Experiment 3

In addition to the interactions reported in the main text, we separately examined effects of value on list 1 and effects of value on recall averaged across lists 2-8. For list 1, a 2 x 3 (item value x value sensitivity) mixed ANOVA with repeated measures on the first factor, on recall for items recalled in the correct S/P form, showed a main effect of value, $F(1, 107) = 45.28$, $p < .001$, $\eta_p^2 = .30$, no main effect of value sensitivity, $F(2, 107) = 2.35$, $p = .101$, $\eta_p^2 = .04$, and an interaction between item value and value sensitivity, $F(2, 107) = 3.31$, $p = .040$, $\eta_p^2 = .06$. Planned comparisons with paired-samples t-tests were consistent with these results, as we found a marginal effect of value in the Weak group, $t(13) = 1.90$, $p = .080$, $d = .51$, a significant effect of value in the Moderate group, $t(37) = 3.89$, $p < .001$, $d = .63$, and significant effect of value in the Strong group, $t(57) = 9.17$, $p < .001$, $d = 1.20$. Thus, even on the first list, value effects tended to be stronger in individuals who reported being more sensitive to value.

We next turn to the analogous set of analyses for lists 2-8. Here, a 2 x 3 (item value x value sensitivity) mixed ANOVA showed a main effect of item value, $F(1, 107) = 208.69$, $p < .001$, $\eta_p^2 = .66$, a main effect of value sensitivity, $F(2, 107) = 9.04$, $p < .001$, $\eta_p^2 = .14$, and an interaction between item value and value sensitivity, $F(2, 107) = 23.55$, $p < .001$, $\eta_p^2 = .31$. Planned comparisons with paired-samples t-tests found that individuals in the Weak value sensitivity group showed a significant effect of value on recall with correct S/P form, $t(13) = 2.37$, $p = .034$, $d = .63$, as did those in the Moderate group, $t(37) = 9.56$, $p < .001$, $d = 1.55$, and those in the Strong group, $t(57) = 21.14$, $p < .001$, $d = 2.78$. Thus, all participants did show some value-

related increase in recall with correct S/P form on lists 2-8, although the effect was reliably greater with stronger strategy use.

We can also examine how self-reported strategy use changed the effects of value on item-only recall. On list 1, a 2 x 3 (item value x value sensitivity) mixed ANOVA showed a main effect of item value, $F(1, 107) = 15.33$, $p < .001$, $\eta_p^2 = .13$, a marginal main effect of value sensitivity, $F(2, 107) = 2.89$, $p = .060$, $\eta_p^2 = .05$, but also no interaction between these factors, $F(2, 107) < 1$, $\eta_p^2 = .01$. Despite the lack of an interaction, planned comparisons with paired-samples t-tests showed no reliable effect of value in the Weak group, $t(13) < 1$, $d = .27$, but a significant effect of value in the Moderate group, $t(37) = 2.75$, $p = .009$, $d = .45$, and a significant effect of value in the Strong group, $t(57) = 4.42$, $p < .001$, $d = .58$. Averaging across lists 2 through 8, a 2 x 3 (item value x value sensitivity) mixed ANOVA showed a main effect of item value, $F(1, 107) = 33.67$, $p < .001$, $\eta_p^2 = .24$, no effect of value sensitivity, $F(2, 107) < 1$, $\eta_p^2 = .00$, and a significant item value x value sensitivity interaction, $F(2, 107) = 8.02$, $p = .001$, $\eta_p^2 = .13$. Planned comparisons using paired-samples t-tests show that for the Weak group, there was no effect of item value on recall in this category, $t(13) < 1$, $d = -.06$, but for individuals in the Moderate group, $t(37) = 5.36$, $p < .001$, $d = .87$, and for individuals in the Strong group, $t(57) = 8.19$, $p < .001$, $d = 1.08$, high-value items did show greater item-only recall than do low-value items. Thus, the effect of value on item-only recall is only reliable in the Moderate and Strong groups, both before and after test experience.

