CrossMark

BRIEF REPORT

The puzzle of study time allocation for the most challenging items

Monika Undorf[1] · Rakefet Ackerman[2]

Abstract Learners often allocate more study time to challenging items than to easier ones. Nevertheless, both predicted and actual memory performance are typically worse for difficult than for easier items. The resulting inverse relations between people's predictions of their memory performance (judgments of learning; JOLs) and self-paced study time (ST) are often explained by bottom-up, data-driven ST allocation that is based on fluency. However, we demonstrate robust inverted U-shaped relations between JOLs and ST that cannot be explained by data-driven ST allocation alone. Consequently, we explored how two models of top-down, strategic ST allocation account for curvilinear JOL-ST relations. First, according to the Region of Proximal Learning model, people stop quickly on items for which they experience too little progress in learning. Second, according to the Diminishing Criterion Model, people set a time limit and stop studying when this time limit is reached. In three experiments, we manipulated motivation with different methods and examined which model best described JOL-ST relations. Consistent with the Diminishing Criterion Model but not with the Region of Proximal Learning model, results revealed that curvilinearity was due to people setting a time limit.

✉ Monika Undorf
  undorf@uni-mannheim.de

[1] Department of Psychology, School of Social Sciences, University of Mannheim, Ehrenhof-Ost, 68131 Mannheim, Germany

[2] Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology, Haifa, Israel

Everyday learning is often self-regulated. For example, students preparing for an exam decide when, how, and how long to study each topic. Research on metacognition has paid a great deal of attention to the ways in which people allocate study time (ST) during learning. This work has shown a close link between learners' predictions of their own future memory performance (judgments of learning; JOLs) and time allocation (e.g., Metcalfe & Finn, 2008).

Within a given task, one may expect that allocating more ST to an item would result in higher JOLs. Consistent with this idea, items that are awarded high incentives for recall are studied longer and receive higher JOLs than items awarded low incentives (e.g., Koriat, Ma'ayan, & Nussinson, 2006). Within incentive levels, however, items that are studied longer receive lower JOLs, resulting in inverse JOL-ST relations. Koriat et al. (2006) argued that these inverse relations are data-driven and stem from variability in item difficulty: For each item, ST is determined in a bottom-up manner by the individual learner's memorizing effort or processing fluency, which people use as a basis for JOLs (see also Undorf & Erdfelder, 2011).

Some findings suggest that data-driven ST allocation alone cannot explain JOL-ST relations. First, while data-driven ST allocation predicts strong inverse JOL-ST relations, correlations are often moderate (e.g., −.42, Koriat et al., 2006, Exp. 1; −.30 or less, Undorf & Erdfelder, 2013). Second, while data-driven ST allocation predicts linear decreases in ST with increasing JOLs, a few studies reported inverted U-shaped curvilinear JOL-ST relations (Little & McDaniel, 2015; Mazzoni & Cornoldi, 1993; Metcalfe & Kornell, 2005; Pelegrina, Bajo, & Justicia, 2000). For instance, Undorf and Erdfelder (2013), Experiment 1 asked undergraduates to self-pace their study of related and unrelated word pairs for a cued recall test and provide a JOL for each item. When we reanalyzed the data, ST was similar across the lowest JOL levels and decreased with increasing JOLs thereafter (Fig. 1).
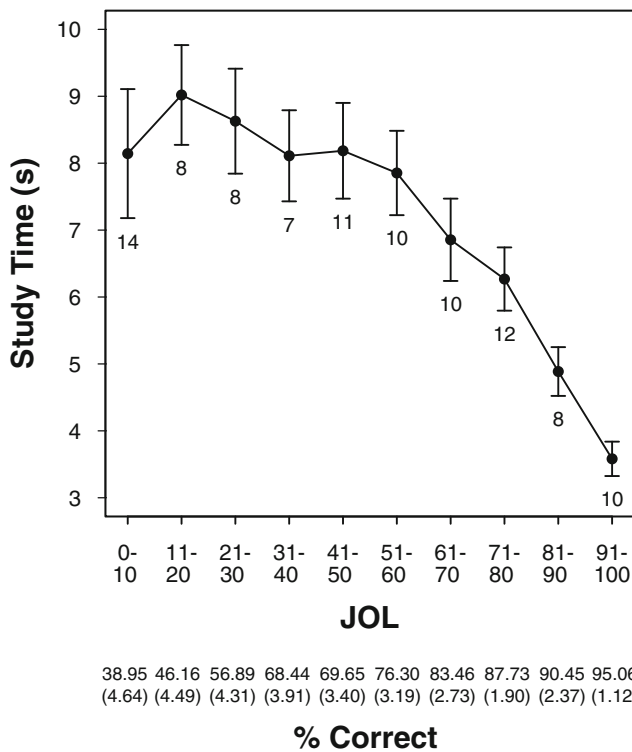
Fig. 1 The graph shows means of median study time (in seconds) for ten judgment of learning (JOL) levels (0–10, 11–20, …, 91–100) found by Undorf and Erdfelder (2013), Exp. 1. Numbers near the graph represent mean percentages of items per JOL level. Error bars represent one standard error of the mean. The table below the graph presents means (and standard deviations) of recall performance (% correct)

