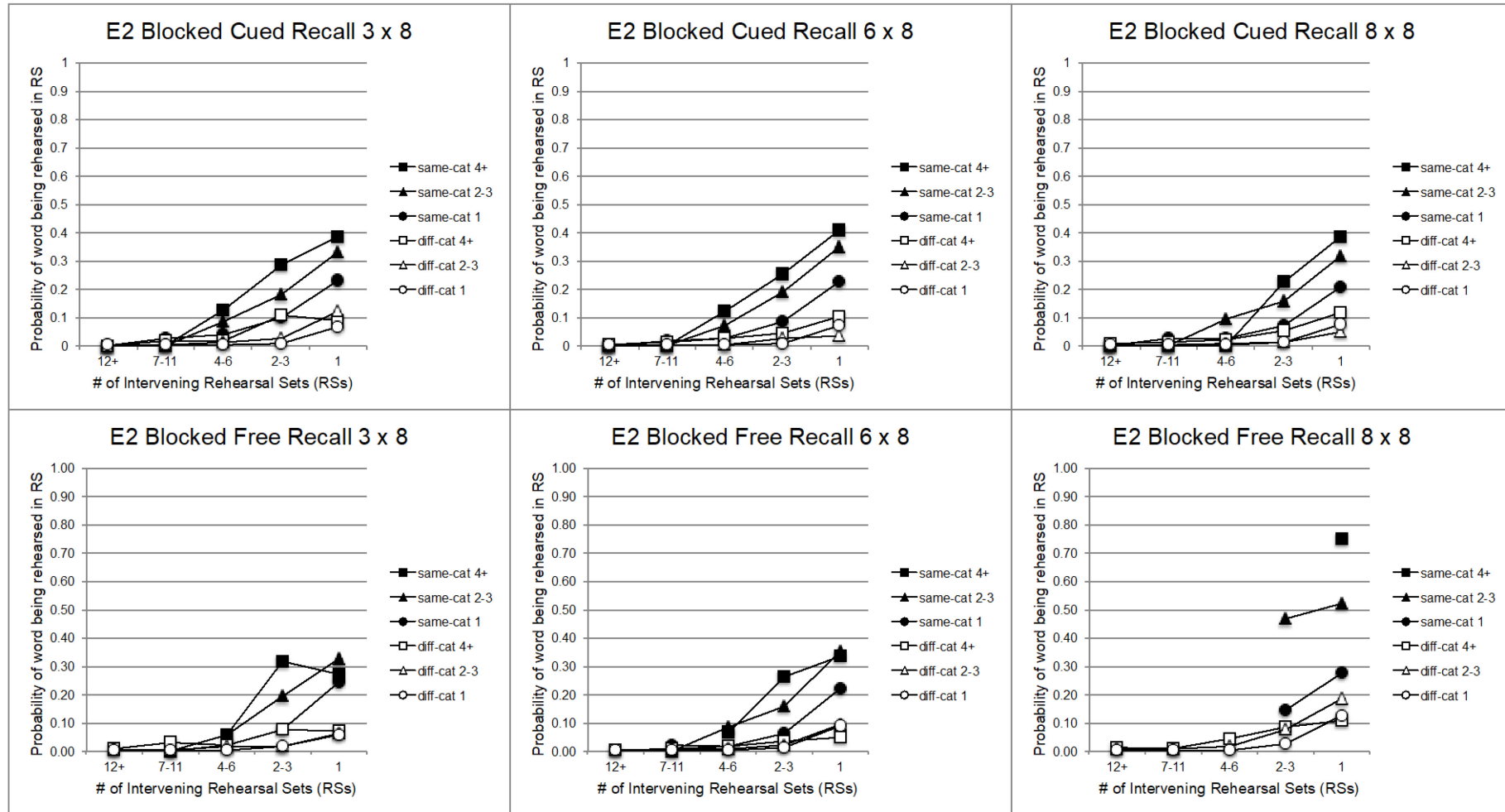
Figure S2: Data from Experiment 2. Probability of Rehearsing an item within a RS in cued recall (top) and free recall (bottom) when the list length was increased by increasing the number of 8-item categories from 3x8 (left) to 6x8 (middle) to 8x8 (right).
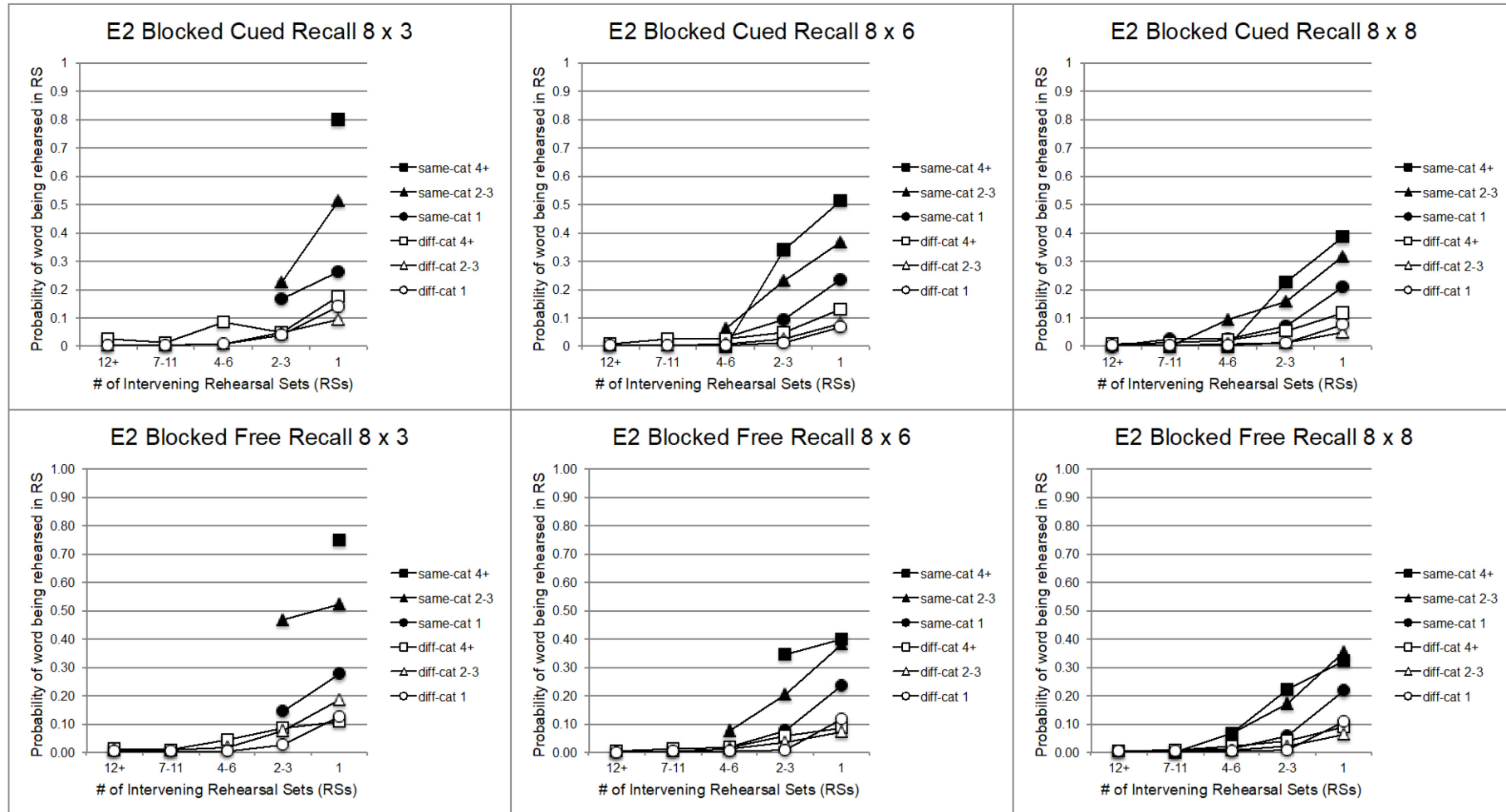
Figure S3: Data from Experiment 2. Probability of Rehearsing an item within a RS in cued recall (top) and free recall (bottom) when the list length was increased by increasing the number of exemplars per category from 8x3 (left) to 8x6 (middle) to 8x8 (right).

