

Influence of the physical effort of reminder-setting on strategic offloading of delayed intentions

Journal:	<i>Quarterly Journal of Experimental Psychology</i>
Manuscript ID	QJE-STD-23-140.R1
Manuscript Type:	Standard Article
Date Submitted by the Author:	26-Jul-2023
Complete List of Authors:	Chiu, Gavin; University College London, Institute of Cognitive Neuroscience Gilbert, Sam; University College London,
Keywords:	memory, effort, metacognition, offloading, reminder

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Influence of the physical effort of reminder-setting on strategic offloading of delayed intentions

Gavin Chiu & Sam J. Gilbert

Institute of Cognitive Neuroscience, University College London, UK

Author note

Data and code to reproduce all analyses, along with the original study pre-registrations can be found at <https://osf.io/qky7j/>

Abstract

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Intention offloading involves the use of external reminders such as diaries, to-do lists, and digital alerts to help us remember delayed intentions. Recent studies have provided evidence for various cognitive and metacognitive factors that guide intention offloading, but little research has investigated the physical cost of reminder-setting itself. Here we present two pre-registered experiments investigating how the cost of physical effort associated with reminder-setting influences strategic intention offloading under different levels of memory load. At all memory loads, reminder-setting was reduced when it was more effortful. The ability to set reminders allowed participants to compensate for the influence of memory load on accuracy in the low-effort condition; this effect was attenuated in the high-effort condition. In addition, there was evidence that participants with less confidence in their memory abilities were more likely to set reminders. Contrary to prediction, physical effort had the greatest effect on reminder-setting at intermediate memory loads. We speculate that the physical costs of reminder-setting might have the greatest impact when participants are uncertain about their strategy choice. These results demonstrate the importance of physical effort as one of the factors relevant to cost-benefit decision-making about cognitive offloading strategies.

Keywords: memory, effort, metacognition, offloading, reminder

1
2
3 Human cognition often relies on the body and external environment to achieve
4 competent behaviour. We point with our fingers to indicate locations, tilt our heads when
5 facing rotated objects (Risko et al., 2014), and arrange objects so that we remember where
6 they were put (Berry et al., 2019). These are examples of *cognitive offloading*, a term defined
7 by Risko and Gilbert (2016) as physical action used to alter the information processing
8 requirements of a task to reduce cognitive demand.
9
10
11
12
13
14
15

16
17 A subset of cognitive offloading concerns the way we remember and execute
18 intentions. Typically, intentions cannot be fulfilled immediately, and there exists an interval
19 between the formation of an intention in our mind and its execution, which is why we often
20 need cues from the environment to trigger retrieval of stored intentions (Crystal & Wilson,
21 2015). The process of creating an artificial cue in the external environment to trigger a
22 delayed intention is known as *intention offloading* (Gilbert et al., 2022). This artificial cue
23 can come in different forms. It can be an organisation of the environment, e.g., when
24 intending to post a letter the next time we leave the house, we might place the letter close to
25 the door. It can be created on other individuals, e.g., asking your colleague to remind you of a
26 meeting tomorrow. But most commonly, we offload intentions by using external props and
27 tools as reminders, such as to-do lists, schedule books, sticky notes, and digital reminders. As
28 a result of offloading, we store a relevant cue in our extended physical and social
29 environments to minimise the chance of memory failure (Gilbert et al., 2022).
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45

46 Intention offloading relies on a process of cost-benefit decision-making (Gilbert,
47 2023; Gilbert et al., 2020). While reminders are undoubtedly helpful, it is not possible to set a
48 reminder for every intention we have. Because of the time and effort it takes to offload an
49 intention, we often have to decide when to use reminders and what to use them for.
50
51 Consistent with this, people are more likely to set reminders for intentions when they are
52 associated with a higher financial reward (Dupont et al., 2021).
53
54
55
56
57
58
59
60

1
2
3 Recent research has begun to investigate other factors that influence this evaluation.
4
5 These include task difficulty (including both memory load and the presence of interruption in
6
7 the ongoing task; Gilbert, 2015a), strategy perseveration (Scarampi & Gilbert, 2020), and
8
9 cognitive effort avoidance (Sachdeva & Gilbert, 2020). Cognitive offloading strategies are
10
11 influenced by metacognitive judgements (i.e., people tend to set reminders when they *think*
12
13 that they will otherwise forget, not necessarily when they will *actually* forget; Gilbert, 2015b;
14
15 Gilbert et al., 2020; Boldt & Gilbert, 2019, 2022) and are associated with age-related change
16
17 (Bulley et al., 2020; Redshaw et al., 2018; Scarampi & Gilbert, 2021; Tsai et al., in press).
18
19 For a review of these findings see Gilbert et al. (2022).
20
21
22

23
24 Gilbert (2023) recently proposed that the process of cognitive offloading can be seen
25
26 as a form of value-based decision-making, balancing the benefits of offloading (e.g., the
27
28 increased likelihood of remembering and the ability to reallocate memory capacity to other
29
30 information) against its costs (e.g., the time and physical activity required to set up a
31
32 reminder). However, these costs have not yet been systematically explored in an experimental
33
34 setting. To investigate this, the present study adapted a previous task used to investigate
35
36 reminder-setting strategies (Gilbert et al., 2020) and manipulated both the number of
37
38 intentions (i.e., memory load) and the physical effort it took to set a reminder. Our main
39
40 interest was whether the physical effort of reminder-setting affected offloading strategies, and
41
42 if so whether this effect varied between different memory loads. Six pre-registered
43
44 hypotheses (<https://osf.io/qky7j/>) corresponding to three groups of analyses were proposed in
45
46 this study:
47
48
49

50
51 *Analysis 1: Task Performance*
52

53 Hypothesis 1: Accuracy will decrease at higher memory loads.
54

55 Hypothesis 2: Accuracy will increase when participants are allowed to set reminders.
56
57

58 *Analysis 2: Offloading Behaviour*
59
60

1
2
3 Hypothesis 3: When given a free choice, participants will set more reminders at
4
5 higher memory loads.
6

7
8 Hypothesis 4: Participants will set fewer reminders when this requires more physical
9
10 effort.
11

12 *Analysis 3: Metacognitive Judgement*
13

14
15 Hypothesis 5: Participants will be underconfident in their predicted accuracy, relative
16
17 to their actual performance. This is based on findings from previous studies using
18
19 related paradigms (Gilbert et al., 2020; Kirk et al., 2021; Engeler & Gilbert, 2020).
20

21
22 Hypothesis 6: Participants with lower confidence in their unaided memory ability will
23
24 be more likely to set reminders (Boldt & Gilbert, 2019; Gilbert, 2015b; Kirk et al.,
25
26 2021).
27

28
29
30 In addition to these main hypotheses, we conducted exploratory analyses to
31
32 investigate whether the influence of physical effort differs between memory loads. We
33
34 predicted that physical effort might suppress reminder-setting more strongly at lower memory
35
36 loads (when reminders are less necessary) than at higher memory loads. We also tested
37
38 whether task performance in unaided trials correlated with offloading behaviour when
39
40 reminders were allowed. This correlation would suggest that people with weaker objective
41
42 memory abilities are more inclined to use reminders to support their memory.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Experiment 1

Methods

Before commencing data collection, all experimental hypotheses, design, procedures, and analyses were pre-registered at <https://osf.io/qky7j/>. All deviations from this pre-registered plan are clearly labelled.

Participants

A total of 52 participants were recruited online via Prolific (<https://www.prolific.co>). We excluded outlying participants whose mean target accuracy, averaged across all memory load and offloading conditions, exceeded 2.5 Median Absolute Deviation units from the full sample (Chen et al., 2021). Three participants were excluded based on this criterion and were replaced with another three participants. The final sample consisted of 52 participants (28 male; 24 female; mean age: 25 years; range: 19-41 years).

The above sample size was estimated by power calculation performed using G*Power 3.1. We wanted to power our study to detect the effect of changing physical effort on individual offloading behaviour. Since no previous literature, to our knowledge, has investigated this effect, we predicted our effect size based on a similar study conducted by Sachdeva and Gilbert (2020). In their study, they used a similar experimental paradigm to demonstrate that providing financial reward as an incentive to use internal memory significantly reduced reminder-setting with an effect size of $d=0.51$. We proposed that increasing physical effort associated with reminder-setting and providing financial reward could have a main effect of comparable magnitude in reducing reminder-setting. We therefore decided to collect a sample so that it would have sufficient power to detect an effect of this size in a within-subjects design ($dz=0.51$), with a two-tailed test, alpha set to 0.05 and a power of 95%. The projected sample size required for this study was $n=52$.

1
2
3 The entire experiment was advertised to last approximately 25 minutes. Each
4 participant was paid £3 after completing the experiment. To incentivise task performance,
5 participants were also told that they would receive an additional £1 bonus if they performed
6 in the top 50% of the entire sample. All participants provided informed consent before
7 participation, and the study was approved by the UCL (University College London) Research
8 Ethics Committee (1584/003).
9
10
11
12
13
14
15
16
17
18

19 *Intention Offloading Task*

20
21 The task used in this study was programmed in Java using Google Web Toolkit
22 version 2.7.0 (<http://www.gwtproject.org>) implemented in the Eclipse development
23 environment version 4.14.0 (<https://www.eclipse.org>). To view the experiment, visit
24 <http://samgilbert.net/demos/GC1>. The code that runs the entire experiment can be found by
25 visiting <https://github.com/MetaOffloading/GC1>.
26
27
28
29
30
31
32

33 See Figure 1 below for a schematic illustration of the task. Participants performed a
34 web-based intention offloading task individually on their computers or tablets but were not
35 permitted to use smartphone devices. The task was modified from the optimal reminders
36 paradigm developed by Gilbert et al. (2020). At the beginning of each trial, participants
37 viewed six numbered yellow circles randomly positioned inside a square on the screen. They
38 were instructed to use their mouse (or their finger if performed on a tablet) to drag the circles
39 to the bottom of the square in ascending numerical order (1, 2, 3, etc.). Each time a circle was
40 dragged to the bottom, it would disappear, and a new circle following the sequence would
41 replace the dragged circle in its original location, e.g., after circle 1 was dragged to the
42 bottom, circle 7 would appear in circle 1's original position.
43
44
45
46
47
48
49
50
51
52
53
54

55 During the ongoing task of circle dragging, participants were provided with delayed
56 intentions. Circles that appeared from circle 7 onwards were special circles. These special
57
58
59
60

1
2
3 circles would appear on screen in different colours corresponding to a specific side of the
4 square (left: blue, right: purple, top: orange). The circles would fade back to yellow after two
5 seconds. As participants were dragging regular circles 1-6, they were tasked to remember the
6 colours of these special circles and drag them to their respective sides upon reaching that
7 number in the sequence. For example, imagine a situation where circle 1 was dragged to the
8 bottom, and circle 7 appeared in purple. After two seconds, circle 7 would fade back to
9 yellow, and the participant would continue dragging circles 2-6 while remembering the
10 special colour associated with circle 7. When the participant finally reached circle 7 in the
11 sequence, they would drag it to the right instead of the bottom because the right side was
12 associated with purple. Hence, a delayed intention was formed when the participant first saw
13 circle 7 in purple. In each trial of the experiment, there were two, four, or six special circles
14 corresponding to three levels of memory load. Since all special circles appeared from circle 7
15 onwards, it meant that each trial contained eight, ten, or twelve circles in total.

