AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Experimental Psychology: General

ISSN: 0096-3445

https://doi.org/10.1037/xge0001255

Thinking About Thinking: People Underestimate How Enjoyable and Engaging Just Waiting Is

Aya Hatano[1, 2], Cansu Ogulmus[3], Hiroaki Shigemasu[1], and Kou Murayama[3, 4, 5]

[1] School of Information, Kochi University of Technology
[2] IdeaLab Inc., Tokyo, Japan
[3] School of Psychology and Clinical Language Sciences, University of Reading
[4] Research Institute, Kochi University of Technology
[5] Hector Research Institute of Education Sciences and Psychology, University of Tübingen

The ability to engage in internal thoughts without external stimulation is a unique characteristic in humans. The current research tested the hypothesis that people metacognitively underestimate their capability to enjoy this process of “just thinking.” Participants (university students; total $N = 259$) were asked to sit and wait in a quiet room without doing anything. Across six experiments, we consistently found that participants’ predicted enjoyment and engagement for the waiting task were significantly less than what they actually experienced. This underappreciation of just thinking also led participants to proactively avoid the waiting task in favor of an alternative task (i.e., Internet news checking), despite their experiences not being statistically different. These results suggest an inherent difficulty in accurately appreciating how engaging just thinking can be, and could explain why people prefer keeping themselves busy, rather than taking a moment for reflection and imagination in our daily life.

Keywords: metacognition, affective forecasting, metamotivation, reward-learning, intrinsic motivation

Supplemental materials: https://doi.org/10.1037/xge0001255.supp

Humans are endowed with a striking capability to immerse themselves in their own thinking. Even without any external stimulation, people can engage in internal thought and reflection, making themselves interested and stimulated. Such spontaneous thinking manifests in a variety of forms, and has been extensively examined in psychological literature. We now know much about the characteristics, triggers, and consequences (i.e., functions) of these types of spontaneous thinking, including daydreaming (Singer, 1975), mind wandering (Smallwood & Schooler, 2015), episodic future thinking (Schacter et al., 2017), recollection of autobiographical memories (Conway & Pleydell-Pearce, 2000), and nostalgia (Sedikides et al., 2008).

The current article extends the literature by investigating an important but overlooked aspect of free thinking activity: How do we think about “just thinking”? More specifically, do we have good metacognition about how enjoyable and engaging the act of just thinking would be? A host of studies on metamemory (Bjork et al., 2013; Yan et al., 2014), affective forecasting (Wilson & Gilbert, 2005), and metamotivation (Scholer et al., 2018) have showed that people are often quite inaccurate at predicting their learning, affective, and motivational states, but past studies have not examined metacognitive (in) accuracy about people’s thinking activity. Understanding one’s metacognitive accuracy is particularly important in this context, because we constantly face situations in our daily life in which we need to make a decision between doing and not doing (i.e., just thinking), especially in the current era of information overload. Imagine a situation when you commute to your workplace by bus. While sitting in a bus, you can immerse yourself in your internal free-floating thinking or check a mobile phone to externally stimulate yourself. You may always check a mobile phone on a bus because you feel that just thinking would be boring and not stimulating. However, if this prediction is inaccurate, you are missing an opportunity to positively engage yourself without relying on such stimulation.

Theoretically, there are reasons to believe that people can actively and positively engage in thinking activities. According to the reward-learning framework of knowledge acquisition (Murayama,

Aya Hatano https://orcid.org/0000-0002-8881-3908

This research was supported by JSPS KAKENHI (Grants 16H06406, 18H01102, and 18K18696), Jacobs Foundation Advanced Research Fellowship, the Leverhulme Trust Research Leadership Award (Grant RL-2016-030), and the Alexander von Humboldt Foundation (the Alexander von Humboldt Professorship endowed by the German Federal Ministry of Education and Research). We thank Yuma Akashi, Naoki Kawashima, Ryo Kaizuka, and Cristina Pascua Martin for helping data collection. Pilot data were collected by Emily Marie Nix and Lidiia Perevoznikova as part of Bachelor's thesis. All the data, code, and markdown output are available in Open Science Framework (https://osf.io/a8bwm/).

Correspondence concerning this article should be addressed to Aya Hatano, School of Information, Kochi University of Technology, 185 Miyanokuchi, Tosayamada, Kami City, Kochi, Japan, or Kou Murayama, Hector Research Institute of Education Sciences and Psychology, University of Tübingen, Europastraße 6, 72072 Tübingen, Germany. Email: hatano.aya0@gmail.com or k.murayama@uni-tuebingen.de

2022; Murayama et al., 2019); for example, people can sustain intrinsically motivated behavior (i.e., behavior not motivated by extrinsic incentives) because they can generate internal reward from the change in mental representation (Baranes et al., 2014; Chater & Loewenstein, 2016; Gruber et al., 2014). Because unconstrained thinking involves constant changes and updates of one's mental representation, just thinking can entail some positive and rewarding experiences. However, it is unlikely that we are explicitly aware of the rewarding nature of free-floating thinking. Murayama (2022) indicated that the internal rewarding process through representational change generally lacks explicit external cues and involves complex interactions with preexisting, self-relevant knowledge. As a result, people underestimate the potential rewarding value of intrinsically rewarding activities. Ruan et al. (2018), for example, found that people tend to underestimate their pleasant experiences when they are forced to think about missing information (e.g., answers to a trivia question), and prefer to see the missing information right away (see also Sansone et al., 1992; Wilson et al., 2005; Woolley & Fishbach, 2015).