Note that the finding that high-value items show better memory than low-value items on list 1, for both recall with S/P form and item-only recall, does not follow directly from our hypotheses. This result suggests that there may be some degree of selective strategy use that does not require experience with a full study-test cycle. Given that college students are expert learners, we speculate that some individuals may have engaged strategies selectively based merely on the expectation of an upcoming free recall test (cf., Middlebrooks et al., in press). Still, given that value effects do become stronger on both types of recall following list 1 for participants in the Moderate and Strong strategy groups, our results also suggest that any engagement of selective strategy use on list 1 was significantly less robust than on subsequent lists, as would be predicted by our hypothesis regarding test-potentiated learning.

Finally, we also report the results of planned comparisons showing effects of value on the two different measures of recognition memory performance for the three value sensitivity groups. Specifically, using paired-samples t-tests, we find that for the plurals test, there was not a significant effect of value on accuracy for individuals in the Weak group, $t(10) = 1.05$, $p = .320$, $d = .32$. In the Moderate group, there was a significant effect of value on this test, $t(18) = 3.25$, $p = .004$, $d = .75$, while the effect of value was not significant in the Strong group, $t(33) = 1.35$, $p = .185$, $d = .23$. For the speeded item test, planned comparisons using paired-samples t-tests showed that value did not affect reliably performance in the Weak group, $t(10) = 1.37$, $p = .200$, $d = .41$, nor in the Moderate group, $t(18) < 1$, $d = .19$, nor in the Strong group, $t(33) = 1.15$, $p = .258$, $d = .20$. Although the relative strength of effects

between groups in these analyses do not replicate patterns shown elsewhere in this dataset, the trends should be interpreted with a great deal of caution. As noted in the main text, there was no item value x value sensitivity interaction on either measure. Additionally, as noted in the Experiment 3 Discussion, there are specific reasons why separating these results by value sensitivity might be noisier than other aspects of the dataset. Thus, while they are reported here for completeness, these analyses ultimately appear to be inconclusive.

Experiment 4

We also break down the interactions from the free recall data in Experiment 4 more fully here. As noted in the main text, there were too few individuals in the Weak value sensitivity group to run meaningful inferential statistics on that group, so the following analyses only included individuals in the Moderate and Strong value sensitivity groups. We first ran a 2 x 2 x 2 (item value x list group x value sensitivity) mixed ANOVA, with repeated measures on the first two factors, on the rate of recall with correct S/P form (Supplemental Table 1). This analysis showed a main effect of value, $F(1, 43) = 72.36$, $p < .001$, $\eta_p^2 = .63$, and an interaction between item value and value sensitivity, $F(1, 43) = 4.16$, $p = .047$, $\eta_p^2 = .09$, indicating that the effect of value on memory was somewhat stronger in the Strong group than in the Moderate group. The analysis also showed a significant interaction between value and list group, $F(1, 43) = 11.05$, $p = .002$, $\eta_p^2 = .20$, indicating that value effects became stronger with practice. However, the 3-way interaction was not significant, $F(1, 43) = 2.45$, $p = .125$, $\eta_p^2 = .05$, indicating that the degree to which value effects

changed with practice did not differ between the Moderate and Strong groups. No other effects in the analysis were significant, all $F < 1.11$, all $\eta_p^2 < .03$.

We also ran a 2 x 2 x 2 (item value x list group x value sensitivity) mixed ANOVA on the corrected rate of item-only recall (Supplemental Table 1). This analysis found a main effect of item value, $F(1, 43) = 41.66$, $p < .001$, $\eta_p^2 = .49$, and an interaction between item value and value sensitivity, $F(1, 43) = 4.10$, $p = .049$, $\eta_p^2 = .09$, again reflecting a somewhat stronger effect of value in the Strong group. The test group x value interaction was marginal here, $F(1, 43) = 3.82$, $p = .057$, $\eta_p^2 = .08$, reflecting a trend for value effects on item-only recall to get stronger with test potentiation. No other effects were at or near significance, all $F < 1$, all $\eta_p^2 < .02$.