16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33 Participants could rely entirely on their unaided memory to perform the task.
34
35 Alternatively, they could offload their intentions by setting reminders. Participants were
36 taught a simple offloading strategy that was integrated into the circle-dragging task. As soon
37 as a special circle appeared on screen, they could immediately drag it close to its target side.
38 This new location would become a perceptual trigger to remind the participant where the
39 circle should go when that number was reached. For instance, when the participant first saw
40 circle 7 in purple, they could drag it towards the right before continuing with circles 2-6.
41
42 Once circle 7 was reached, the participant would know by the circle's location that it should
43 go to the right. By employing this strategy, delayed intentions were offloaded to the external
44 environment as participants no longer had to remember or mentally rehearse the specific
45 colours of the special circles.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 **[Figure 1 about here]**
4
5

6 *Counterbalancing and Effort Manipulation*
7

8
9 The experiment involved manipulations of two key factors – memory load and
10 offloading conditions. Memory load, as mentioned above, was controlled by changing the
11 number of special circles (two, four, or six) presented to the participant. Before each trial,
12 participants were told how many special circles they would get so that they could prepare and
13 formulate a memory strategy. On the other hand, three offloading conditions were presented
14 in total – no reminder (trials where participants were forced to use internal memory), low-
15 effort reminders (trials where reminders were optional and required low physical effort), and
16 high-effort reminders (trials where reminders were optional and required high physical
17 effort).
18
19
20
21
22
23
24
25
26
27
28

29 Each participant performed three blocks of trials in the experiment, corresponding to
30 the three offloading conditions. Each block contained nine trials of the same circle-dragging
31 task. Within the nine trials, three contained two special circles, three contained four special
32 circles, and another three contained six special circles. Trials in each block were presented to
33 the participant in random order. The experiment was randomised in both “reminder
34 condition” and “effort condition” for counterbalancing purposes. Randomising the “reminder
35 condition” meant that some participants were assigned to perform the no-reminder block first,
36 then moved on to the two reminder blocks (“no-reminder first” condition); whereas others
37 performed the two reminder blocks before the no-reminder block (“reminders first”
38 condition). Randomising the “effort condition” meant that within the two reminder blocks,
39 some participants were first taught to use low-effort reminders, then switched to high-effort
40 reminders (“low-effort first” condition); while the rest would do the opposite (“high-effort
41 first” condition). The two reminder blocks were always placed adjacent to one another, to
42 increase the salience of the difference in physical effort.
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 In terms of how physical effort was manipulated, we incorporated a physically
4 demanding clicking task after each reminder was set in the high-effort block (see Figure 1
5 above). Upon moving a special circle close to its target side, a window would pop up on
6 screen and the participant had to click and close the window five times before they could
7 proceed to move the next circle. A real-world analogue of this high-effort reminder might be
8 a smartphone reminder app with a poorly designed user interface that requires users to
9 perform multiple unnecessary clicks to set a reminder. On the contrary, in the low-effort
10 block, participants would not engage in any clicking task after each reminder, and they could
11 freely proceed to the next circle after setting a reminder.
12
13
14
15
16
17
18
19
20
21
22
23
24
25

26 *Procedures*

27
28 Participants began by performing four practice trials in the following order: no special
29 circle, two special circles, four special circles, six special circles. All circles except the next
30 one in sequence were fixed on screen, meaning that participants were forced to rely on their
31 own memory. Afterwards, they were asked to evaluate how accurately they thought they
32 could perform the task at each memory load (i.e., self-predicted accuracy). Since this
33 evaluation took place before any offloading strategy was taught, it reflected participants'
34 metacognitive confidence in their internal memory.
35
36
37
38
39
40
41
42
43

44 Next, the offloading strategy was taught. Participants in the “low-effort first”
45 condition were first taught the low-effort strategy without the additional clicking task,
46 whereas those in the “high-effort first” condition were taught the strategy with the clicking
47 task. All participants performed one practice trial with four special circles using their
48 respective taught strategies.
49
50
51
52
53
54
55

56 Participants in the “no-reminder first” condition then proceeded to perform a block of
57 nine trials of the circle-dragging task. In this block, they were instructed not to use any
58
59
60

1
2
3 offloading strategy. The system also disabled reminder-setting by fixing all circles that were
4 not next in the sequence. Participants were thus required to use internal memory to remember
5 as many special circles as possible. After that, they moved on to the next part of the
6 experiment where reminders were allowed.
7
8
9
10
11

12 For those in the “low-effort first” condition, they performed another block of nine
13 trials, this time with the option to freely choose between setting reminders with the taught
14 low-effort strategy or using internal memory. After completion, they were informed that
15 reminder-setting would now become more effortful and were introduced to a modified
16 strategy with the additional clicking task. They then performed another practice with four
17 special circles, and advanced to the final block of the experiment where they repeated the
18 previous block but with high-effort reminders.
19
20
21
22
23
24
25
26
27

28 For those in the “high-effort first” condition, procedures were essentially the same,
29 except participants first used the high-effort strategy, then were told reminder-setting would
30 become less effortful and subsequently switched to the low-effort strategy. Similarly,
31 participants in the “reminders first” condition also went through the same procedures but with
32 one difference: they performed the two reminder blocks first, before attempting the no-
33 reminder block. At the end of the experiment, all participants were thanked for their time and
34 paid.
35
36
37
38
39
40
41
42
43
44
45
46

47 *Data Analysis*

48
49 All data were analysed using R version 4.1.2 and RStudio version 2021.09.0. The data
50 and code to reproduce all analyses can be found at <https://osf.io/qky7j/>. Data were analysed
51 by repeated measures analysis of variance (ANOVA), with follow-up pairwise comparisons
52 using paired samples *t*-tests. Results were two-tailed with a significance threshold of $p < .05$.
53
54
55
56
57
58
59
60 Degrees of freedom in the repeated measures ANOVAs were adjusted by a Greenhouse-

1
2
3 Geisser correction where the assumption of sphericity was not met. All analyses were
4
5 conducted as described in our pre-registered plan (<https://osf.io/qky7j/>). In addition, several
6
7 exploratory analyses were performed which were not included in the plan. They will be
8
9 described in the Results section below and clearly labelled as not being pre-registered.
10
11 Independent measures were three levels of memory load (two, four, six special circles) and
12
13 three levels of offloading conditions (no reminder, low-effort reminders, high-effort
14
15 reminders) manipulated in a within-subject manner. Key dependent measures were:
16
17 Target accuracy: the proportion of special circles correctly dragged to their target sides of
18
19 matching colour, instead of to the bottom of the square (adopted from Gilbert, 2015a).
20
21 Externalizing proportion: the proportion of special circles for which the participant set an
22
23 external reminder. In common with previous studies (e.g., Gilbert, 2015a), this was measured
24
25 by calculating the likelihood of dragging a target circle when it was not the next in the
26
27 numerical sequence, then subtracting the equivalent likelihood from nontarget circles. This
28
29 corrects for any general tendency to accidentally click the wrong circle.
30
31 Metacognitive evaluation: a self-predictive measure with a range of 0% to 100% showing
32
33 how confident participants thought they could accurately perform the task without external
34
35 reminders (adopted from Gilbert et al., 2020).
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Results

Target Accuracy

A summary of the target accuracy data is illustrated in Figure 2. We first conducted a 3x3 repeated measures ANOVA with the dependent variable of target accuracy and factors of memory load (two, four, six) and offloading condition (no-reminder, low-effort, high-effort). There was a significant main effect of memory load ($F(1.81, 92.50) = 47.13, p < .001, \eta_p^2 = .48$). We followed up by performing one-way repeated measures ANOVAs and pairwise comparisons between memory loads at each offloading condition. Results are displayed in Table 1. In line with our prediction, significant pairwise differences across all memory loads were observed in the no-reminder condition, showing that participants performed increasingly poorly as memory load increased. The same effect was also observed in participants using high-effort reminders; on the contrary, accuracy differences across memory loads in the low-effort condition remained nonsignificant.

[Figure 2 about here]

[Table 1 about here]

In the same 3x3 ANOVA, a significant main effect of offloading condition was also found ($F(1.73, 88.29) = 38.90, p < .001, \eta_p^2 = .43$). We performed pairwise comparisons between the three offloading conditions at each memory load, as shown in Table 2 below. We predicted that reminder-allowed accuracy, in any effort condition, would be significantly higher than no-reminder accuracy at all memory loads. Results partially supported our hypothesis, with significant differences observed only in memory loads 4 and 6, but not in 2.