Fig. 2 Mean z-transformed study time for six judgment of learning (JOL) levels for participants with short and long overall study time (ST), reported by Metcalfe and Kornell (2005). Participants with short overall ST stopped quickly on the most challenging items. This produced inverted U-shaped JOL-ST relations with shorter ST for items with lowest JOLs than for items with moderately low JOLs. In contrast, participants with long overall ST did not stop quickly, resulting in linear negative JOL-ST relations. From "A region of proximal learning model of study time allocation," by J. Metcalfe and N. Kornell, 2005, *Journal of Memory and Language*, *52*, Fig. 6, p. 473. Copyright 2005 by Elsevier Inc.

Curvilinear patterns may be artifacts of aggregation that occur when each individual's JOL-ST relation is linear negative, but some individuals spend little ST and make low JOLs. To test this possibility, we fitted a multilevel regression model to Undorf and Erdfelder's (2013) data (see 'Procedure' section of Experiment 1 for details). Results revealed that JOL-ST relations were curvilinear at the individual level.

As explained above, data-driven ST allocation predicts inverse linear JOL-ST relations and cannot account for curvilinearity as depicted in Fig. 1. Consequently, curvilinear JOL-ST relations may inform theories of top-down strategic ST allocation. According to the influential *Discrepancy Reduction Model* (Nelson & Narens, 1990), people study each item until its JOL meets a preset target level. Because people stop studying irrespective of invested ST, this model predicts that JOLs and ST are unrelated. Hence, the discrepancy reduction model cannot account for curvilinear JOL-ST relations.

According to the *Region of Proximal Learning Model* (RPL; Metcalfe & Kornell, 2005), people stop studying each item when the perceived rate of learning falls below a stopping criterion. When items are extremely difficult, this occurs very quickly. Thus, learners stop quickly and make very low JOLs (see Fig. 2, short ST). This results in inverted U-shaped JOL-ST
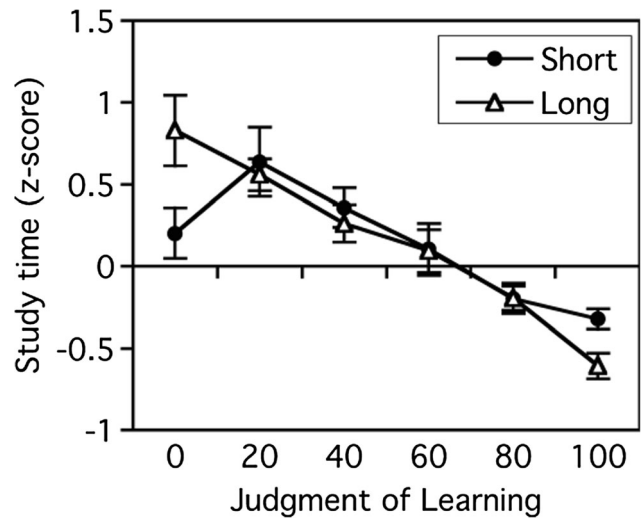
relations with shorter ST for items with lowest JOLs than with moderately low JOLs.

The *Diminishing Criterion Model* (*DCM*, Ackerman, 2014) was proposed to explain negative correlations between problem-solving time and confidence (e.g., Koriat et al., 2006). Like the discrepancy reduction model, the DCM assumes that people aim at a target level of confidence. Over time, however, they are willing to compromise and provide responses accompanied by lower confidence. According to the DCM, curvilinear JOL-ST relations are due to people setting a time limit for each item (see Fig. 3). If an item hits the time limit, people stop studying and make a rather low JOL. ST thus reaches a plateau at the time limit for items with lowest and moderately low JOLs, resulting in curvilinear JOL-ST relations.

In sum, both the RPL and the DCM predict curvilinear JOL-ST relations as presented in Fig. 1. However, the two models make different predictions for the exact shape of the curve and the effects of varying motivation on JOL-ST relations. As explained above, the RPL attributes curvilinearity to stopping quickly on the most challenging items. Therefore, ST should be shorter for items with lowest JOLs (quick stops) than for items with moderately low JOLs (slow stops, Metcalfe, 2009). By this model, high motivation increases the willingness to invest ST despite a slow learning progress. Thus, curvilinearity in JOL-ST relations should decrease with increasing motivation. Indeed, Metcalfe and Kornell (2005, Fig. 2) found curvilinear JOL-ST relations among participants who worked
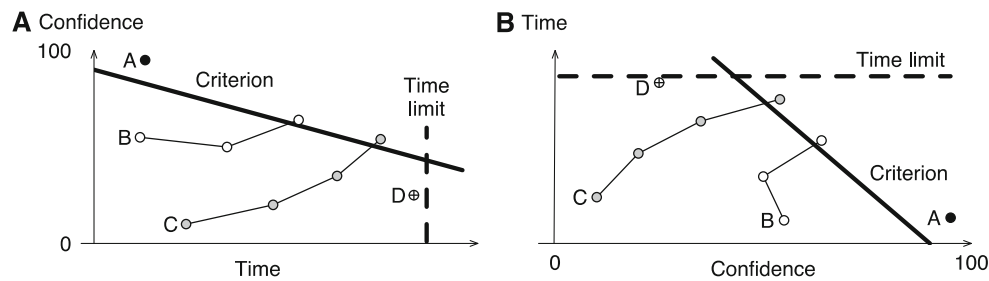
Fig. 3 The Diminishing Criterion Model proposed by Ackerman (2014). Panel **A** was adapted from "The Diminishing Criterion Model for metacognitive regulation of time investment," by R. Ackerman, 2014, *Journal of Experimental Psychology: General*, 143(3), 1349–1368, Fig. 2, p. 1351. Copyright 2013 American Psychological Association. Panel **B** represents the model with switched axes in order to be consistent with the other figures in the present article. Circles represent