Figure S4: Data from Experiment 2. Final Cued recall performance for the participants who undertook the immediate tests of free recall. The large grey circles represent the mean final cued recall performance for each of the list types. The dashed lines represent the means of the subset of recalls for words that had earlier been recalled (filled circles) and that had earlier been unrecalled (open circles) in the immediate test of free recall.
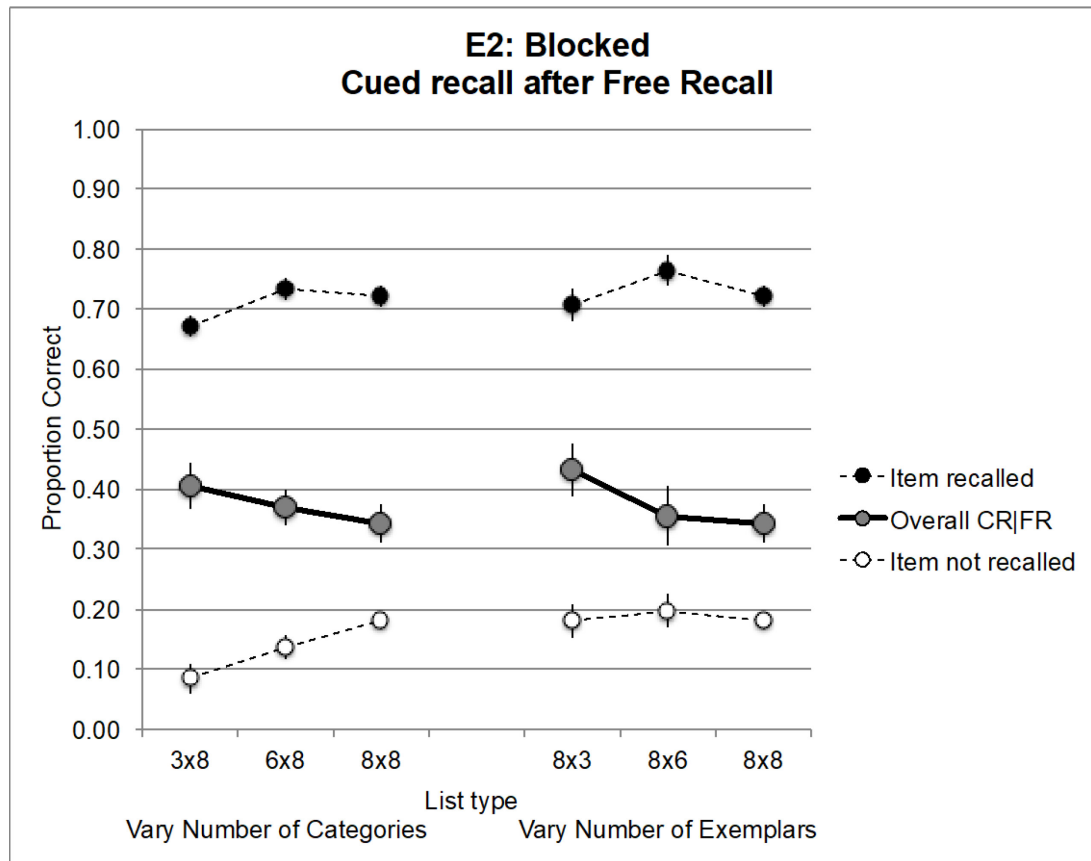
Figure S5: Data from Experiment 2. The Lag LRS – CRP plots for the free recall data, irrespective of the category of transition. The Lag LRS is calculated by subtracting the Last RS of each output word from the Last RS of its predecessor. A Lag LRS = 0 refers to successive recalled words that were co-rehearsed during their Last RS.
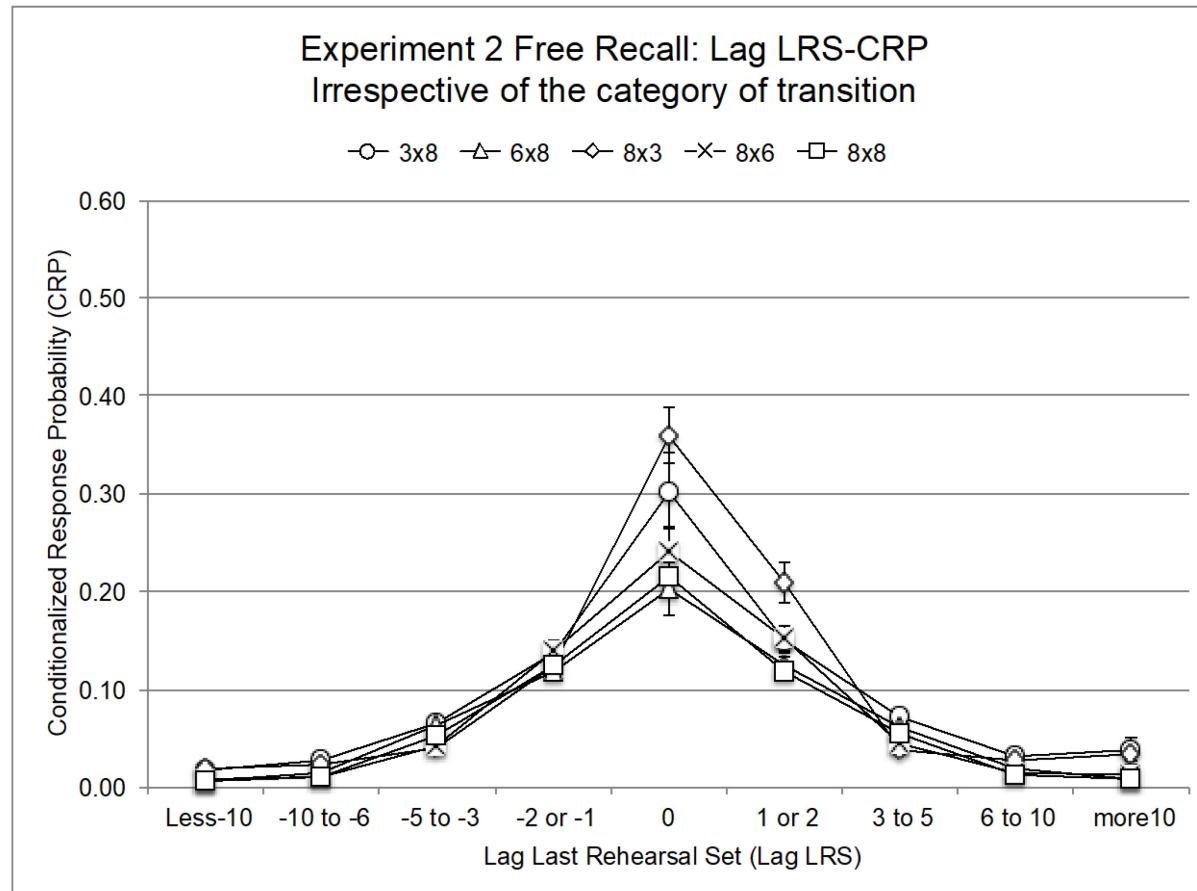
Figure S6: Data from Experiment 2. The Lag LRS – CRP plots for the free recall data, further conditionalized by the category of