1
2
3 Interestingly, at memory load 6, target accuracy in the high-effort condition was significantly
4 lower than that in the low-effort condition. This was evidence that the increase in physical
5 effort associated with setting a reminder negatively influenced participants' target accuracy at
6 high memory loads.
7
8
9
10
11
12
13

14 **[Table 2 about here]**
15
16
17
18
19
20

21 The memory load x offloading condition interaction in this 3x3 ANOVA was
22 significant ($F(3.09, 157.38) = 13.19, p < .001, \eta_p^2 = .21$). We conducted additional 2x3
23 repeated measures ANOVAs as follow-up to examine interactions for each pairwise
24 comparison between two of the three offloading conditions. Interactions were consistently
25 significant across each pair (no-reminder & low-effort: $F(1.84, 93.96) = 36.70, p < .001, \eta_p^2$
26 = .42; no-reminder & high-effort: $F(1.74, 88.64) = 5.03, p = .011, \eta_p^2 = .09$; low-effort &
27 high-effort: $F(1.91, 97.40) = 6.87, p = .002, \eta_p^2 = .12$). These interactions demonstrated that
28 the effect of memory load on target accuracy was greater in the no-reminder condition than
29 the other two reminder-allowed conditions. The effect was also greater in the high-effort
30 condition compared to the low-effort condition.
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 *Externalizing Proportion*

50 Next, we investigated participants' offloading behaviour across memory loads and
51 between the two effort conditions. Figure 3 below shows a summary of our externalizing
52 proportion data. We used a two-factor 2x3 repeated measures ANOVA with the dependent
53 variable of externalizing proportion and factors of memory load (two, four, six) and effort
54 conditions (low-effort, high-effort). As expected, main effect of memory load was significant
55
56
57
58
59
60

1
2
3 ($F(1.88, 95.92) = 38.83, p < .001, \eta_p^2 = .43$). Follow-up pairwise comparisons confirmed
4 significant differences in externalizing proportion across memory loads at each effort
5 condition, with only one exception in the comparison of load 2 vs. 4 in the high-effort
6 condition. Details are shown in Table 3 below. This result provided evidence that people
7 generally set more reminders when memory load was higher.
8
9
10
11
12
13
14
15
16

17 **[Figure 3 about here]**
18
19
20
21

22 **[Table 3 about here]**
23
24
25
26
27

28 We then turned to the key analysis of our study: the effect of increasing physical
29 effort on offloading behaviour. A significant main effect of effort condition was found ($F(1,$
30 $51) = 90.81, p < .001, \eta_p^2 = .64$). Consistent with this, follow-up analyses with paired samples
31 t -tests demonstrated significant differences in externalizing proportion between low-effort
32 and high-effort conditions at all memory loads. The differences are shown in Table 4 below.
33 Results therefore supported our hypothesis: increasing physical effort of reminder-setting
34 significantly reduced offloading behaviour at all memory loads.
35
36
37
38
39
40
41
42
43

44 As part of our exploratory analyses, we further examined the interaction between
45 memory load and effort condition to study if people were more selective in offloading when
46 effort increased. The interaction was significant ($F(1.85, 94.27) = 7.45, p = .001, \eta_p^2 = .13$),
47 suggesting that the reduction in externalizing proportion between low- and high-effort
48 conditions was not constant across memory loads. However, contrary to what we predicted,
49 data did not show a larger reduction in offloading at memory load 2 compared to load 6.
50 Instead, an opposite effect was observed (see Table 4). Our results demonstrated that
51
52
53
54
55
56
57
58
59
60

1
2
3 increasing physical effort had a greater effect on reducing offloading behaviour at higher
4 memory loads. Hence, our exploratory hypothesis concerning effort-induced selectivity was
5 not supported by this interaction analysis.
6
7
8
9
10
11

12 **[Table 4 about here]**
13
14
15

16 *Metacognitive Evaluation*

17
18 Figure 4 below shows participants' mean metacognitive judgements and objective
19 unaided accuracies, i.e., mean performance in the no-reminder condition, at each memory
20 load. The difference between these two measures indicated metacognitive bias, i.e., the extent
21 of underconfidence.
22
23
24
25
26
27
28
29

30 **[Figure 4 about here]**
31
32
33
34

35 A 2x3 repeated measures ANOVA was conducted with a dependent variable of
36 accuracy and factors of memory load (two, four, six) and measure (metacognitive judgement,
37 objective unaided accuracy). Results demonstrated a significant effect of memory load
38 ($F(1.45, 73.80) = 98.60, p < .001, \eta_p^2 = .66$). More importantly, main effect of measure was
39 also significant ($F(1, 51) = 118.78, p < .001, \eta_p^2 = .70$), and follow-up paired samples *t*-tests
40 comparing the two measures at each load all produced significant differences (see Table 5
41 below). Self-predicted accuracies, i.e., metacognitive judgement, were significantly lower
42 than objective unaided accuracies at all memory loads. It was thus evident that participants
43 were generally underconfident in their internal memory abilities. Besides, there was a
44 significant interaction between memory load and measure ($F(1.56, 79.71) = 7.81, p = .002,$
45 $\eta_p^2 = .13$), implying that the extent of underconfidence increased as memory load increased.
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 [Table 5 about here]
4
5
6

7 Our final hypothesis concerning the correlation between confidence and offloading
8 was not supported by our results. Correlation analyses between mean metacognitive
9 judgements/bias and mean externalizing proportion, averaged across all memory load and
10 effort conditions, were first performed. Both correlations were nonsignificant despite being
11 negative (metacognitive judgement: $r(50) = -.16, p = .24$; metacognitive bias: $r(50) = -.09, p$
12 $= .52$). We then followed up by running the same correlations separately at each memory
13 load, averaged across effort conditions. Most of the correlations under these different
14 conditions were also found to be nonsignificant (see Tables 1 and 2 in the online
15 Supplementary Material for details of correlations).
16
17
18
19
20
21
22
23
24
25
26
27
28

29 *Other analyses*

30
31 In our final exploratory hypothesis, we wanted to examine if participants who scored
32 lower in no-reminder trials would set more reminders when they were allowed. A negative
33 correlation between these two variables was therefore predicted. Analysis between mean
34 unaided accuracy and mean externalizing proportion, averaged across all memory load and
35 effort conditions, did reveal a significant negative correlation ($r(50) = -.34, p = .01$).
36
37 However, when we performed the same analysis separately at each memory load and effort
38 condition, correlations became mostly nonsignificant (see Table 3 in the online
39 Supplementary Material). Hence, from the results of this study, there seemed to be a negative
40 relationship between internal memory ability and offloading behaviour only when collapsing
41 across all memory load and effort conditions, but this relationship was weak and
42 nonsignificant when looking individually at each condition. We note that the original power
43 calculation was based on the effect of physical effort on reminder setting, not these
44 exploratory analyses, hence the study may have been underpowered to detect correlations.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Discussion

This experiment demonstrated clear influence of physical effort on intention offloading. We had multiple predictions prior to the investigation, most of which were supported by our results. First, in situations where reminders were not allowed or were effortful, task performance was found to be negatively associated with memory load. The drop in accuracy was smaller in the high-effort condition compared to the no-reminder condition (see effect sizes in Table 1) but was nevertheless significant. These findings were in line with existing literature showing a similar negative effect of memory load on target accuracy (Gilbert, 2015a; Redshaw et al., 2018; Scarampi & Gilbert, 2021). Two additional improvements were made in our experiment with regard to methodology: we manipulated three levels of memory load (two, four, six targets) instead of two levels in previous studies (one-target vs. three-targets) (Gilbert, 2015a; Redshaw et al., 2018; Scarampi & Gilbert, 2021) to demonstrate a decreasing trend of accuracy as memory load increased; we also embedded intentions *during* rather than *before* the circle-dragging task (Gilbert, 2015a; Redshaw et al., 2018; Scarampi & Gilbert, 2021) to create greater real-world resemblance where intentions are usually formed and held as we engage in ongoing tasks at hand.

Notably, at the low-effort condition, there was no statistically significant change in accuracy across memory loads (see Table 1, Figure 2). This suggests that when physical effort was low, the availability of reminders enabled compensation for the increase in task difficulty. Participants were aware of the benefits brought by setting reminders and therefore managed to achieve similar levels of accuracy by strategically using more reminders at higher memory loads. This observation was further substantiated by the increase in externalizing proportion as memory load increased (see Figure 3). By contrast, the decrease in accuracy across loads in the high-effort condition suggests that participants failed to fully adjust for

1
2
3 task difficulty when reminders became more effortful. It was also found that at load 6,
4 accuracy in the high-effort condition was significantly lower than that in the low-effort
5 condition. These findings hinted that increasing physical effort associated with reminder-
6 setting made people more reluctant to offload and led to poorer task performance. This effect
7 was especially pronounced at higher memory loads.
8
9

10
11
12
13
14
15 Moreover, we anticipated that the option to set reminders, regardless of effort, would
16 improve accuracy at all memory loads. Our results showed that reminders improved target
17 accuracies at memory loads 4 and 6, but the same could not be observed at load 2 (see Table
18 2). Participants achieved near-perfect accuracies at this load even when reminders were not
19 allowed (mean = .96, SD = .10), meaning that they could easily remember the two special
20 circles without a need for external reminders. However, as memory load increased, using
21 reminders did improve task performance. This finding is consistent with past literature
22 showing the benefits of reminders or other external tools in augmenting our memory (Gilbert
23 et al., 2020; Storm & Stone, 2015).
24
25
26
27
28
29
30
31
32
33
34

35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Participants generally set more reminders as memory load increased (see Table 3, Figure 3), which was in line with our hypothesis as well as previous studies (Gilbert, 2015a; Scarampi & Gilbert, 2021). This was evidence that participants were strategic in their use of external aid to compensate for increases in task difficulty. However, notably at the high-effort condition, a significant increase in externalizing proportion was only observed from loads 4 to 6, but not from loads 2 to 4 (see Table 3). In other words, higher effort reminders compromised the compensatory offloading of intentions at low memory loads, and participants were more likely to change their offloading strategy, i.e., set more reminders, when they faced six special circles. This finding suggests that people might be more resistant to increase offloading in response to task difficulty when reminders demanded high effort.

1
2
3 Our main line of investigation concerns how increasing physical effort influences
4 intention offloading. As predicted, high-effort reminders significantly reduced offloading at
5 all memory loads. The differences in externalizing proportion between low- and high-effort
6 conditions were large (see effect sizes in Table 4). This supports our suggestion that cognitive
7 offloading can be seen as a form of cost/benefit or value-based decision-making (Gilbert et
8 al., 2020; Gilbert, 2023).