hypothetical confidence ratings for Items A, B, C, and D. The thick line shows how the stopping criterion diminishes as the time limit approaches. The dashed line represents a time limit—the maximum time the participant is willing to invest in an item, regardless of the level of confidence. Thin lines show how confidence for Items B and C changes over time until the stopping criterion is reached

quickly, but not among those who worked slowly. The authors concluded that only highly motivated learners tried to master the most challenging items despite perceiving slow progress. The DCM, in contrast, attributes curvilinearity to similar ST across items with lowest and moderately low JOLs that hit the time limit (Fig. 3B). By the DCM, high motivation raises the time limit so that ST increases equally for items with lowest and moderately low JOLs. Importantly, this should leave curvilinearity unaffected.[1] In the following experiments, we manipulated motivation with three methods and examined which model best described the pattern of JOL-ST relations.

Experiment 1

In Experiment 1, we tested whether high motivation would reduce curvilinearity as predicted by the RPL or would leave curvilinearity unaffected as predicted by the DCM. We therefore awarded different incentives to the recall of different items (cf. Koriat et al., 2006). Based on Koriat et al.'s (2006) findings, we expected higher JOL, ST, and recall for high-incentive items than for low-incentive items.

Method

Participants and materials Participants were 50 University of Mannheim undergraduates. Stimuli were 60 pairs of German 4–10 letter nouns, half of which were related and half of which were unrelated. Four additional practice pairs were excluded from the analyses.

Procedure Participants were told that the experiment was about studying topics of different importance. For practice, participants studied four stories that were either more (5 points) or less (1 point) important (cf. Koriat et al., 2006). Participants then studied the word pairs one at a time. They were instructed to allocate their time so that they could recall the target words while spending as little ST as possible. For each participant, 5 points were awarded to 20 randomly chosen pairs (ten related and ten unrelated) and 1 point was awarded to the remaining pairs. At study, related and unrelated 1-point and 5-point pairs were randomly intermixed. On each trial, the incentive value appeared on the screen 2 s before the pair appeared and both remained until participants clicked a button. Immediately afterwards, the JOL prompt "Chance to recall (0–100%)?" and a percentage scale appeared. Participants indicated their JOL by clicking the scale. After study, they worked on a 90-s filler math task and then took a cued recall test in which cue words were presented in a random order and participants typed targets.

Results and discussion

Table 1 presents a summary of results (for all experiments, correlations between ST, JOL, recall, and item difficulty are shown in the Appendix). Replicating Koriat et al. (2006), ST, JOL, and recall were greater for 5-point than for 1-point items. Figure 4 reveals curvilinear JOL-ST relations in both incentive levels.

To investigate JOL-ST relations, we fitted a multilevel regression model (level 1: items, level 2: participants) using the *R* packages lme4 and lmerTest (Bates, Maechler, & Bolker, 2015; Kuznetsova, Brockhoff, & Christensen, 2015; R Core Team, 2015). Prior to the analyses, we log-transformed ST to reduce the skewness of the ST distribution[2] and we centered

---

[1] When motivation is extremely high, both models predict linear negative JOL-ST relations. This is due to no quick stops occurring (RPL) or so high a time limit that no item hits it (DCM). Also, both models predict that JOL and ST are unrelated when motivation is extremely low (RPL: quick stops on all items, DCM: all items hit time limit). However, we do not expect such extreme levels of motivation in the current experiments.

[2] In all experiments, results were unchanged when untransformed ST was used.