We also used planned comparisons to further break down these ANOVAs, separately examining the Moderate and Strong value sensitivity groups. Examining the likelihood of items recalled in the correct S/P form, a 2 x 2 (list group x value) repeated measures ANOVA on individuals in the Moderate group showed a main effect of value, $F(1, 10) = 20.53$, $p = .001$, $\eta_p^2 = .67$, but no main effect of list group, $F(1, 10) < 1$, $\eta_p^2 = .02$, and no interaction, $F(1, 10) = 1.16$, $p = .307$, $\eta_p^2 = .10$. Still, despite the failure to reach significance, the numeric trend is for effects of value to become stronger with practice in this group. An analogous ANOVA on those in the Strong group finds a main effect of value, $F(1, 33) = 103.53$, $p < .001$, $\eta_p^2 = .76$, no reliable effect of list group $F(1, 33) = 2.10$, $p = .156$, $\eta_p^2 = .06$, and a significant interaction, $F(1, 33) = 23.62$, $p < .001$, $\eta_p^2 = .42$, showing that the effect of value does become reliably stronger with practice in this group.

After similarly breaking down the analysis for item-only recall into separate

2 x 2 (list group x item value) ANOVAs based on self-reported value sensitivity, we

found, in the Moderate group, a main effect of item value, $F(1, 10) = 12.71$, $p = .005$,

$\eta_p^2 = .56$, no main effect of list group, $F(1, 10) < 1$, $\eta_p^2 = .06$, and a significant

interaction, $F(1, 10) = 6.29$, $p = .031$, $\eta_p^2 = .39$. In the Strong group, an analogous 2 x

2 ANOVA showed a main effect of value, $F(1, 33) = 64.05$, $p < .001$, $\eta_p^2 = .66$, no main

effect of list group, $F(1, 33) < 1$, $\eta_p^2 = .01$, and no reliable interaction, $F(1, 33) = 1.85$,

$p = .183$, $\eta_p^2 = .05$. Thus, there is a trend for value effects to get stronger with test

potentiation across the Moderate and Strong strategy use groups, although the

interaction was only statistically reliable within the Moderate group.

Finally, we separately examined effects of value on recall scores from list 1,

and scores computed based on all lists after list 1, with self-reported value

sensitivity as an additional factor. For list 1, a 2 x 2 (item value x value sensitivity)

mixed ANOVA examining effects of value on recall with correct S/P form showed a

main effect of item value, $F(1, 43) = 20.40$, $p < .001$, $\eta_p^2 = .32$, but no main effect of

value sensitivity, $F(1, 43) < 1$, $\eta_p^2 = .01$, and no interaction, $F(1, 43) < 1$, $\eta_p^2 = .01$.

Planned comparisons with separate paired-samples t-tests confirm that effects of

value were significant in the Moderate group, $t(10) = 4.25$, $p = .002$, $d = 1.28$, and in

the Strong group, $t(33) = 4.72$, $p < .001$, $d = .81$. A similar analysis on the recall data

from the two tested lists among lists 2-8 showed a similar overall pattern of effects,

but an increase in their magnitude. A 2 x 2 (item value x value sensitivity) mixed

ANOVA found a main effect of item value, $F(1, 43) = 85.18$, $p < .001$, $\eta_p^2 = .66$, no

main effect of value sensitivity, $F(1, 43) = 1.54$, $p = .222$, $\eta_p^2 = .03$, and an interaction,

indicating that the effect of value was significantly stronger in the Strong value sensitivity group, $F(1, 43) = 7.32$, $p = .010$, $\eta_p^2 = .15$. Still, both the Moderate group, $t(10) = 3.26$, $p = .009$, $d = .98$, and the Strong group, $t(33) = 12.72$, $p < .001$, $d = 2.18$, show highly significant effects of value on recall of items in the correct S/P form.

We also examined item-only recall, separating list 1 from subsequent lists. On list 1, a 2 x 2 (item value x value sensitivity) ANOVA showed a main effect of item value, $F(1, 43) = 7.55$, $p = .009$, $\eta_p^2 = .15$, no main effect of value sensitivity, $F(1, 43) = 1.48$, $p = .230$, $\eta_p^2 = .03$, and a marginal trend towards an interaction, $F(1, 43) = 3.11$, $p = .085$, $\eta_p^2 = .07$. Planned comparisons with paired-samples t-tests show that this effect of value, prior to any test experience, is not significant in the Moderate group, $t(10) < 1$, $d = .27$, but is significant in the Strong group, $t(33) = 4.20$, $p < .001$, $d = .72$. For later lists, a 2 x 2 (item value x value sensitivity) ANOVA showed a main effect of item value, $F(1, 43) = 27.86$, $p < .001$, $\eta_p^2 = .39$, but no main effect of value sensitivity, $F(1, 43) < 1$, $\eta_p^2 = .00$, and no item value x value sensitivity interaction, $F(1, 43) < 1$, $\eta_p^2 = .02$. Planned comparisons showed that both the Moderate group, $t(10) = 4.16$, $p = .002$, $d = 1.25$, and the Strong group, $t(33) = 5.69$, $p < .001$, $d = .98$, had stronger item-only recall for high-value items.