9
10
11
12
13
14
15
16
17 In terms of metacognition, this experiment replicated the results of past studies (Boldt
18 & Gilbert, 2019; Engeler & Gilbert, 2020; Gilbert, 2015b; Gilbert et al., 2020; Grinschgl et
19 al., 2021), demonstrating underconfidence in participants' unaided memory at all three
20 memory loads (see Table 5, Figure 4). The extent of underconfidence, or metacognitive bias,
21 also increased as memory load increased. This finding agrees with the existing literature
22 showing that underconfidence may be more pronounced in more difficult tasks (Cherkaoui &
23 Gilbert, 2017; Gilbert, 2015b). However, the predicted negative correlation between
24 confidence and reminder-setting was not obtained, at least when averaged over memory loads
25 and effort conditions. It is unclear whether this reflects a difference between the present
26 paradigm and previous ones (Boldt & Gilbert, 2019; Gilbert, 2015b; Gilbert et al., 2020) or
27 simply a lack of power, considering that only 52 participants took part in this study.

28
29
30
31
32
33
34
35
36
37
38
39
40
41
42 Finally, we were interested in how physical effort influences people's *selectivity* in
43 intention offloading. We predicted that more effortful reminders would have a larger effect in
44 reducing offloading at low memory loads, when the need for reminders is least. We
45 demonstrated a significant interaction between memory load and effort condition in our
46 externalizing proportion data, but surprisingly found that participants had greater reductions
47 in offloading when there were six targets compared to when there were two (see Table 4),
48 opposite to our prediction.
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 A possible explanation for this result is that our manipulation of effort was
4
5 confounded with our manipulation of memory load. Since the additional clicking task was
6
7 implemented every time a reminder was set, the effort required at low memory loads was
8
9 smaller (two-targets: 10 clicks) than that at high memory loads (six-targets: 30 clicks). Any
10
11 interaction in our externalizing proportion data could be a result of differences in physical
12
13 effort between memory loads rather than strategic changes in offloading behaviour
14
15 (selectivity). Hence, we cannot conclusively interpret this interaction as an indicator of effort-
16
17 induced selectivity. The purpose of Experiment 2 was to remove this confound and
18
19 investigate whether a fixed difference in physical effort affects the selectivity in intention
20
21 offloading.
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Experiment 2

Experiment 2 aimed to evaluate whether the impact of physical effort differed between memory loads when the manipulation required the same number of clicks at each load, rather than scaling with the total number of targets. This experiment repeated the procedures of Experiment 1 with one key change: the effort manipulation was adjusted by fixing the clicking task in the high-effort condition so that it required 15 clicks, but only for the *first* reminder set on each trial. This meant that the additional effort was constant at each memory load. We predicted that people would more selectively reduce offloading at low memory loads, i.e., two-target condition. By contrast, there would be less reduction at high memory loads, i.e., six-target condition, because of a stronger need for reminders.

Methods

All experimental hypotheses, design, procedures, and analyses were pre-registered at (<https://osf.io/qky7j/>) before data collection. Any deviations from this plan are clearly labelled.

Participants

52 participants were recruited online via Prolific (<https://www.prolific.co>) for consistency with Experiment 1. The earlier experiment found a significant interaction ($p = .001$) in externalizing proportion between memory load and effort condition with a partial eta squared of .13. A sample size of 52 participants would give > 99% power to detect an effect of the same size. Three participants whose mean target accuracy, averaged across all memory load and offloading conditions, exceeded 2.5 Median Absolute Deviation units were excluded and replaced with another three participants. The final sample included 52 participants (27 male; 25 female; mean age: 29 years; range: 19-57 years).

Procedures and Data Analysis

All procedures, analyses, and exclusion criteria were identical to Experiment 1, with the exception that the clicking task in the high-effort condition only appeared upon the first time a reminder was set on that trial, and it required the participant to click and close a window 15 times before they could proceed to the next circle. To view the experiment, visit <http://samgilbert.net/demos/GC2>. The code that runs the entire experiment can be found by visiting <https://github.com/MetaOffloading/GC2>.

Results

Results of target accuracy and metacognitive judgements in this experiment were broadly similar to those in Experiment 1, and they will be displayed first. The main comparison in externalizing proportion between Experiments 1 and 2 will follow.

Target Accuracy

See Figure 5 for a summary of results. As in Experiment 1, a 3x3 repeated measures ANOVA with two factors of memory load (three levels) and offloading condition (three levels) revealed a significant main effect of memory load on target accuracy ($F(1.84, 93.71) = 38.70, p < .001, \eta_p^2 = .43$). Follow-up pairwise comparisons (Table 6) showed mostly significant results in the no-reminder condition, which was in line with findings in Experiment 1 showing that participants' unaided task performance decreased as memory load increased. Whereas Experiment 1 showed a decrease in accuracy at higher memory loads in the high-effort but not low-effort condition, Experiment 2 showed a decrease in both conditions but only in the comparison between two- and six-target trials.

1
2
3 **[Figure 5 about here]**
4
5
6
7

8 **[Table 6 about here]**
9
10
11

12 A significant main effect of offloading condition on target accuracy was also found
13 ($F(1.65, 84.26) = 42.93, p < .001, \eta_p^2 = .46$). Follow-up analyses demonstrated significant
14 differences in accuracy at all loads between no-reminder and low-effort conditions, as well as
15 between low-effort and high-effort conditions, while the comparison between no-reminder
16 and high-effort conditions was only significant at load 6 (Table 7). Participants performed
17 best when using low-effort reminders, compared to when using high-effort reminders or
18 when reminders were not available. The interaction between memory load and offloading
19 condition in our 3x3 ANOVA was also significant ($F(3.01, 153.31) = 13.74, p < .001, \eta_p^2$
20 $= .21$). For each pairwise comparison between two of the three offloading conditions,
21 interactions were significant when the no-reminder condition was involved (no-reminder &
22 low-effort: $F(1.93, 98.32) = 23.48, p < .001, \eta_p^2 = .32$; no-reminder & high-effort: $F(1.81,$
23 $92.10) = 12.46, p < .001, \eta_p^2 = .20$), suggesting that the effect of memory load on accuracy
24 was the greatest in the no-reminder condition. Interaction was not significant in the
25 comparison involving low-effort and high-effort conditions ($F(1.76, 89.79) = .75, p = .46, \eta_p^2$
26 $= .01$).
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48

49 **[Table 7 about here]**
50
51

52 *Metacognitive Evaluation*

53

54 As in Experiment 1, we calculated the metacognitive biases of each participant from
55 their metacognitive judgements and unaided accuracies at each memory load. Figure 6 below
56 illustrates the comparison between the two variables.
57
58
59
60

1
2
3 A 2x3 repeated measures ANOVA with factors of memory load (three levels) and
4
5 measure (two levels) demonstrated significant main effects from both factors (Memory load:
6
7 $F(1.70, 86.85) = 120.98, p < .001, \eta_p^2 = .70$; Measure: $F(1, 51) = 51.48, p < .001, \eta_p^2 = .50$).
8
9
10 Follow-up analyses further showed significant differences between the two measures at all
11
12 loads, which indicated participants' underconfidence in their unaided memory abilities (Table
13
14 8, Figure 6). We also found a significant interaction between memory load and measure
15
16 ($F(1.66, 84.45) = 10.23, p < .001, \eta_p^2 = .17$), as the extent of underconfidence increased as
17
18 memory load increased. These findings replicated the results of Experiment 1.
19
20
21
22
23
24

25 **[Table 8 about here]**

26
27 **[Figure 6 about here]**
28
29
30
31

32
33 Significant negative correlations between metacognitive evaluations and externalizing
34
35 proportion, averaged across all memory load and effort conditions, were found in this
36
37 experiment (metacognitive judgement: $r(50) = -.47, p < .001$; metacognitive bias: $r(50) =$
38
39 $-.34, p = .01$). Follow-up correlation analyses at each memory load also showed mostly
40
41 significant results (see Tables 4 and 5 in the online Supplementary Material for details of all
42
43 correlations). These results were contrasted with findings in Experiment 1, which provided
44
45 inconclusive evidence regarding the relationship between confidence and offloading
46
47 behaviour. Here, data in Experiment 2 supported our original hypothesis: participants who
48
49 were less confident in their memory abilities tended to use more reminders.
50
51
52
53
54
55
56
57
58
59
60

Other analyses

Similar to Experiment 1, this experiment did not find consistently significant correlations between participants' unaided accuracy and externalizing proportion (see Table 6 in the online Supplementary Material for details of all correlations). Averaged across all memory loads and effort conditions there was only a marginally significant effect.

Main Comparison: Externalizing Proportion

With effort now fixed across all memory loads, we wanted to investigate how this modification changes people's strategic decision in using high-effort reminders. The same analyses as in Experiment 1 were run on the externalizing proportion data. A 2x3 repeated measures ANOVA with factors memory load (three levels) and effort condition (two levels) demonstrated a significant main effect of memory load on externalizing proportion ($F(1.69, 86.28) = 50.75, p < .001, \eta_p^2 = .50$). Pairwise comparisons further gave significant differences across all memory loads at each effort condition (see Table 9, Figure 7). We thus replicated findings from Experiment 1, suggesting that participants consistently set more reminders as memory load increased.