Table 1 Means (and standard deviations) of study time (ST), judgment of learning (JOL), and memory performance (% correct), separately for each condition of Experiments 1–3

Experiment and condition	ST	JOL	% correct
Experiment 1			
1-point items	3.57 (2.49)	45.64 (16.02)	60.55 (23.06)
5-point items	4.42*** (3.38)	53.95*** (16.13)	67.00** (24.76)
Experiment 2			
Standard	4.78 (3.19)	52.31 (10.97)	71.39 (21.56)
Accuracy	7.55* (5.27)	57.40 (14.85)	73.85 (18.24)
Experiment 3			
Standard	3.67 (2.02)	52.40 (10.15)	54.93 (12.63)
Time pressure	1.82 (0.48)***	51.52 (9.57)	48.96 (11.58)

Note. The table presents means of median ST

*$p < .05$, **$p < .01$, ***$p < .001$, for differences between conditions

JOL at the group mean (see Cohen, Cohen, West, & Aiken, 2003). In two separate regression models, one for each incentive level, we regressed ST on JOL and JOL^2. The linear predictor JOL denotes the overall slope of the JOL-ST relation, with a negative regression coefficient indicating inverse linear
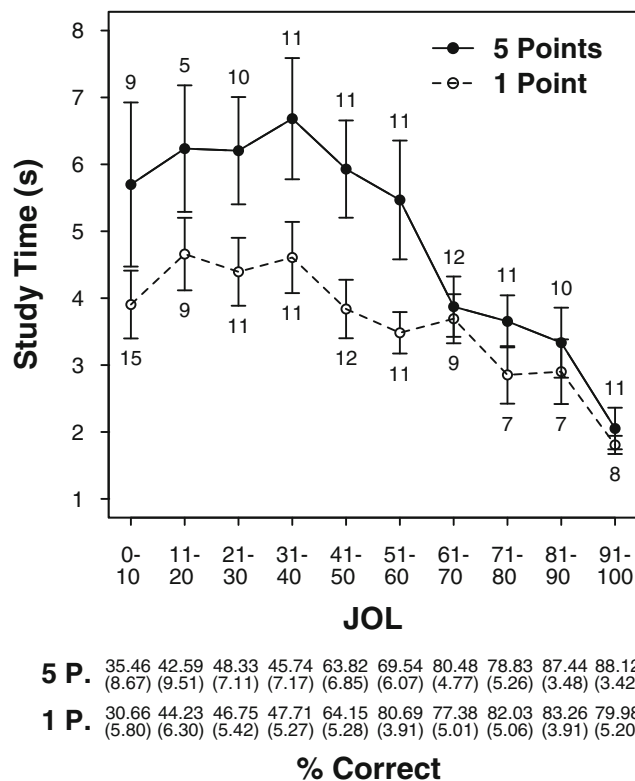
JOL-ST relations. The quadratic predictor JOL^2 denotes curvilinear JOL-ST relations, with a negative coefficient indicating an inverted U-shaped relation, a positive coefficient indicating a U-shaped relation, and an insignificant coefficient indicating that there is no curvilinearity. Including both predictors in each model allows examining slope (JOL) and curvilinearity (JOL^2) separately, while holding the other predictor constant. Random intercepts were specified for items and participants, allowing for variance in intercept across items and participants.

Table 2 presents regression results. Negative effects of JOL revealed negative JOL-ST slopes in both incentive levels. Negative effects of JOL^2 showed that JOL-ST relations were curvilinear in both incentive levels. Allowing for variance in slope and curvilinearity across participants through including random effects for JOL and JOL^2 did not improve model fit, both χ^2s < 1.

To test for differences between 5-point and 1-point items in slope and curvilinearity, we predicted ST in both incentive levels in a single regression model that included condition and condition interaction terms for JOL and JOL^2. Nonsignificant interactions of incentive level with JOL, $t(2906) = 1.49$, $p = .137$, and JOL^2, $t(2901) = 1.90$, $p = .057$, revealed that neither slope nor curvilinearity differed across incentive levels. Notably, the marginal difference in curvilinearity was in the opposite direction to that predicted by the RPL. To further examine the lack of a curvature difference across incentive levels, we compared the full model described above to a null model without interaction terms using Bayesian analyses (R package BayesFactor; Morey & Rouder, 2015). Results revealed a posterior probability of > .99 for the null model,

Fig. 4 The graphs show means of median study time (in seconds) for 10 judgment of learning (JOL) levels (0–10, 11–20, …, 91–100) for 1-point and 5-point items in Experiment 1. Numbers near the graphs represent mean percentages of items per JOL level. Error bars represent one standard error of the mean. The table below the graphs presents means (and standard deviations) of recall performance (% correct)

Table 2 Fixed effects regression estimates (and standard errors) for Experiments 1–3

Experiment and condition	Effect	
	JOL	JOL^2
Experiment 1		
1-point items	−0.19*** (0.05)	−0.73*** (0.16)
5-point items	−0.27*** (0.08)	−1.17*** (0.26)
Experiment 2		
Standard	−0.40*** (0.14)	−1.29*** (0.20)
Accuracy	−0.52***[a] (0.08)	−1.62*** (0.22)
Experiment 3		
Standard	−0.51*** (0.06)	−1.63*** (0.18)
Time pressure	−0.20***[a] (0.06)	−0.98***[a] (0.20)

Note. JOL = judgment of learning

*$p < .05$. **$p < .01$. ***$p < .001$

[a] $p < .05$, for differences between conditions

indicating very strong evidence in favor of equal curvilinearity across incentive levels (cf. Raftery, 1995).