transition (same-category or different-category). The Lag LRS is calculated by subtracting the Last RS of each output word from the

Last RS of its predecessor. A Lag LRS = 0 refers to successive recalled words that were co-rehearsed during their Last RS.
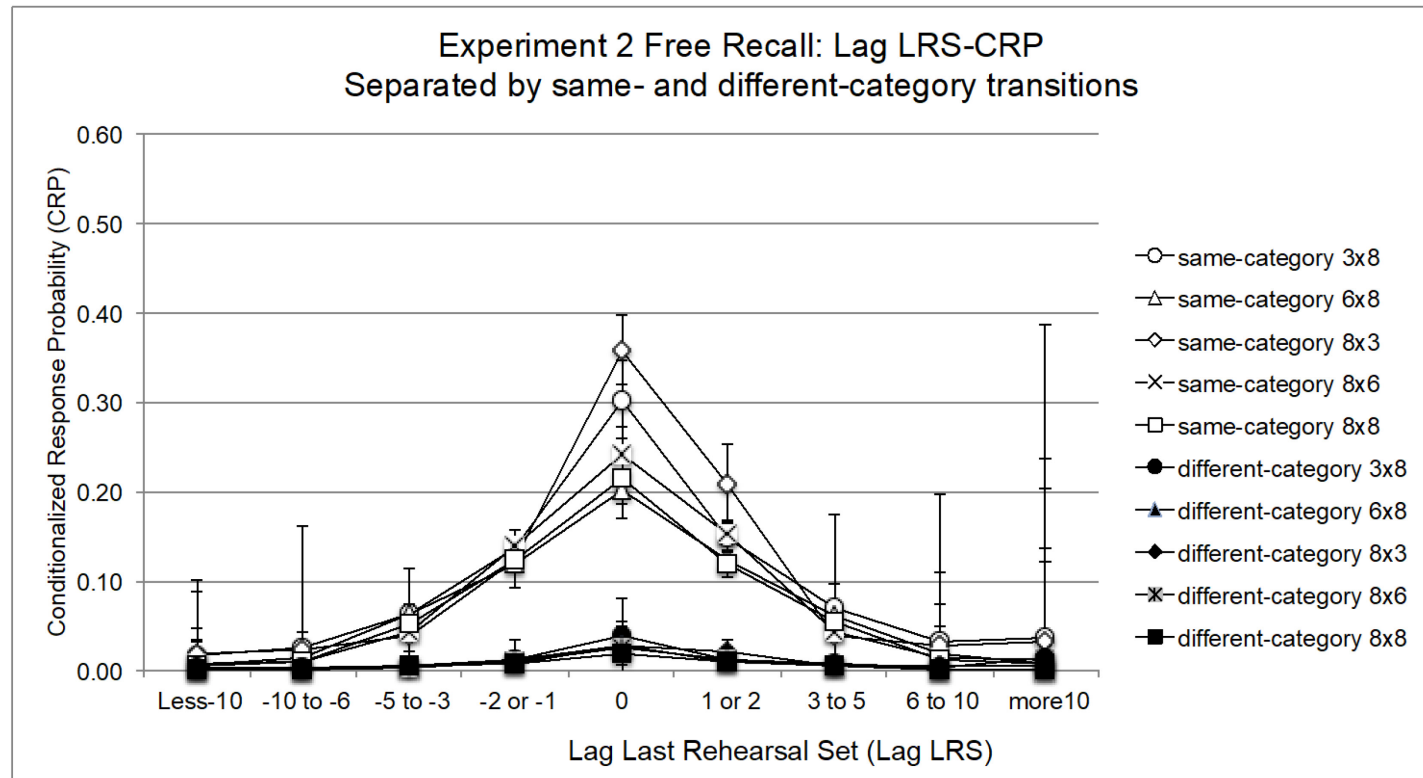
Figure S7: Data from Experiment 3. The effect of list length on the Probability of Rehearsing an item within a RS in cued recall (top) and free recall (bottom) when list length was increased by increasing the number of categories (Left) and by increasing the number of exemplars per category (Right).
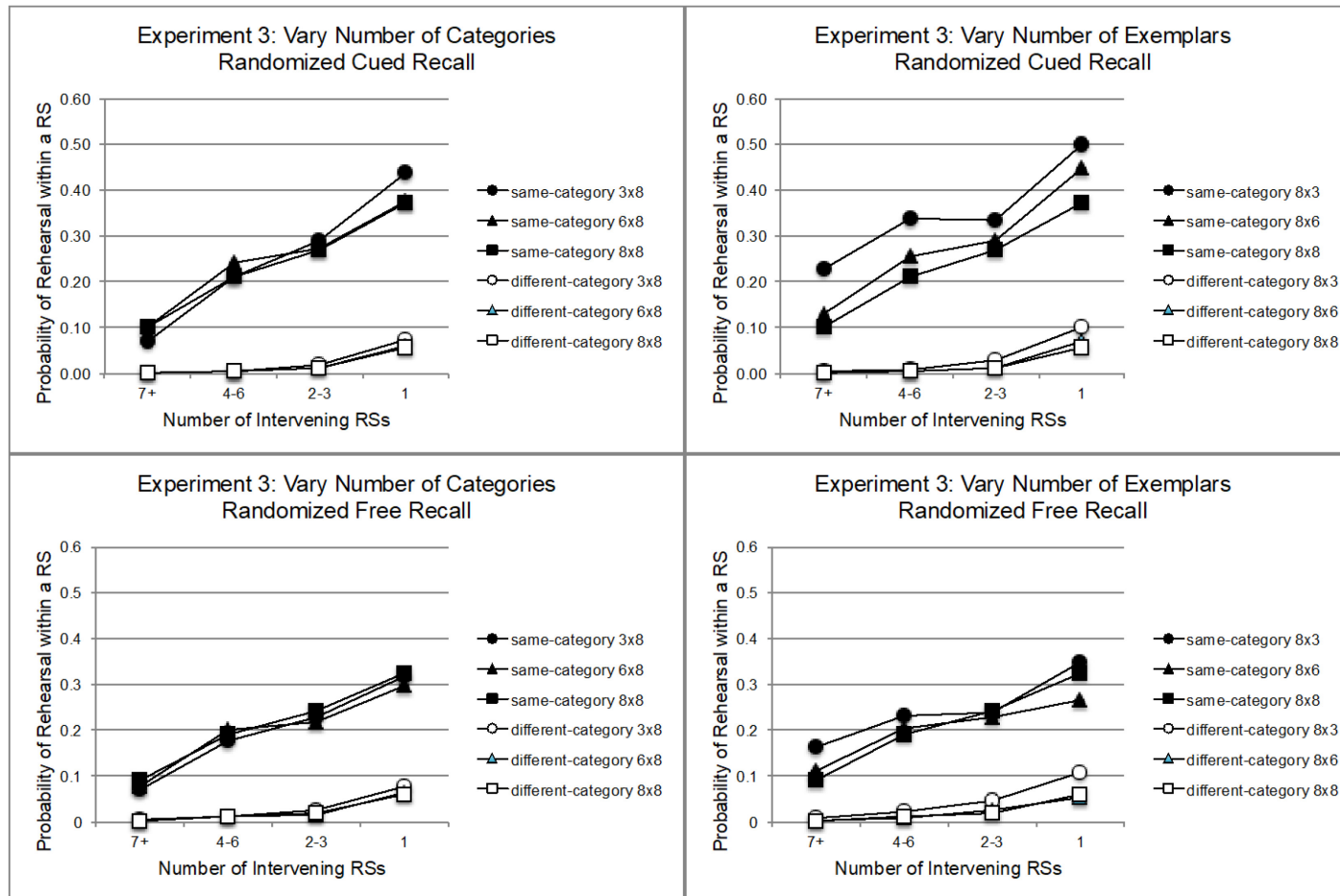
Figure S8: Data from Experiment 3. Probability of Rehearsing an item within a RS in cued recall (top) and free recall (bottom) when the list length was increased by increasing the number of 8-item categories from 3x8 (left) to 6x8 (middle) to 8x8 (right).