[Table 9 about here]

[Figure 7 about here]

[Table 10 about here]

We also replicated the significant main effect of effort condition on externalizing proportion ($F(1, 51) = 38.34, p < .001, \eta_p^2 = .43$), with follow-up analyses confirming significant differences between low- and high-effort conditions at each load (Table 10). Once again, participants offloaded significantly less when reminders became more effortful at all memory loads. However, unlike Experiment 1, we did not find a significant interaction between memory load and effort condition in this experiment ($F(1.91, 97.52) = 2.19, p = .12$,

1
2
3 $\eta_p^2 = .04$). Therefore, the interaction did not support our prediction that higher-effort
4 reminders should selectively decrease offloading at lower memory loads.
5
6

7
8 To further investigate this unexpected result, we looked into the mean differences in
9 externalizing proportion between low- and high-effort conditions at each load, and compared
10 the differences with those in Experiment 1. This investigation was not included in our pre-
11 registered plan. The comparison is displayed in Table 11.
12
13
14
15
16
17

18
19 **[Table 11 about here]**
20
21
22

23
24 A different pattern can be observed from the two experiments. In Experiment 1, the
25 reduction in offloading was larger at loads 4 and 6 (.50 and .46 respectively) and smallest
26 (.29) at load 2. In Experiment 2, where physical effort associated with reminder-setting was
27 fixed, reduction in offloading was largest at load 4 (.34), whereas loads 2 and 6 had similarly
28 small reductions of .21 and .23. To evaluate this statistically, we ran a polynomial contrast for
29 the differences in Experiment 2 and confirmed that the mean differences in externalizing
30 proportion across three loads followed a significant quadratic effect but no linear effect
31 (linear: $t(102) = .27, p = .79$; quadratic: $t(102) = 2.08, p = .04$). In Experiment 1, this was not
32 the case, as the same analysis revealed a highly-significant linear trend, as well as a quadratic
33 one (linear: $t(102) = 2.99, p = .004$; quadratic: $t(102) = 2.44, p = .02$).
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Discussion

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Target accuracy results were similar to those in Experiment 1. When reminders were not allowed, participants' performance in the task decreased as memory load increased. The drop in accuracy was effectively compensated by the use of low-effort reminders. Interestingly, in this experiment, accuracies in the high-effort condition did not show significant differences across memory loads despite still demonstrating a negative trend (Figure 5). This was not the case in Experiment 1, and the contrast could potentially be attributed to how effort was manipulated differently in the two experiments. When effort was unfixed (up to 30 clicks in Experiment 1), participants may be easily discouraged by the high number of additional clicks at high memory loads. Consequently, this could translate into increasingly poor task performance as difficulty increased. On the other hand, when effort was fixed (15 clicks in Experiment 2), participants may be generally more willing to set reminders regardless of changes in memory load, which resulted in more consistent task performance in the high-effort condition.

In terms of metacognitive confidence, we also replicated findings from Experiment 1. Participants demonstrated underconfidence in their unaided memory abilities, and the extent of metacognitive bias increased with memory load, consistent with previous results (Cherkaoui & Gilbert, 2017; Gilbert, 2015b). Furthermore, we demonstrated a significant negative correlation between metacognitive evaluations and externalizing proportion in this experiment, which was not observed in Experiment 1. This correlation provided evidence that individuals with lower confidence in their memory tend to employ more reminders when they are available, as found in previous studies (Boldt & Gilbert, 2019; Gilbert, 2015b; Gilbert et al., 2020).

Turning to externalizing proportion, we established significant main effects of memory load and effort condition on offloading. This supports our initial hypotheses that

1
2
3 people set more reminders when memory load increases, and that offloading is generally
4
5 reduced when reminders become more effortful. Participants were more willing to use high-
6
7 effort reminders when the effort manipulation was fixed (Experiment 2), compared to when it
8
9 was unfixed (Experiment 1). The mean differences between low- and high-effort conditions
10
11 at all memory loads were also smaller in Experiment 2 than in Experiment 1 (Table 11). In
12
13 other words, there was an overall smaller reduction in offloading when physical effort was
14
15 fixed. This further supports the above interpretation that participants readily set more high-
16
17 effort reminders when effort manipulation was not influenced by memory load.
18
19
20

21 The key investigation of this experiment concerns whether a fixed increase in physical
22
23 effort induces selectivity in participants' offloading behaviour in response to changing
24
25 memory loads. Although we predicted that increasing physical effort would most strongly
26
27 reduce offloading at the lowest memory load, results of this experiment did not support this
28
29 hypothesis. If anything, higher effort most strongly reduced offloading behaviour at an
30
31 intermediate memory load (i.e., load 4 in this experiment), but this result should be
32
33 interpreted with caution because the relevant analysis was not pre-registered.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

General Discussion

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

This study was the first to explore the role of physical effort associated with setting a reminder in strategic intention offloading. Past literature has revealed a number of cognitive and metacognitive factors that influence our cost-benefit analysis when making offloading decisions. The present results show that physical effort is another factor that should be taken into consideration.

Experiments 1 and 2 manipulated physical effort of reminder-setting in two different ways: effort was unfixed and influenced by the number of special circles in Experiment 1, whereas Experiment 2 demanded a constant amount of additional effort when setting reminders. Both experiments showed that participants had the poorest task performance when they were not allowed to set reminders, and their performance decreased as memory load increased. The availability of low-effort reminders effectively compensated for the rising task difficulty, and participants were able to consistently achieve near-perfect accuracies with the aid of reminders. Other studies have also demonstrated how the provision of reminders improves prospective memory task performance (Henry et al., 2012; Jones et al., 2021; Maylor, 1990). Our finding in this study further supports the practical effectiveness of intention offloading.

Increasing physical effort, however, negatively impacts this compensatory effect of offloading, which in turn also leads to declining task performance. There were significant reductions in offloading behaviour at all memory loads when reminders became more effortful, and the reductions were even greater when effort manipulation was unfixed. This effect of physical effort as a cost in offloading is comparable to the findings of another study conducted by Grinschgl et al. (2020). In their study, participants engaged in less offloading when using mouse-based (higher effort) compared to touch-based controls (lower effort), and when the reminder interface was less responsive (higher temporal cost). Whereas Grinschgl

1
2
3 and colleagues (2020) investigated the nature of reminder-setting in a working memory task,
4
5 our study demonstrated a similar effect of physical effort in the context of intention
6
7 offloading.
8
9

10 We predicted that the effect of increased physical effort would be most evident at low
11
12 memory loads because the stronger need for reminders at high loads would offset the cost of
13
14 physical effort. To our surprise, Experiment 1 showed an opposite result, as participants had
15
16 the smallest reduction in offloading at load 2. Experiment 2, which removed the confound
17
18 between memory load and effort manipulation by fixing the number of additional clicks, also
19
20 did not provide conclusive evidence for our predicted selectivity. While Experiment 1
21
22 showed a strong linear trend in the mean differences in externalizing proportion across the
23
24 three loads, Experiment 2 did not. This might reflect the confound between memory load and
25
26 the effort manipulation which was present in Experiment 1 but not Experiment 2.
27
28

29
30 Furthermore, both experiments showed a significant quadratic effect with the greatest
31
32 reduction in high-effort reminder-setting at the intermediate memory load. We tentatively
33
34 suggest one possible explanation for this: participants may be more sensitive to the physical
35
36 effort of reminder-setting when they are most unsure about which strategy to use. When the
37
38 memory load was lowest or highest, participants may have been relatively confident in their
39
40 strategy choice – with only two special circles, there was no need to set reminders especially
41
42 since they required extra time and effort; whereas at load 6, participants might always rely on
43
44 reminders because of how difficult the task was. By contrast, they may have been less
45
46 certain of the correct strategy at the intermediate memory load (load 4), making them more
47
48 sensitive to other factors such as physical effort. This hypothesis could be tested in future
49
50 experiments by asking participants to provide confidence ratings about their strategy choices.
51
52
53
54
55

56 The current study adds to the existing evidence showing that intention offloading is a
57
58 delicate process of cost-benefit decision-making. Remembering an intention internally and
59
60

1
2
3 setting (effortful) reminders externally incur respective costs and benefits. The costs of
4 remembering involve an expense of cognitive effort and an increased likelihood to forget, but
5 this has the benefit of saving the time/physical effort spent setting reminders. Offloading, on
6 the contrary, takes more physical effort, but almost guarantees that the intention would be
7 remembered and fulfilled. Deciding which type of effort to expend therefore likely depends
8 on the context of the situation as well as task characteristics. One key factor governing this
9 decision might be the subjective value of intentions. Existing literature has shown that people
10 are more inclined to offload high-value intentions (Dupont et al., 2021). Hence, when an
11 intention is subjectively important to an individual, the costs of forgetting might be too high
12 such that they would resort to external tools even if extra physical effort is demanded.
13
14 Furthermore, a second factor would be the length of time between creating the intention and
15 its execution (Conte & McBride, 2018; McBride et al., 2013). A person would more likely set
16 a reminder for a meeting two weeks later, compared to one that takes place tomorrow. Longer
17 delay periods between intention formation and fulfilment mean that more cognitive effort is
18 required to hold the intention, which again raises the chance of forgetting. In this case, the
19 preference to avoid cognitive effort outweighs the physical effort of setting a reminder, and it
20 would be more reasonable to store the intention externally. In sum, these factors are
21 interesting lines of research that warrant separate studies to investigate how physical effort
22 plays into individual cost-benefit calculations of offloading.

23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47 Finally, understanding the influence of physical effort on intention offloading has
48 direct real-world implications. In the age of technological and digital boom, our cognitive
49 systems are becoming ever more integrated with external tools. These tools demand different
50 levels or types of effort depending on their design and interfaces. Unlike an experimental
51 task, we often have greater flexibility in devising strategies to offload in our everyday lives,
52 and the amount of effort for setting a reminder can be adjusted personally. Understanding the
53
54
55
56
57
58
59
60

1
2
3 factors that influence decisions whether or not to use those tools, and designing the tools in a
4 way that optimises such decisions, can play a positive role in facilitating the adaptive use of
5
6 cognitive technology.
7
8
9

10 11 12 **Supplementary Material**

13
14 The Supplementary Material is available at: qjep.sagepub.com
15
16

17 18 19 **Data Accessibility Statement**

20
21 The data from the present experiment are publicly available at the Open Science Framework
22
23 website: <https://osf.io/qky7j/>
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

References

- Berry, E., Allen, R. J., Mon-Williams, M., & Waterman, A. H. (2019). Cognitive offloading: structuring the environment to improve children's working memory task performance. *Cognitive Science*, 43(8), e12770. <https://doi.org/10.1111/cogs.12770>
- Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. *Cognitive Research: Principles and Implications*, 4(1), 45. <https://doi.org/10.1186/s41235-019-0195-y>
- Boldt, A., & Gilbert, S. J. (2022). Partially overlapping neural correlates of metacognitive monitoring and metacognitive control. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 42(17), 3622–3635. <https://doi.org/10.1523/JNEUROSCI.1326-21.2022>
- Bulley, A., McCarthy, T., Gilbert, S. J., Suddendorf, T., & Redshaw, J. (2020). Children devise and selectively use tools to offload cognition. *Current Biology*, 30(17), 3457–3464.e3. <https://doi.org/10.1016/j.cub.2020.06.035>
- Chen, Z., Song, S., Wei, Z., Fang, J., & Long, J. (2021). Approximating median absolute deviation with bounded error. *Proc. VLDB Endow.*, 14(11), 2114–2126. <https://doi.org/10.14778/3476249.3476266>
- Cherkaoui, M., & Gilbert, S. J. (2017). Strategic use of reminders in an 'intention offloading' task: Do individuals with autism spectrum conditions compensate for memory difficulties?. *Neuropsychologia*, 97, 140–151. <https://doi.org/10.1016/j.neuropsychologia.2017.02.008>
- Conte, A. M., & McBride, D. M. (2018). Comparing time-based and event-based prospective memory over short delays. *Memory (Hove, England)*, 26(7), 936–945. <https://doi.org/10.1080/09658211.2018.1432662>