To explore ST for items with lowest and moderately low JOLs, JOLs were grouped into bins of 20%. We compared ST for items from the two lowest bins (JOL of 20 or below vs. JOL from 21 to 40; Table 3). A 2 (incentive level) × 2 (JOL bin) within-participant ANOVA revealed that 5-point items were studied longer than 1-point items, $F(1, 27) = 7.69$, $p = .010$, $\eta_p^2 = .22$. Neither the main effect of JOL bin, $F(1, 27) = 2.59$, $p = .119$, nor the interaction, $F < 1$, were significant. Thus, ST was similar for items with lowest and moderately low JOLs.

In sum, equally curvilinear JOL-ST relations and similar ST across items with low JOLs in both incentive levels supported the DCM but not the RPL.

Experiment 2

Experiment 1 revealed that awarding high incentives to a subset of items did not affect curvilinearity. Experiment 2 tested whether high motivation would impact curvilinearity in a between-participants design. This prevents carryover effects between incentive levels and approximates motivational differences across participants as observed by Metcalfe and Kornell (2005).

In the standard condition, we emphasized both speed and accuracy of learning, which is common in self-paced learning studies (e.g., Koriat et al., 2006; Mazzoni & Cornoldi, 1993). In the accuracy condition, participants were instructed to attain complete mastery of all items and were allowed unlimited time to do so (e.g., Nelson & Leonesio, 1988). Nelson and Leonesio (1988) found that

Table 3 Means (and standard deviations) of study time (in seconds) for items from judgment of learning (JOL) bins 1 (JOLs of 20 or below) and 2 (JOLs from 21 to 40)

Experiment and condition	JOL bin	
	1	2
Experiment 1		
1-point items	4.72 (3.25)	5.22 (3.55)
5-point items	6.11 (4.49)	6.90 (5.34)
Experiment 2		
Standard	5.07 (4.54)	4.83 (3.48)
Accuracy	9.29 (7.17)	8.18 (5.04)
Experiment 3		
Standard	4.04 (2.35)	4.71 (3.30)
Time pressure	2.27 (1.46)	2.10 (0.79)

Note. The table presents means of median ST

accuracy instructions increased ST but affected neither metamemory judgments nor recall. We therefore expected higher ST in the accuracy than in the standard group.

Method

Participants and materials Sixty-two University of Mannheim undergraduates were randomly assigned to the standard group ($n = 30$) or to the accuracy group ($n = 32$). Stimuli were the same as in Experiment 1.

Procedure For the standard group, the procedure was identical to Experiment 1 except that there was no practice phase or incentive values. The accuracy group was instructed to study hard and to spend as much time as needed to master each pair. Participants were informed that they would receive five points for each correctly recalled target and that prizes (€30.00, €20.00, and €10.00) would be awarded to the three best participants.

Results and discussion

The accuracy group studied longer than the standard group, while JOLs and recall did not differ across groups (Table 1). Figure 5 reveals curvilinear JOL-ST relations in both groups.

As in Experiment 1, we regressed ST in each group on JOL and JOL². Results revealed negative JOL-ST slopes and curvilinear JOL-ST relations in both groups (Table 2). Allowing for variance in slope and curvilinearity across participants did not improve model fit, $\chi^2_{standard} < 1$, $\chi^2(2)_{accuracy} = 1.80$, $p = .407$. When testing for differences across groups, a significant condition × JOL interaction, $t(3622) = 4.76$, $p < .001$, revealed a more pronounced JOL-ST slope in the accuracy group. A nonsignificant condition × JOL² interaction, $t(3619) = 1.49$, $p = .137$, demonstrated that curvilinearity did not differ across groups. Bayesian analyses indicated very strong evidence in favor of equal curvilinearity across groups, posterior probability of a null model without condition interaction terms > .99.

ST for the two lowest JOL bins (Table 3) was submitted to a mixed 2 (condition) × 2 (JOL bin) ANOVA. Results revealed longer ST in the accuracy group than in the standard group, $F(1, 45) = 6.92$, $p = .012$, $\eta_p^2 = .13$. Neither the main effect of JOL bin, $F(1, 52) = 1.63$, $p = .208$, nor the interaction, $F < 1$, were significant.