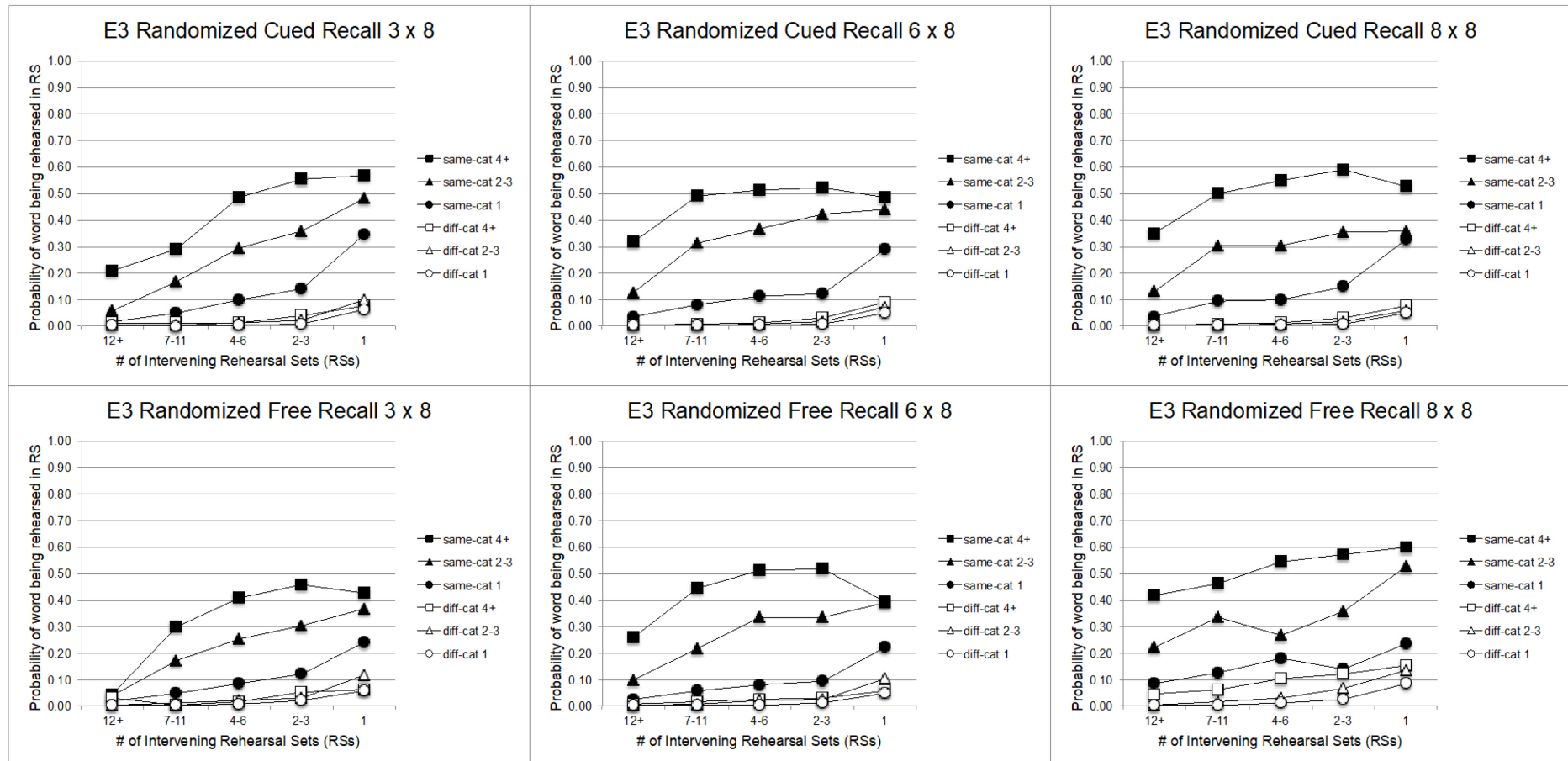
Figure S9: Data from Experiment 3. Probability of Rehearsing an item within a RS in cued recall (top) and free recall (bottom) when the list length was increased by increasing the number of exemplars per category from 8x3 (left) to 8x6 (middle) to 8x8 (right).
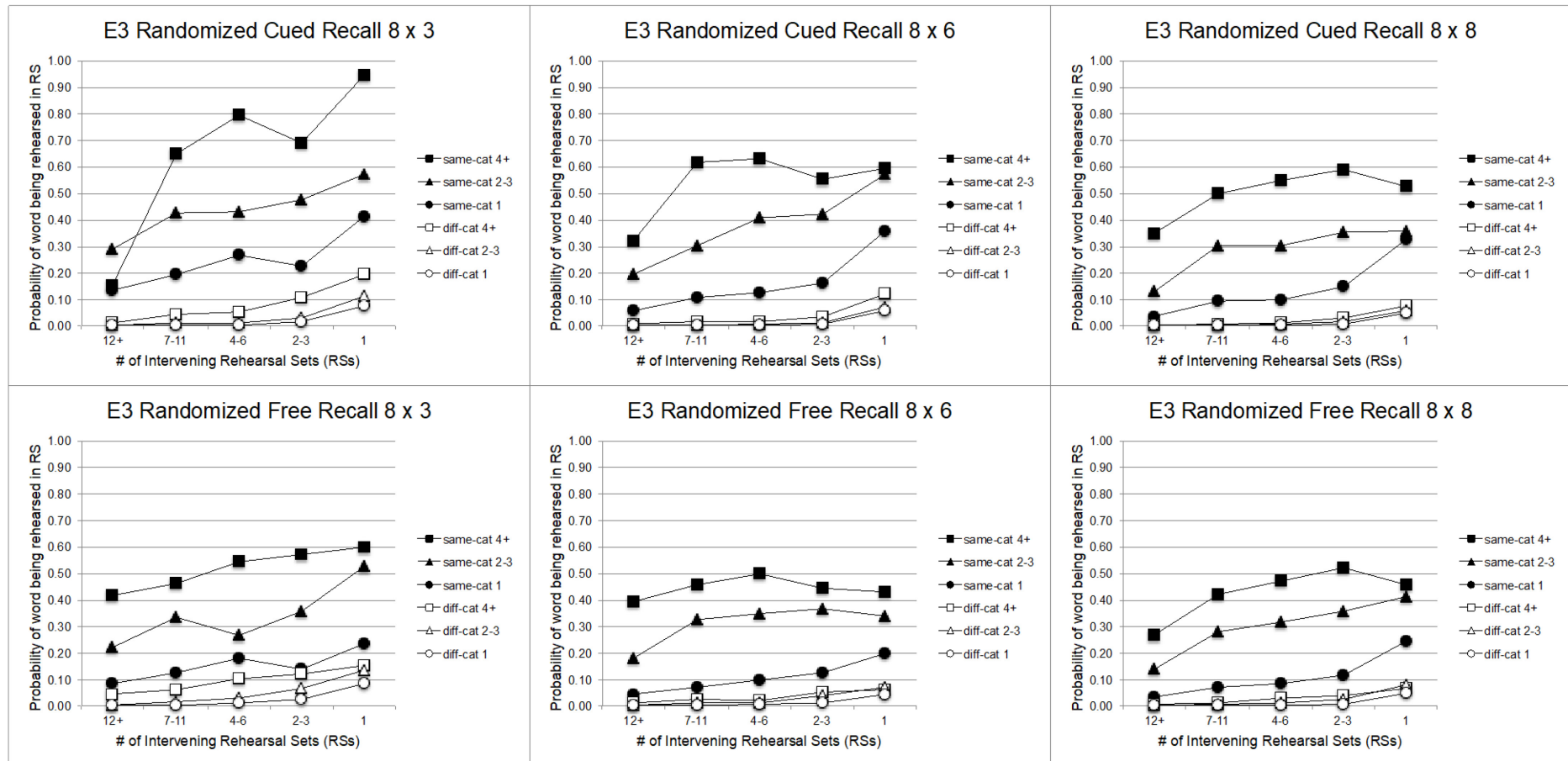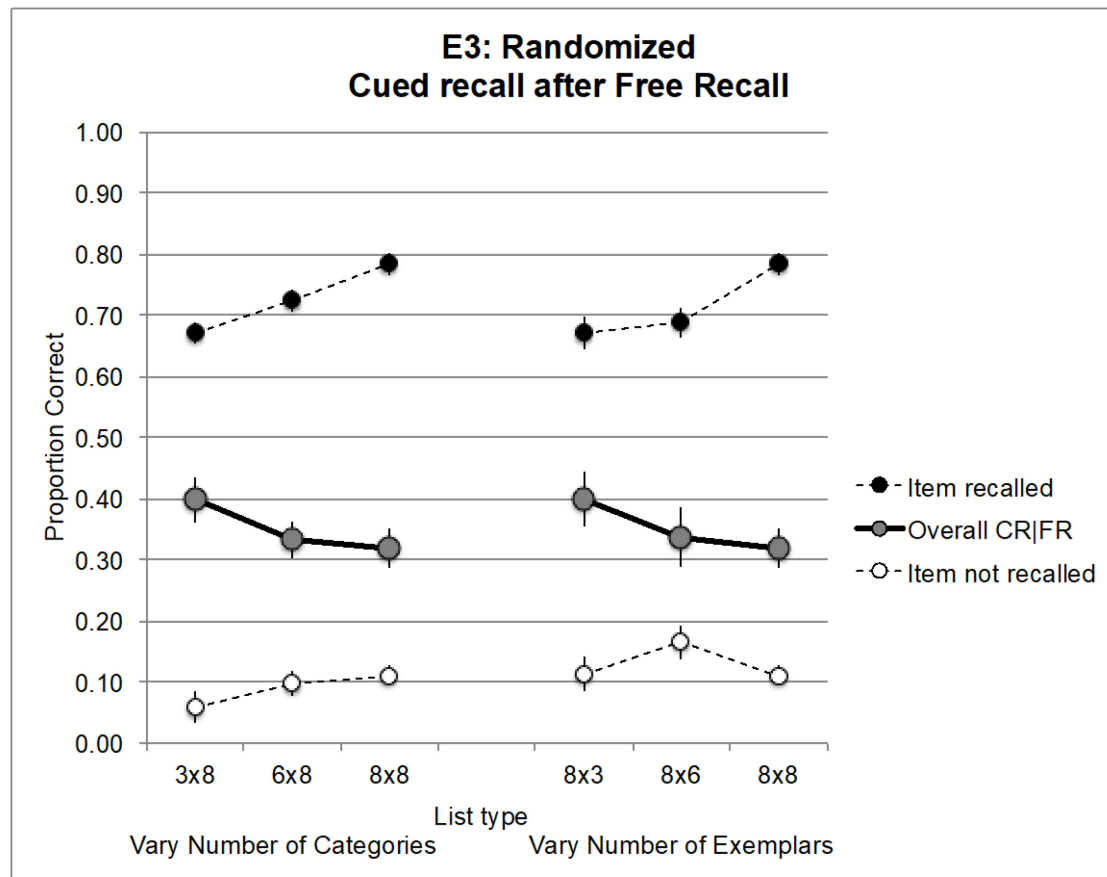
Figure S10: Data from Experiment 3. Final Cued recall performance for the participants who undertook the immediate tests of free recall. The large grey circles represent the mean final cued recall performance for each of the list types. The dashed lines represent the means of the subset of recalls for words that had been recalled (filled circles) and unrecalled (open circles) in the immediate test of free recall.