- 1
2
3 Crystal, J. D., & Wilson, A. G. (2015). Prospective memory: a comparative perspective.
4
5 *Behavioural Processes*, 112, 88–99. <https://doi.org/10.1016/j.beproc.2014.07.016>
6
7
8 Dupont, D., Zhu, Q., & Gilbert, S. (2021). Value-based routing of delayed intentions into
9
10 brain-based vs external memory stores. *PsyArXiv*.
11
12 <https://doi.org/10.31234/osf.io/wq7rs>
13
14
15 Engeler, N. C., & Gilbert, S. J. (2020). The effect of metacognitive training on confidence
16
17 and strategic reminder setting. *PloS One*, 15(10), e0240858.
18
19 <https://doi.org/10.1371/journal.pone.0240858>
20
21
22 Gilbert, S. J. (2015a). Strategic offloading of delayed intentions into the external
23
24 environment. *Quarterly Journal of Experimental Psychology (2006)*, 68(5), 971–992.
25
26 <https://doi.org/10.1080/17470218.2014.972963>
27
28
29 Gilbert, S. J. (2015b). Strategic use of reminders: Influence of both domain-general and task-
30
31 specific metacognitive confidence, independent of objective memory ability.
32
33 *Consciousness and Cognition*, 33, 245–260.
34
35 <https://doi.org/10.1016/j.concog.2015.01.006>
36
37
38 Gilbert, S. J. (2023). Cognitive offloading is value-based decision making: Modelling
39
40 cognitive effort and the expected value of memory. <https://psyarxiv.com/5e8mg/>
41
42
43 Gilbert, S. J., Bird, A., Carpenter, J. M., Fleming, S. M., Sachdeva, C., & Tsai, P. C. (2020).
44
45 Optimal use of reminders: metacognition, effort, and cognitive offloading. *Journal of*
46
47 *Experimental Psychology. General*, 149(3), 501–517.
48
49 <https://doi.org/10.1037/xge0000652>
50
51
52 Gilbert, S. J., Boldt, A., Sachdeva, C., Scarampi, C., & Tsai, P. C. (2022). Outsourcing
53
54 memory to external tools: a review of 'intention offloading'. *Psychonomic Bulletin &*
55
56 *Review*, 10.3758/s13423-022-02139-4. Advance online publication.
57
58 <https://doi.org/10.3758/s13423-022-02139-4>
59
60

- 1
2
3 Grinschgl, S., Meyerhoff, H. S., & Papenmeier, F. (2020). Interface and interaction design:
4
5 how mobile touch devices foster cognitive offloading. *Computers in Human*
6
7 *Behavior, 108*, 106317. [https://doi.org/https://doi.org/10.1016/j.chb.2020.106317](https://doi.org/10.1016/j.chb.2020.106317)
8
9
- 10 Grinschgl, S., Meyerhoff, H. S., Schwan, S., & Papenmeier, F. (2021). From metacognitive
11
12 beliefs to strategy selection: does fake performance feedback influence cognitive
13
14 offloading?. *Psychological Research, 85*(7), 2654–2666.
15
16 <https://doi.org/10.1007/s00426-020-01435-9>
17
18
- 19 Henry, J. D., Rendell, P. G., Phillips, L. H., Dunlop, L., & Kliegel, M. (2012). Prospective
20
21 memory reminders: a laboratory investigation of initiation source and age effects.
22
23 *Quarterly Journal of Experimental Psychology (2006), 65*(7), 1274–1287.
24
25
- 26 Jones, W. E., Bengel, J. F., & Scullin, M. K. (2021). Preserving prospective memory in daily
27
28 life: A systematic review and meta-analysis of mnemonic strategy, cognitive training,
29
30 external memory aid, and combination interventions. *Neuropsychology, 35*(1), 123–
31
32 140. <https://doi.org/10.1037/neu0000704>
33
34
- 35 Kirk, P. A., Robinson, O. J., & Gilbert, S. J. (2021). Trait anxiety does not correlate with
36
37 metacognitive confidence or reminder usage in a delayed intentions task. *Quarterly*
38
39 *Journal of Experimental Psychology (2006), 74*(4), 634–644.
40
41 <https://doi.org/10.1177/1747021820970156>
42
43
- 44 Maylor E. A. (1990). Age and prospective memory. *The Quarterly Journal of Experimental*
45
46 *Psychology. A, Human Experimental Psychology, 42*(3), 471–493.
47
48 <https://doi.org/10.1080/14640749008401233>
49
50
- 51 McBride, D. M., Coane, J. H., Drwal, J., & LaRose, S. A. (2013). Differential effects of delay
52
53 on time-based prospective memory in younger and older adults. *Neuropsychology,*
54
55 *Development, and Cognition. Section B, Aging, Neuropsychology and Cognition,*
56
57 *20*(6), 700–721.
58
59
60

- 1
2
3 Morey, R. (2008). Confidence intervals from normalized data: a correction to Cousineau
4
5 (2005). *Tutorials in Quantitative Methods for Psychology*, 4.
6
7 <https://doi.org/10.20982/tqmp.04.2.p061>
8
9
10 Redshaw, J., Vandersee, J., Bulley, A., & Gilbert, S. J. (2018). Development of children's use
11
12 of external reminders for hard-to-remember intentions. *Child Development*, 89(6),
13
14 2099–2108. <https://doi.org/10.1111/cdev.13040>
15
16
17 Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. *Trends in Cognitive Sciences*,
18
19 20(9), 676–688. <https://doi.org/10.1016/j.tics.2016.07.002>
20
21
22 Risko, E. F., Medimorec, S., Chisholm, J., & Kingstone, A. (2014). Rotating with rotated
23
24 text: a natural behavior approach to investigating cognitive offloading. *Cognitive*
25
26 *Science*, 38(3), 537–564. <https://doi.org/10.1111/cogs.12087>
27
28
29 Sachdeva, C., & Gilbert, S. J. (2020). Excessive use of reminders: metacognition and effort-
30
31 minimisation in cognitive offloading. *Consciousness and Cognition*, 85, 103024.
32
33 <https://doi.org/10.1016/j.concog.2020.103024>
34
35
36 Scarampi, C., & Gilbert, S. J. (2021). Age differences in strategic reminder setting and the
37
38 compensatory role of metacognition. *Psychology and Aging*, 36(2), 172–185.
39
40 <https://doi.org/10.1037/pag0000590>
41
42
43 Storm, B. C., & Stone, S. M. (2015). Saving-enhanced memory: the benefits of saving on the
44
45 learning and remembering of new information. *Psychological Science*, 26(2), 182–
46
47 188. <https://doi.org/10.1177/0956797614559285>
48
49
50 TSAI, P.-C., Scarampi, C., Kliegel, M., & Gilbert, S. (in press). Optimal cognitive
51
52 offloading: Increased reminder usage but reduced pro-reminder bias in older adults.
53
54 *Psychology and Aging*. <https://doi.org/10.1037/pag0000751>
55
56
57
58
59
60

Figure Captions

Figure 1

Schematic representation of the intention offloading task used in the current study (low-effort & high-effort conditions)

Figure 2

Experiment 1: Results of mean target accuracies ($n=52$) across three levels of memory load, separately at each offloading condition. Error bars represent standard errors corrected for the within-subjects design using the method of Morey (2008).

Figure 3

Experiment 1: Results of mean externalizing proportions ($n=52$) across three levels of memory load, separately at each effort condition. Error bars represent standard errors corrected for the within-subjects design using the method of Morey (2008).

Figure 4

Experiment 1: Comparison between metacognitive judgements and objective unaided accuracy ($n=52$), separately at each memory load. Error bars represent standard errors corrected for the within-subjects design using the method of Morey (2008).

Figure 5

Experiment 2: Results of mean target accuracies ($n=52$) across three levels of memory load, separately at each offloading condition. Error bars represent standard errors corrected for the within-subjects design using the method of Morey (2008).

Figure 6

Experiment 2: Comparison between metacognitive judgements and objective unaided accuracy ($n=52$), separately at each memory load. Error bars represent standard errors corrected for the within-subjects design using the method of Morey (2008).

Figure 7

Experiment 2: Results of mean externalizing proportions ($n=52$) across three levels of memory load, separately at each effort condition. Error bars represent standard errors corrected for the within-subjects design using the method of Morey (2008).