In sum, the between-participant manipulation of motivation increased ST but not JOLs or recall. Replicating Experiment 1, equally curvilinear JOL-ST relations and
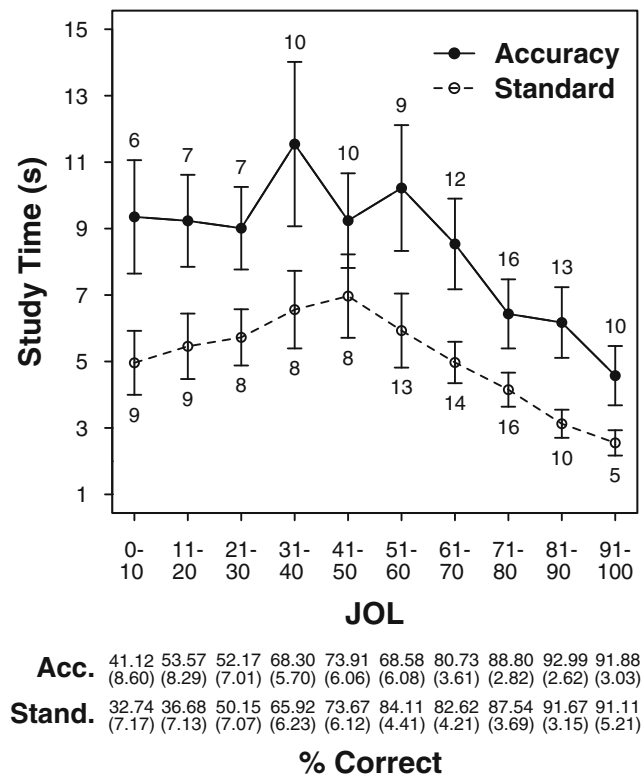
Fig. 5 The graphs show means of median study time (in seconds) for ten judgment of learning (JOL) levels (0–10, 11–20, …, 91–100) for the standard and accuracy groups in Experiment 2. Numbers near the graphs represent mean percentages of items per JOL level. Error bars represent one standard error of the mean. The table below the graphs presents means (and standard deviations) of recall performance (% correct)

similar ST across items with low JOLs in both groups support the DCM but not the RPL.

Experiment 3

It may be premature to conclude that the DCM best explains curvilinear JOL-ST relations. In the previous experiments, differences in item difficulty may have been too subtle for participants to quickly identify the most challenging items. Moreover, prior research suggested that strategy changes are particularly likely when the level of motivation is decreased in comparison to a standard condition (Ackerman, 2014) and that time pressure decreases motivation (e.g., Thiede & Dunlosky, 1999). Therefore, in Experiment 3, we used materials with a wider range of difficulty and examined the effects of time pressure on JOL-ST relations. Finally, we used a different population in order to verify the generalizability of our results.

In the time pressure condition, we imposed a severe constraint on the overall study time available. The RPL predicts that time pressure promotes stopping quickly when the perceived learning progress is slow (i.e., raises the stopping criterion) and thus increases curvilinearity in JOL-ST relations. In contrast, the DCM predicts that time pressure lowers the time limit. If the time limit remains considerably above the time invested in high JOL items, curvilinearity remains unchanged. Alternatively, if the time limit gets close to the time invested in high JOL items, ST variability and curvilinearity decrease.

Method

Participants and materials Sixty Technion undergraduates were randomly assigned to the standard group or to the time pressure group. Stimuli were 60 item-category pairs with very large differences in difficulty used before by Ackerman (2014), Experiment 1. The easiest pairs were known concepts and categories with known associations (e.g., CUBA–ISLAND). The most difficult items were unknown concepts associated with unknown categories (e.g., ERUCARIA–CRUCIFERAE). Two additional pairs were used for practice.

Procedure The procedure was identical to that of Experiment 2 with the following exceptions. The standard group was instructed to allocate their time (30 min) as they wished. The time pressure group was told that they had only 6 min for study. Instructions explained that they must manage their time efficiently and study about 12 items per minute. The 6-min time frame corresponded to 75% of mean overall ST in Ackerman's (2014) Experiment 1. The experimenter indicated when 3 min and 1 min remained. Before the cued recall test, participants worked on a 60-s filler task (numeric puzzle).

Results and discussion

The time pressure group studied shorter than the standard group, while JOLs and recall did not differ between groups (Table 1). Figure 6 reveals curvilinear JOL-ST relations in both conditions. Notably, JOL-ST relations in the standard groups of Experiment 3 and Experiment 2 (see Fig. 5) were very similar despite differences in materials and populations.

As before, we regressed ST in each group on JOL and JOL^2. Results revealed negative JOL-ST slopes and curvilinear JOL-ST relations in both groups (Table 2).
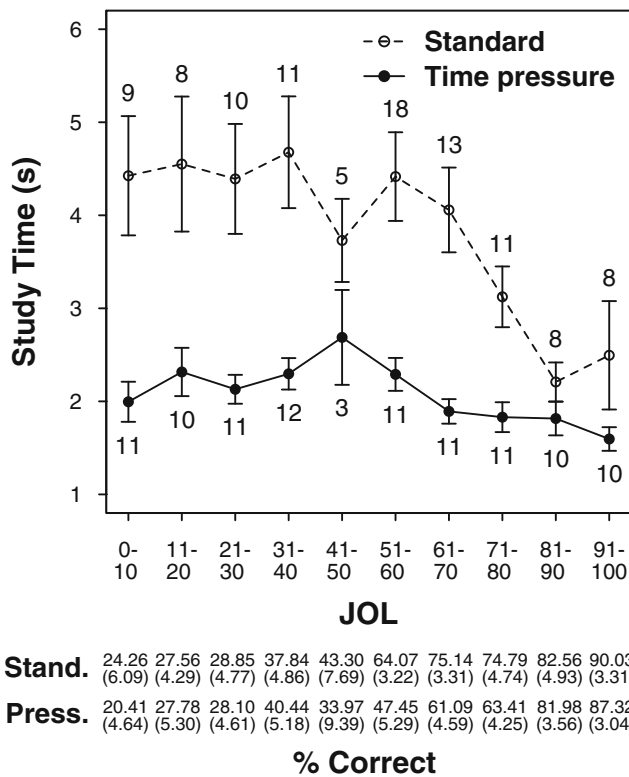
Fig. 6 The graphs show means of median study time (in seconds) for ten judgment of learning (JOL) levels (0–10, 11–20, …, 91–100) for the standard and time pressure groups in Experiment 3. Numbers near the graphs represent mean percentages of items per JOL level. Error bars represent one standard error of the mean. The table below the graphs presents means (and standard deviations) of recall performance (% correct)