Appendix

Appendix S1.1. Rehearsal data from Experiment 2

Rehearsal data from Experiment 2 examining the effects of increasing number of categories on the mean probability of rehearsing an item in a Rehearsal Set (RS) averaged across all the RSs for the lists. Only items from the 3x8, 6x8 and 8x8 trial types were included in this analysis. Summary of mixed ANOVA: the between-subject independent variable was task (cued recall and free recall), the within-subjects independent variables were Semantic Category (whether the earlier item was from the same or different semantic category as the just-presented item), the number of intervening RSs (the number of rehearsal sets since that item was last rehearsed), and the number of 8-exemplar categories (3x8, 6x8, and 8x8 list types). Only 18 participants in the cued recall group and 16 participants in the free recall had data in each of the 24 cells.

Source	*F*-value	*MSE*	η_p^2	*p*
Task	*F* (1, 32) = 0.83	.033	.025	*p* = .369
Semantic Category	**F (1, 32) = 46.05**	**.016**	**.590**	**p < .001**
Number of Intervening (Int.) RSs	**F (3, 96) = 24.90**	**.012**	**.438**	**p < .001**
Number of Categories (Cats.)	**F (2, 62) = 7.18**	**.001**	**.183**	**p = .002**
Task x Semantic Category	*F* (1, 32) = 2.08	.016	.020	*p* = .159
Task x Intervening RSs	*F* (3, 96) = 0.15	.012	.005	*p* = .928
Task x Number of Cats	*F* (2, 64) = 0.65	.001	.020	*p* = .523
Semantic Category X Intervening RSs	**F (3, 96) = 16.76**	**.007**	**.344**	**p < .001**
Semantic Category X Number of Cats	**F (2, 64) = 5.40**	**.001**	**.144**	**p = .007**
Intervening RSs x Number of Cats.	*F* (6, 192) = 2.09	.001	.061	*p* = .057
Task x Semantic Category X Intervening RSs	*F* (3, 96) = 0.98	.007	.002	*p* = .977
Task x Semantic Category X Number of Cats.	*F* (6, 192) = 1.09	.001	.033	*p* = .368
Semantic Category X Intervening RSs X Number of Cats	*F* (6, 192) = 1.52	.001	.045	*p* = .172
Task x Semantic Category X Intervening RSs X Number of Cats	*F* (6, 192) = 1.23	.001	.037	*p* = .291

Appendix S1.2. Rehearsal data from Experiment 2

Rehearsal data from Experiment 2 examining the effects of increasing number of exemplars on the mean probability of rehearsing an item in a Rehearsal Set (RS) averaged across all the RSs for the lists. Only items from the 8x3, 8x6 and 8x8 trial types were included in this analysis. Summary of mixed ANOVA: the between-subject independent variable was task (cued recall and free recall), the within-subjects independent variables were Semantic Category (whether the earlier item was from the same or different semantic category as the just-presented item), the number of intervening RSs (the number of rehearsal sets since that item was last rehearsed, 1 or 2-3), and the number of 8-exemplar categories (8x3, 8x6, and 8x8 list types).

Source	*F*-value	*MSE*	$\eta_\square^2$	*p*
Task	$F(1, 46) = 0.04$	.108	.001	$p = .850$
Semantic Category	**$F(1, 46) = 76.10$**	**.048**	**.623**	**$p < .001$**
Number of Intervening RSs	**$F(1, 46) = 28.86$**	**.032**	**.369**	**$p < .001$**
Number of Exemplars (Exem)	**$F(2, 92) = 41.55$**	**.005**	**.475**	**$p < .001$**
Task x Semantic Category	$F(1, 46) = 0.02$	.048	.001	$p = .880$
Task x Intervening RSs	$F(1, 46) = 0.004$	.032	<.001	$p = .948$
Task x Number of Exem	$F(2, 92) = 1.37$	.005	.029	$p = .260$
Semantic Category X Intervening RSs	**$F(1, 46) = 12.56$**	**.011**	**.214**	**$p = .001$**
Semantic Category X Number of Exem	**$F(2, 92) = 9.76$**	**.005**	**.175**	**$p < .001$**
Intervening RSs x Number of Exem.	$F(2, 92) = 0.29$	.003	.314	$p = .750$
Task x Semantic Category X Intervening RSs	$F(1,46) = 0.11$	.008	.002	$p = .738$
Task x Semantic Category X Number of Exem.	$F(2, 92) = 0.95$	.010	.020	$p = .392$
Task x Intervening RSs X Number of Exem	$F(2, 92) = 0.45$	.003	.010	$p = .640$
Semantic Category X Intervening RSs X Number of Exem	$F(2,92) = 0.47$	.002	.010	$p = .626$
Task x Semantic Category X Intervening RSs X Number of Exem.	$F(2, 92) = 0.44$	.002	.009	$p = .645$