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

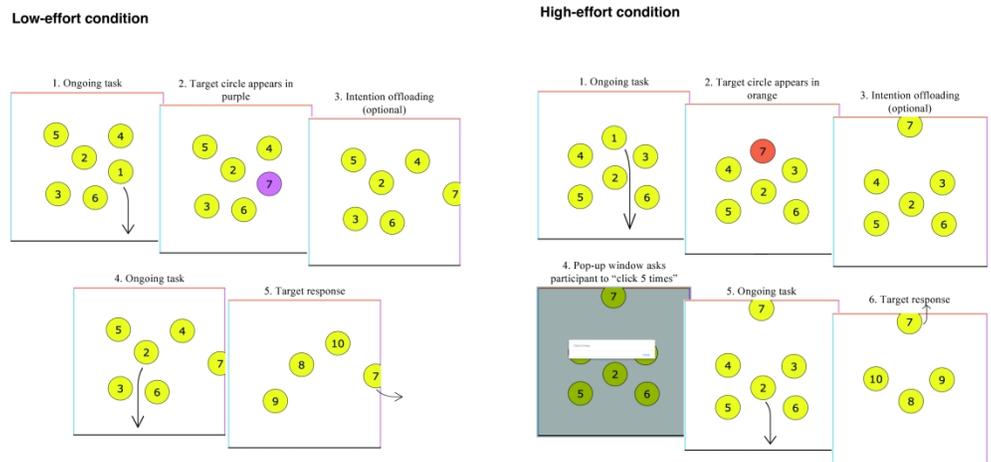


Figure 1. Schematic representation of the intention offloading task used in the current study (low-effort & high-effort conditions)

297x145mm (300 x 300 DPI)

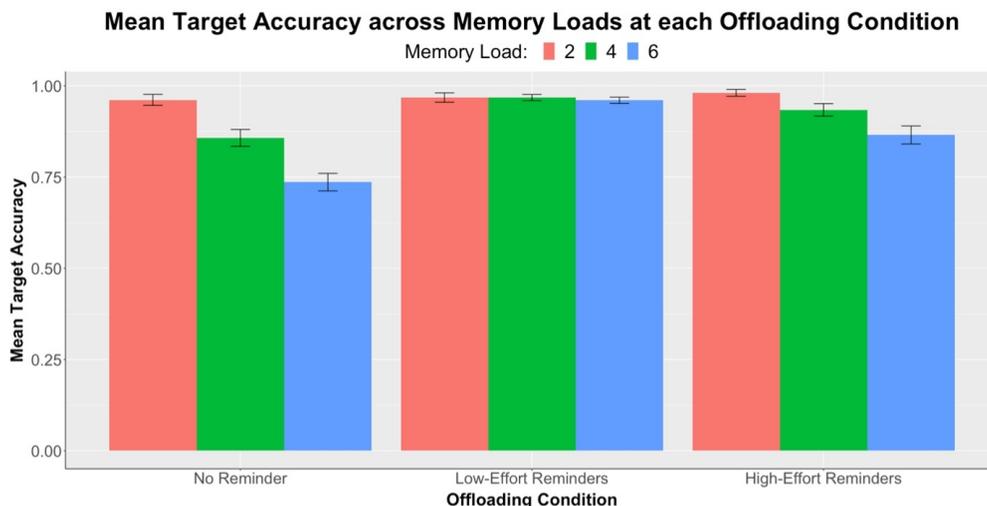


Figure 2. Experiment 1: Results of mean target accuracies (n=52) across three levels of memory load, separately at each offloading condition. Error bars represent standard errors corrected for the within-subjects design using the method of Morey (2008).

159x80mm (220 x 220 DPI)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

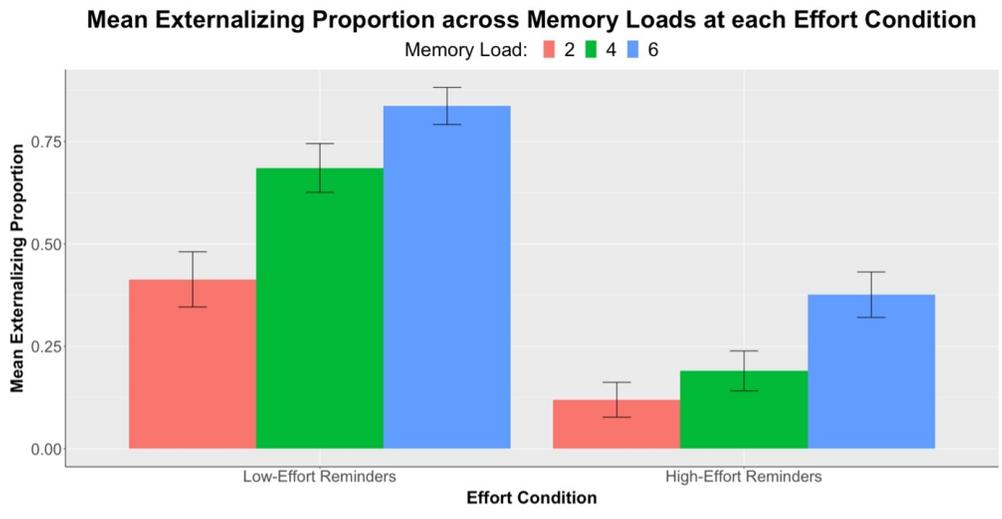


Figure 3. Experiment 1: Results of mean externalizing proportions (n=52) across three levels of memory load, separately at each effort condition. Error bars represent standard errors corrected for the within-subjects design using the method of Morey (2008).

159x80mm (220 x 220 DPI)

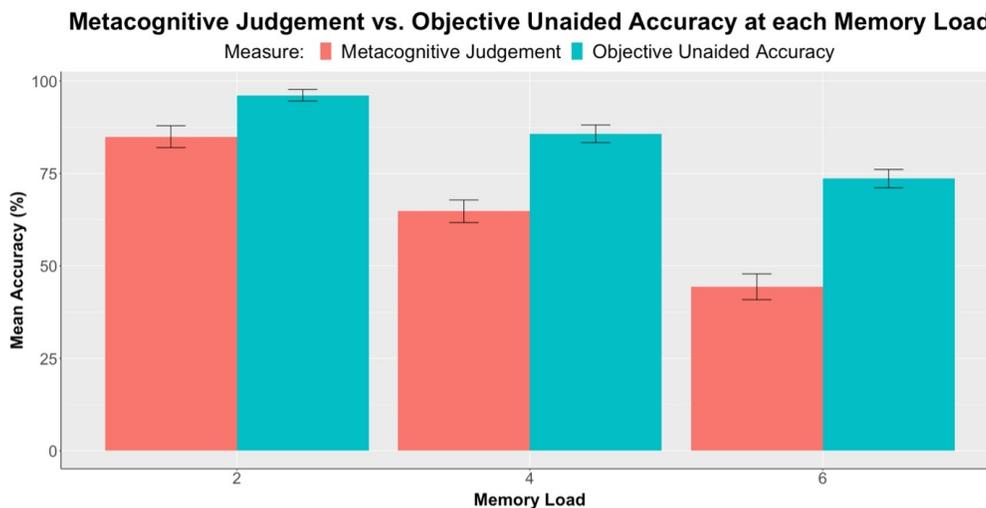


Figure 4. Experiment 1: Comparison between metacognitive judgements and objective unaided accuracy (n=52), separately at each memory load. Error bars represent standard errors corrected for the within-subjects design using the method of Morey (2008).

159x80mm (220 x 220 DPI)

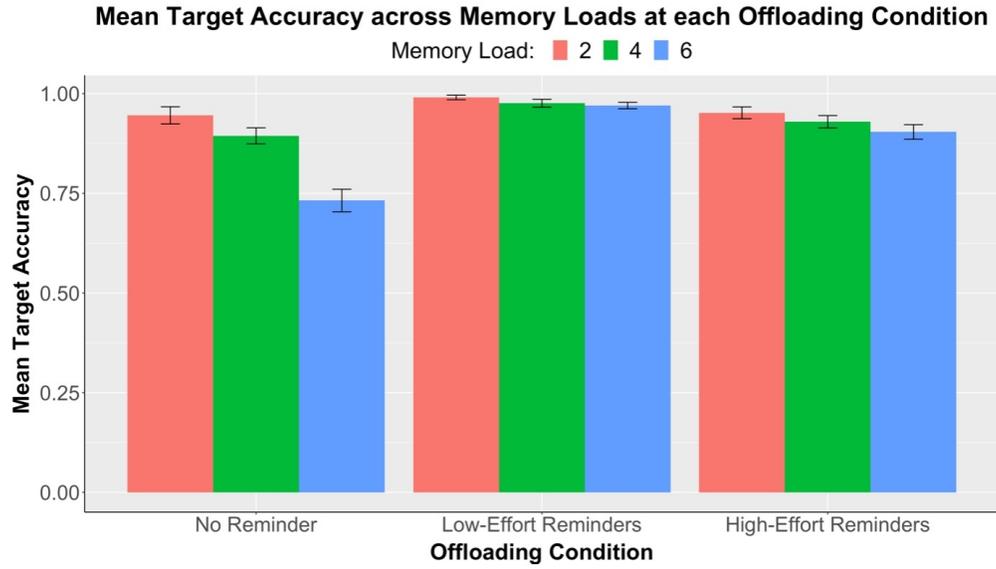


Figure 5. Experiment 2: Results of mean target accuracies (n=52) across three levels of memory load, separately at each offloading condition. Error bars represent standard errors corrected for the within-subjects design using the method of Morey (2008).

159x90mm (220 x 220 DPI)

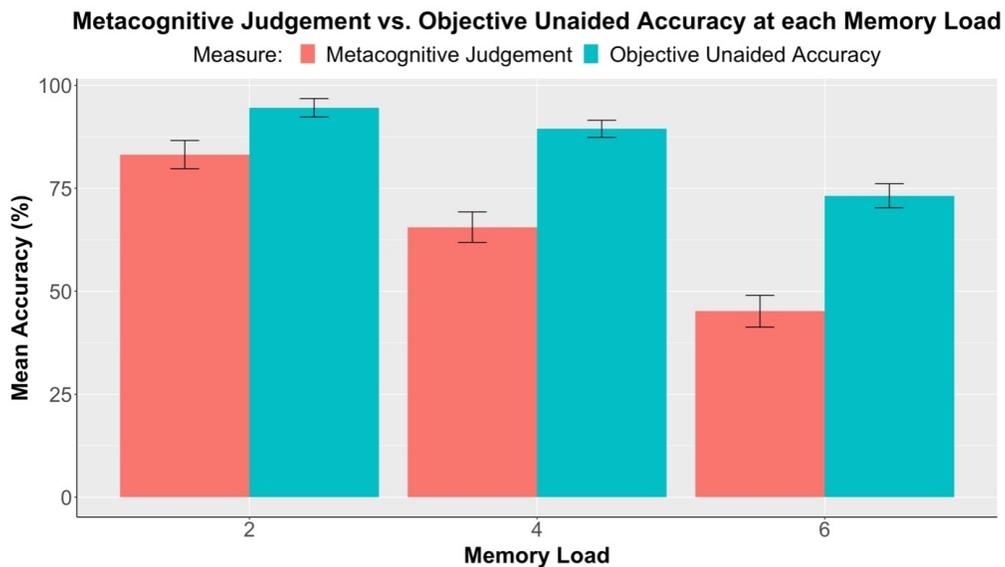
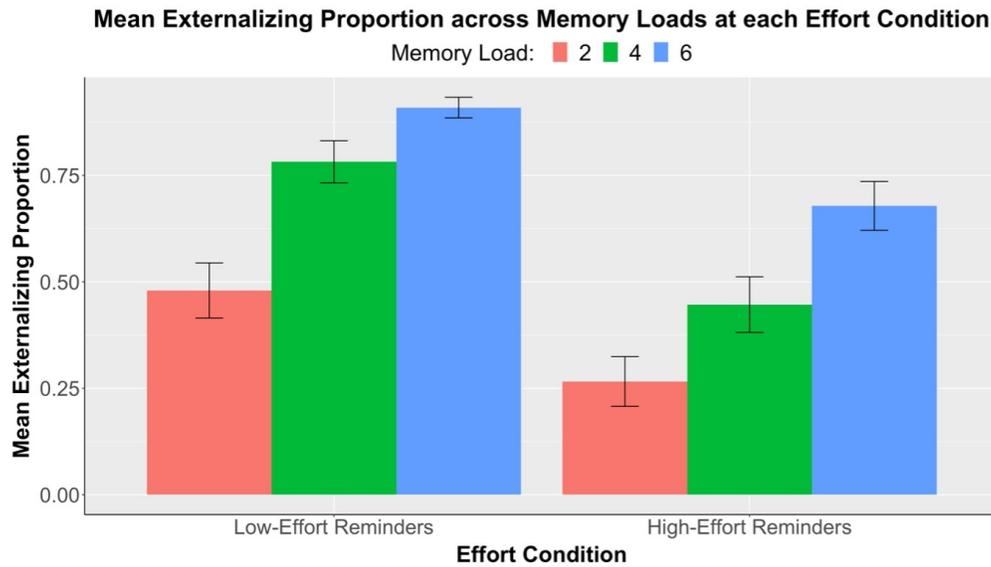