across items with low JOLs in both groups support the DCM but not the RPL.

General discussion

In this study, we addressed the puzzle of time allocation for the most challenging items in learning tasks. Our starting points were moderately negative correlations between JOL and ST that have been often reported in the literature (e.g., Ackerman, 2014; Koriat et al., 2006; Undorf & Erdfelder, 2013) and the less common finding of inverted U-shaped relations between JOL and ST (e.g., Metcalfe & Kornell, 2005).

We compared two models (RPL and DCM) that may account for curvilinear JOL-ST relations and make different predictions about how motivation affects curvilinearity. Three experiments revealed robust curvilinear JOL-ST relations. Neither increasing motivation to master a subset of items (Experiment 1) nor manipulating motivation between subjects (Experiment 2) affected curvilinearity. Severe time pressure decreased, but did not eliminate curvilinearity (Experiment 3). These findings are inconsistent with the predictions of the RPL. However, they are compatible with the DCM. Importantly, this was true both when the DCM predicted unchanged curvilinearity (Experiment 1 and Experiment 2) and differences in curvilinearity (Experiment 3).

Much research has shown that, compared with extrinsic cues such as motivation and the available time, intrinsic cues such as item difficulty have stronger effects on metacognition (e.g., Koriat, 1997; Rabinowitz, Ackerman, Craik, & Hinchley, 1982). Our finding of robust curvilinear JOL-ST relations suggests that item difficulty also determines the way of ST allocation. Easy items trigger bottom-up data-driven ST allocation that is governed by processing fluency. In contrast, challenging items trigger top-down strategic ST allocation that takes extrinsic cues into account.

Finally, our results nicely converge with findings from the metacognition of problem solving (Ackerman, 2014) and reveal previously unrecognized parallels between time allocation in memorizing and more complex problem-solving tasks. It remains to be seen whether curvilinearity in judgment-time relations can also inform theorizing about metacognition in other domains, such as decision making or problem solving (meta-reasoning; Ackerman & Thompson, 2015).

Allowing for variance in slope and curvilinearity across participants did not improve model fit, $\chi^2_{standard} < 1$, $\chi^2(2)_{time\ pressure} = 3.14$, $p = .208$. When testing for differences across groups, a significant condition × JOL interaction, $t(3527) = 4.83$, $p < .001$, demonstrated a reduced JOL-ST slope in the time pressure group. A significant condition × JOL^2 interaction, $t(3543) = 2.83$, $p = .005$, demonstrated attenuated curvilinearity in the time pressure group. Bayesian analyses indicated positive evidence for attenuated curvilinearity under time pressure, posterior probability of the full model with condition interaction terms = .86.

ST for the two lowest JOL bins (Table 3) was shorter in the time pressure group than in the standard group, $F(1, 52) = 18.42$, $p < .001$, $\eta_p^2 = .26$. Neither the main effect of JOL bin, $F < 1$, nor the interaction, $F(1, 52) = 2.56$, $p = .115$, were significant.

In sum, time pressure reduced ST but not JOLs or recall. Decreased curvilinearity under time pressure and similar ST

Appendix

Table 4 shows within-person gamma correlations between ST, JOL, recall, and item difficulty in Experiments 1–3.

Across all conditions, JOL-ST correlations were negative, whereas JOL-recall correlations were positive. ST-recall correlations were negative in Experiment 3 and numerically but not significantly negative in Experiments 1 and 2. Item difficulty correlated negatively with JOLs and recall but positively with ST. Results were the same for Pearson correlations between JOL and ST.

Neither correlation differed across conditions, except that, in Experiment 3, the JOL-ST correlation was more pronounced in the standard than in the time pressure group.