Appendix S2.1. Rehearsal data from Experiment 3

Rehearsal data from Experiment 3 examining the effects of increasing number of categories on the mean probability of rehearsing an item in a Rehearsal Set (RS) averaged across all the RSs for the lists. Only items from the 3x8, 6x8 and 8x8 trial types were included in this analysis. Summary of mixed ANOVA examining the mean probability of rehearsing an item in a Rehearsal Set (RS) averaged across all the RSs for the lists. Only items from the 3x8, 6x8 and 8x8 trial types were included in this analysis. Summary of mixed ANOVA: the between-subject independent variable was task (cued recall and free recall), the within-subjects independent variables were Semantic Category (whether the earlier item was from the same or different semantic category as the just-presented item), the number of intervening RSs (the number of rehearsal sets since that item was last rehearsed), and the number of 8-exemplar categories (3x8, 6x8, and 8x8 list types).

Source	*F*-value	*MSE*	η_{d}^{2}	*p*
Task	$F\,(1,46) = 0.172$	.112	.004	$p = .680$
Semantic Category	**$F\,(1,46) = 108.0$**	**.077**	**.701**	**$p < .001$**
Number of Intervening RSs	**$F\,(3,138) = 39.94$**	**.019**	**.465**	**$p < .001$**
Number of Categories	$F\,(2,92) = 1.19$	.010	.025	$p = .309$
Task x Semantic Category	$F\,(1,46) = 0.17$	.077	.004	$p = .685$
Task x Intervening RSs	$F\,(3,138) = 0.06$	.019	.001	$p = .983$
Task x Number of Cats	$F\,(2,92) = 0.02$	.003	$< .001$	$p = .958$
Semantic Category X Intervening RSs	**$F\,(3,138) = 30.88$**	**.008**	**.402**	**$p < .001$**
Semantic Category X Number of Cats	$F\,(2,92) = 1.72$	.004	.036	$p = .186$
Intervening RSs x Number of Cats.	$F\,(6,276) = 1.45$	.002	.031	$p = .195$
Task x Semantic Category X Intervening RSs	$F\,(3,138) = 0.04$	.008	.001	$p = .990$
Task x Semantic Category X Number of Cats.	$F\,(2,92) = 0.01$	.004	$< .001$	$p = .986$
Task x Int. RSs X Number of Cats	$F\,(6,276) = 0.03$	.004	.001	$p > .999$
Semantic Category X Intervening RSs X Number of Cats	$F\,(6,276) = 0.631$	.002	.015	$p = .631$
Task x Semantic Category X Intervening RSs X Number of Cats	$F\,(6,276) < 0.01$	.002	$< .001$	$p > .999$

Appendix S2.2. Rehearsal data from Experiment 3
Rehearsal data from Experiment 3 examining the effects of increasing number of exemplars on the mean probability of rehearsing an item in a Rehearsal Set (RS) averaged across all the RSs for the lists. Only items from the 8x3, 8x6 and 8x8 trial types were included in this analysis. Summary of mixed ANOVA: the between-subject independent variable was task (cued recall and free recall), the within-subjects independent variables were Semantic Category (whether the earlier item was from the same or different semantic category as the just-presented item), the number of intervening RSs (the number of rehearsal sets since that item was last rehearsed), and the number of 8-exemplar categories (8x3, 8x6, and 8x8 list types).

Source	*F*-value	*MSE*	$\eta_{\square}^2$	*p*
Task	$F(1, 46) = 0.19$	.154	.004	$p = .665$
Semantic Category	**$F(1, 46) = 96.33$**	**.101**	**.677**	**$p < .001$**
Number of Intervening RSs	**$F(3, 138) = 35.42$**	**.017**	**.435**	**$p < .001$**
Number of Exemplars (Exem)	**$F(2, 92) = 18.74$**	**.006**	**.289**	**$p < .001$**
Task x Semantic Category	$F(1, 46) = 0.17$	.101	.004	$p = .683$
Task x Intervening RSs	$F(3, 138) = 0.10$	.021	.017	$p = .961$
Task x Number of Exem	$F(2, 92) = 0.10$	.003	.006	$p = .902$
Semantic Category X Intervening RSs	**$F(3, 138) = 17.37$**	**.007**	**.274**	**$p < .001$**
Semantic Category X Number of Exem	$F(2, 92) = 1.69$	.007	.036	$p = .190$
Intervening RSs x Number of Exem.	**$F(6, 276) = 2.71$**	**.003**	**.056**	**$p = .014$**
Task x Semantic Category X Intervening RSs	$F(3, 138) = 0.09$	.007	.002	$p = .964$
Task x Semantic Category. X Number of Exem.	$F(2, 92) = 0.04$	.007	.001	$p = .958$
Task x Intervening RSs X Number of Exem	$F(6, 276) = 0.04$	.003	.001	$p > .999$
Semantic Category X Intervening RSs X Number of Exem	**$F(6, 276) = 5.64$**	**.002**	**.109**	**$p < .001$**
Task x Semantic Category X Intervening RSs X Number of Exem.	$F(6, 276) < 0.01$	.002	.001	$p > .999$