Figure 6. Experiment 2: Comparison between metacognitive judgements and objective unaided accuracy (n=52), separately at each memory load. Error bars represent standard errors corrected for the within-subjects design using the method of Morey (2008).

159x90mm (220 x 220 DPI)



25 Figure 7. Experiment 2: Results of mean externalizing proportions ($n=52$) across three levels of memory
26 load, separately at each effort condition. Error bars represent standard errors corrected for the within-
27 subjects design using the method of Morey (2008).

28
29
30 156x88mm (220 x 220 DPI)

Table 1

Experiment 1: One-way ANOVAs (*F*-values, significance levels, and partial eta-square) and pairwise comparisons (*t*-values, effect sizes, and significance levels) of mean target accuracy between memory loads at each offloading condition. Refer to Figure 2 for mean differences.

		Offloading Condition					
		No reminder		Low-effort reminders		High-effort reminders	
One-way repeated measures ANOVA		$F(1.88, 96.43) = 43.87^{***}$, $\eta_p^2 = .46$		$F(1.79, 91.11) = .21$, $\eta_p^2 = .004$		$F(1.48, 75.67) = 11.64^{***}$, $\eta_p^2 = .186$	
Pairwise comparisons between memory loads		<i>t</i>(51)	<i>d_z</i>	<i>t</i>(51)	<i>d_z</i>	<i>t</i>(51)	<i>d_z</i>
Memory Loads	2 vs. 4	4.91***	.68	< .01	< .01	2.83**	.32
	4 vs. 6	4.97***	.69	.70	.10	2.35*	.33
	2 vs. 6	8.54***	1.18	.53	.07	4.67***	.65

Note: #*p* < .1. **p* < .05. ***p* < .01. ****p* < .001.

Table 2

Experiment 1: Pairwise comparisons (*t*-values, effect sizes, and significance levels) of mean target accuracy between offloading conditions at each memory load. Refer to Figure 2 for mean differences.

Pairwise comparisons between offloading conditions		Memory Load					
		2		4		6	
		<i>t</i> (51)	<i>d_z</i>	<i>t</i> (51)	<i>d_z</i>	<i>t</i> (51)	<i>d_z</i>
Offloading Conditions	No reminder vs. Low-effort reminders	.39	.05	5.28***	.73	9.16***	1.27
	No reminder vs. High-effort reminders	1.10	.15	3.22**	.45	4.33***	.60
	Low-effort reminders vs. High-effort reminders	.81	.11	1.93#	.27	3.82***	.53

Note: #*p* < .1. **p* < .05. ***p* < .01. ****p* < .001.

Table 3

Experiment 1: Pairwise comparisons (t-values, effect sizes, and significance levels) of mean externalizing proportion between memory loads at each effort condition. Refer to Figure 3 for mean differences.

Pairwise comparisons between memory loads		Effort Condition			
		Low-effort reminders		High-effort reminders	
		<i>t</i> (51)	<i>d_z</i>	<i>t</i> (51)	<i>d_z</i>
Memory Loads	2 vs. 4	5.57***	.77	1.72 [#]	.24
	4 vs. 6	3.30**	.46	4.29***	.60
	2 vs. 6	7.44***	1.03	5.33***	.74

Note: [#]*p* < .1. **p* < .05. ***p* < .01. ****p* < .001.

Table 4

Experiment 1: Comparison of externalizing proportion between effort conditions at each memory load (mean differences, *t*-values, effect sizes, and significance levels).

Paired samples <i>t</i> -tests comparing low-effort vs. high-effort reminders at each load				
		<i>Mean difference</i>	<i>t</i> (51)	<i>d_z</i>
Memory Loads	2	.29	5.09***	.71
	4	.50	9.33***	1.29
	6	.46	8.88***	1.23

Note: #*p* < .1. **p* < .05. ***p* < .01. ****p* < .001.

Table 5

Experiment 1: Comparison between metacognitive judgements and objective unaided accuracy at each memory load (t-values, effect sizes, and significance levels). Refer to Figure 4 for mean differences.

Paired samples <i>t</i>-tests comparing metacognitive judgements vs. objective unaided accuracy at each load			
		<i>t</i>(51)	<i>d_z</i>
Memory Loads	2	3.78***	.52
	4	6.62***	.92
	6	8.22***	1.14

*Note: #*p* < .1. **p* < .05. ***p* < .01. ****p* < .001.*

Table 6

Experiment 2: One-way ANOVAs (*F*-values, significance levels, and partial eta-square) and pairwise comparisons (*t*-values, effect sizes, and significance levels) of mean target accuracy between memory loads at each offloading condition. Refer to Figure 5 for mean differences.

		Offloading Condition					
		No reminder		Low-effort reminders		High-effort reminders	
One-way within-subjects ANOVA		$F(1.88, 96.10) = 33.53^{***}$, $\eta_p^2 = .40$		$F(1.97, 100.56) = 2.33$, $\eta_p^2 = .04$		$F(1.82, 93.01) = 2.73^{\#}$, $\eta_p^2 = .05$	
Pairwise comparisons between memory loads		<i>t</i>(51)	<i>d_z</i>	<i>t</i>(51)	<i>d_z</i>	<i>t</i>(51)	<i>d_z</i>
Memory Loads	2 vs. 4	1.83 [#]	.25	1.42	.20	1.10	.15
	4 vs. 6	6.82 ^{***}	.95	.61	.08	1.45	.20
	2 vs. 6	7.21 ^{***}	1.00	2.22 [*]	.31	2.06 [*]	.29

Note: [#] $p < .1$. ^{*} $p < .05$. ^{**} $p < .01$. ^{***} $p < .001$.

Table 7

Experiment 2: Pairwise comparisons (t-values, effect sizes, and significance levels) of mean target accuracy between offloading conditions at each memory load. Refer to Figure 5 for mean differences.

Pairwise comparisons between offloading conditions		Memory Load					
		2		4		6	
		<i>t</i> (51)	<i>d_z</i>	<i>t</i> (51)	<i>d_z</i>	<i>t</i> (51)	<i>d_z</i>
Offloading Conditions	No reminder vs. Low-effort reminders	2.08*	.29	4.03***	.56	9.28***	1.29
	No reminder vs. High-effort reminders	.29	.04	1.42	.20	6.13***	.85
	Low-effort reminders vs. High-effort reminders	2.47*	.34	2.98**	.41	3.57***	.49

Note: #*p* < .1. **p* < .05. ***p* < .01. ****p* < .001.

Table 8

Experiment 2: Comparison between metacognitive judgements and objective unaided accuracy at each memory load (*t*-values, effect sizes, and significance levels). Refer to Figure 6 for mean differences.

Paired samples <i>t</i> -tests comparing metacognitive judgements vs. objective unaided accuracy at each load			
		<i>t</i> (51)	<i>d_z</i>
Memory Loads	2	4.09***	.57
	4	6.08***	.84
	6	6.71***	.93

Note: #*p* < .1. **p* < .05. ***p* < .01. ****p* < .001.

Table 9

Experiment 2: Pairwise comparisons (t-values, effect sizes, and significance levels) of mean externalizing proportion between memory loads at each effort condition. Refer to Figure 7 for mean differences.

Pairwise comparisons between memory loads		Effort Condition			
		Low-effort reminders		High-effort reminders	
		<i>t</i> (51)	<i>d_z</i>	<i>t</i> (51)	<i>d_z</i>
Memory Loads	2 vs. 4	5.73***	.79	3.96***	.55
	4 vs. 6	3.12**	.43	4.32***	.60
	2 vs. 6	6.99***	.97	7.17***	.99

Note: #*p* < .1. **p* < .05. ***p* < .01. ****p* < .001.

Table 10

Experiment 2: Comparison of externalizing proportion between effort conditions at each memory load (mean differences, *t*-values, effect sizes, and significance levels).

Paired samples <i>t</i> -tests comparing low-effort vs. high-effort reminders at each load				
		<i>Mean difference</i>	<i>t</i> (51)	<i>d_z</i>
Memory Loads	2	.21	4.00***	.55
	4	.34	5.48***	.76
	6	.23	4.48***	.62

Note: #*p* < .1. **p* < .05. ***p* < .01. ****p* < .001.

Peer Review Version

Table 11
Comparison of mean externalizing proportion between Experiments 1 and 2 at each effort condition.

Effort Condition	Memory Load	Mean Externalizing Proportion	
		Experiment 1	Experiment 2
Low-effort reminders	2	.41	.48
	4	.69	.78
	6	.84	.91
High-effort reminders	2	.12	.27
	4	.19	.45
	6	.38	.68
Difference between low- and high-effort conditions (Low - High)	2	.29	.21
	4	.50	.34
	6	.46	.23