We also split the data from the recognition tests by value sensitivity, comparing individuals from the Moderate and Strong groups. A 2 x 2 (item value x value sensitivity) repeated measures ANOVA found that on the plurals test, there was a main effect of value, $F(1, 43) = 13.79$, $p = .001$, $\eta_p^2 = .24$, but no main effect of value sensitivity, $F(1, 43) < 1$, $\eta_p^2 = .00$, and no significant interaction between item value and value sensitivity, $F(1, 43) = 2.25$, $p = .141$, $\eta_p^2 = .05$. For the speeded test,

an analogous ANOVA found a main effect of value, $F(1, 43) = 8.56$, $p = .005$, $\eta_p^2 = .17$, a main effect of value sensitivity showing better overall performance on this test in the Strong group, $F(1, 43) = 5.15$, $p = .028$, $\eta_p^2 = .11$, and no interaction between value and value sensitivity, $F(1, 43) < 1$, $\eta_p^2 = .01$.

We also examined planned comparisons separating the Moderate and Strong groups' recognition data. For the plurals test, there was not a reliable effect of item value in the Moderate value sensitivity group, $t(10) = 1.56$, $p = .149$, $d = .47$, but the analogous effect in the Strong group was significant, $t(33) = 5.02$, $p < .001$, $d = .86$. For the speeded test, there was also no reliable effect of value in the Moderate group, $t(10) = 1.55$, $p = .153$, $d = .47$, but there was a significant effect of value in the Strong group, $t(33) = 2.71$, $p = .011$, $d = .46$. Still, although only the Strong value sensitivity group showed reliable value-related benefits on either recognition test, the Moderate group also shows a trend for value to benefit memory on both tests, and, at least on the speeded test, the effect sizes for the two groups are comparable. Additionally, as noted in the main text, there was no interaction between item value and value sensitivity on either measure. Thus, while not conclusive, these results are at least not inconsistent with the idea that value does strengthen both measures for individuals who show either Moderate or Strong value sensitivity.

Supplemental Table 1

Mean proportion correct (and standard error) on recall, plurals test (recollection) and

speeded item test (familiarity), split by self-reported value sensitivity, in Experiment 4.

The rightmost columns display data from the combination of the Moderate and Strong

value sensitivity groups, in order to match analyses reported in the main text.

Value Sensitivity	Weak (n = 3)		Moderate (n = 11)		Strong (n = 34)		Moderate + Strong combined	
Item Value	High	Low	High	Low	High	Low	High	Low
List 1, Item + S/P Form Recall	.389 (.139)	.111 (.111)	.303 (.030)	.159 (.029)	.348 (.034)	.157 (.020)	.337 (.027)	.157 (.017)
List 2 & 3, Item + S/P Form Recall	.625 (.214)	.181 (.091)	.352 (.055)	.133 (.031)	.479 (.026)	.077 (.014)	.448 (.025)	.091 (.013)
List 1, Item-only Recall (Corrected)	.352 (.081)	.083 (.083)	.072 (.023)	.053 (.024)	.136 (.019)	.048 (.012)	.120 (.016)	.049 (.011)
List 2 & 3, Item-only Recall (Corrected)	.019 (.019)	.035 (.018)	.133 (.022)	.035 (.013)	.150 (.023)	.015 (.005)	.146 (.018)	.020 (.005)
Plurals Recognition	.544 (.048)	.600 (.033)	.612 (.045)	.558 (.035)	.662 (.021)	.533 (.014)	.650 (.019)	.539 (.013)
Speeded Recognition	.233 (.100)	.322 (.091)	.500 (.040)	.421 (.024)	.537 (.013)	.482 (.015)	.528 (.014)	.467 (.013)