Table 4 Means (and standard deviations) of gamma correlations in Experiments 1–3

	G(JOL, ST)	G(JOL, recall)	G(ST, recall)	G(JOL, difficulty)	G(ST, difficulty)	G(recall, difficulty)
Experiment 1						
1-point items	−.13*** (.19)	.48*** (.33)	−.08 (.30)	−.76*** (.24)	.29*** (.28)	−.73*** (.30)
5-point items	−.17*** (.23)	.54***[b] (.35)	−.04[b] (.35)	−.75*** (.28)	.37*** (.31)	−.72***[b] (.47)
Experiment 2						
Standard	−.19*** (.17)	.52***[c] (.26)	−.11[c] (.29)	−.50*** (.58)	.30*** (.37)	−.50***[c] (.68)
Accuracy	−.22*** (.23)	.45*** (.26)	−.05 (.31)	−.52*** (.49)	.33*** (.43)	−.52*** (.54)
Experiment 3						
Standard	−.19*** (.14)	.56*** (.19)	−.12** (.19)	−.47*** (.14)	.24*** (.13)	−.59*** (.13)
Time Pressure	−.11*** [a] (.16)	.54*** (.17)	−.14*** (.20)	−.44*** (.13)	.18*** (.13)	−.59*** (.14)

Note. JOL = judgment of learning; difficulty = Experiment 1 and Experiment 2: 1 = related; 0 = unrelated; Experiment 3: 5 = concepts, categories, and associations known, 4 = concepts and categories known, associations unknown, 3 = concepts known, categories and associations unknown, 2 = categories known, concepts and associations unknown, 1 = concepts, categories, and associations unknown. Correlations could not be calculated for some participants in Experiment 1 and Experiment 2 due to insufficient variability in recall performance

**p < .01. ***p < .001

[a] p < .05, for differences between conditions

[b] df = 46

[c] df = 27

References

Ackerman, R. (2014). The diminishing criterion model for metacognitive regulation of time investment. *Journal of Experimental Psychology: General, 143*(3), 1349–1368. doi:10.1037/a0035098

Ackerman, R., & Thompson, V. A. (2015). Meta-reasoning: What can we learn from meta-memory? In A. Feeney & V. A. Thompson (Eds.), *Reasoning as memory* (pp. 164–182). Hove: Psychology Press.

Bates, D. M., Maechler, M., & Bolker, B. (2015). lme4: Linear mixed-effects models using S4 classes (R package version 1.1-6) [Software].

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences.* Mahwah, NJ: Erlbaum.

Core Team, R. (2015). *R: A language and environment for statistical computing [Software]*. Vienna: R Foundation for Statistical Computing.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349–370. doi:10.1037/0096-3445.126.4.349

Koriat, A., Ma'ayan, H., & Nussinson, R. (2006). The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior. *Journal of Experimental Psychology: General, 135*(1), 36–69. doi:10.1037/0096-3445.135.1.36

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package) (R package version 2.0-6) [Software].

Little, J. L., & McDaniel, M. A. (2015). Metamemory monitoring and control following retrieval practice for text. *Memory & Cognition, 43*(1), 85–98. doi:10.3758/s13421-014-0453-7

Mazzoni, G., & Cornoldi, C. (1993). Strategies in study time allocation: Why is study time sometimes not effective? *Journal of Experimental Psychology: General, 122*(1), 47–60. doi:10.1037//0096-3445.122.1.47

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18*(3), 159–163. doi:10.1111/j.1467-8721.2009.01628.x

Metcalfe, J., & Finn, B. (2008). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review, 15*(1), 174–179. doi:10.3758/PBR.15.1.174

Metcalfe, J., & Kornell, N. (2005). A region of proximal learning model of study time allocation. *Journal of Memory and Language, 52*(4), 463–477. doi:10.1016/j.jml.2004.12.001

Morey, R. D., & Rouder, J. N. (2015). BayesFactor (R package version 0.9-12) [Software].

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect". *Journal of Experimental Psychology Learning Memory and Cognition, 14*(4), 676–686. doi:10.1037/0278-7393.14.4.676

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 26, pp. 125–173). New York: Academic Press.

Pelegrina, S., Bajo, M. T., & Justicia, F. (2000). Differential allocation of study time: Incomplete compensation for the difficulty of the materials. *Memory, 8*(6), 377–392. doi:10.1080/09658210050156831

Rabinowitz, J. C., Ackerman, B. P., Craik, F. I. M., & Hinchley, J. L. (1982). Aging and metamemory: The roles of relatedness and imagery. *Journal of Gerontology, 37*(6), 688–695. doi:10.1093/geronj/37.6.688

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25,* 111–163. doi:10.2307/271063

Thiede, K. W., & Dunlosky, J. (1999). Toward a general model of self-regulated study: An analysis of selection of items for study and self-paced study time. *Journal of Experimental Psychology Learning Memory and Cognition, 25*(4), 1024–1037. doi:10.1037/0278-7393.25.4.1024

Undorf, M., & Erdfelder, E. (2011). Judgments of learning reflect encoding fluency: Conclusive evidence for the ease-of-processing hypothesis. *Journal of Experimental Psychology Learning Memory and Cognition, 37*(5), 1264–1269. doi:10.1037/a0023719

Undorf, M., & Erdfelder, E. (2013). Separation of encoding fluency and item difficulty effects on judgements of learning. *The Quarterly Journal of Experimental Psychology, 66*(10), 2060–2072. doi:10.1080/17470218.2013.777751