Based on these observations (and pilot data we collected before the study), we hypothesized that people's metacognition *underappreciates* the potential enjoyment and engagement associated with the activity of free thinking. In fact, despite these potential engaging aspects of thinking, empirical literature found that people often *proactively* avoid just waiting to think. Hsee et al. (2010), for example, showed that people have a tendency to avoid just waiting, especially when there are justifications for being busy ("idleness aversion"). People prefer an immediate smaller reward rather than waiting to receive a larger one (Frederick et al., 2002; Kable & Glimcher, 2007; McClure et al., 2007). Other studies showed that people even incur a cost to proactively avoid just waiting (Bishai & Lang, 2000; Frederick et al., 2002). Although these studies did not assess people's metacognitive evaluation, the results are consistent with a metacognitive belief that thinking itself is not intrinsically motivating.

Note that we are not arguing that just thinking is a very enjoyable task. Thinking exercises involve some mental cost and people need both motivation and skill to fully appreciate its rewarding nature (Alahmadi et al., 2017; Westgate & Wilson, 2018; Westgate et al., 2017). In fact, empirical studies have indicated that just thinking is experienced as, on average, neither too boring nor too intrinsically motivating. For example, when participants were asked to engage in just thinking without performing any overt activities, they reported that this thinking period was somewhat enjoyable and somewhat boring, with the average enjoyment ratings on approximately the midpoint of the Likert scale (Wilson, Reinhard, et al., 2014). This is substantially less than what people experience when they engage in enjoyable external activities (Wilson, Gilbert, et al., 2014). Our hypothesis, however, is that people exhibit *relative* underestimation of this moderate level of positive and motivating experiences during the thinking period.

In the current experiments, participants were asked to wait and just think for a certain period (e.g., 20 min). Critically, before the task, participants made a prediction about how enjoyable and engaging their experiences would be at the end of the waiting task. This prediction was directly compared with their experienced enjoyment and engagement they reported at the end of the waiting task. We expected that participants would experience greater enjoyment and engagement than what they predicted. To our knowledge, two of the previous articles (Alahmadi et al., 2017, Studies 5 and S1; Wilson, Reinhard, et al., 2014, Study 11) assessed the prediction of enjoyment for the waiting task. These studies collected the data for different purposes, and they did not directly compare the predicted and experienced enjoyment; our reanalyses of the archived data (https://osf.io/cgwdy/) from Wilson, Reinhard, et al. (2014) and descriptive data reported in Alahmadi et al. (2017) does not seem to show reliable differences between the conditions. This may be because participants in these studies made a prediction online after reading a passage of a hypothetical experiment, without being put in a situation of doing the actual waiting task (we will discuss another possibility in the General Discussion). As you will see, when participants made a prediction as part of a real on-site experiment, they exhibited robust and strikingly large underestimations.

Experiment 1A

Method

All studies were approved by the ethics committee at Kochi University of Technology (No. 148) and University of Reading (No. 2016-125-KM). We report all relevant experiments that were conducted. In this and the following experiments, all data, code, and markdown outputs are available via the Open Science Framework (OSF; https://osf.io/a8bwm/). The graphs in the Results sections in this article include individual raw data points and were generated using Open-visualizations in *R* and Python (van Langen, 2020).

Participants

Thirty university students in Japan participated in the experiment (female = 9, $M_{\text{age}} = 19.63$, $SD = 1.13$). In this and the following experiments, convenience sampling was used (although data was collected in two different countries to ensure generalizability). We had pilot experiment data ($N = 18$) showing a significant underestimation effect with the same measure. Based on the estimated effect size from the pilot data (Cohen's $dz = .65$), we set the minimal sample size that achieved at least 80% statistical power and collected data as much as possible within a term period. The final sample size gives us a statistical power of 93%. Data from one participant were excluded because of missing values.

Measures

In this and the following experiments, we assessed participants' enjoyment and engagement for the task with a 12-item scale consisting of four subcategories (three items for each subcategory): task enjoyment (e.g., "I enjoy waiting very much."; Elliot & Harackiewicz, 1996), task engagement (e.g., "I lose track of time."; Elliot & Harackiewicz, 1996), task interest (e.g., "I find waiting interesting."; adopted from (Wigfield & Eccles, 2000), and task boredom (e.g., "Waiting bores me."; adopted from Pekrun et al., 2002). All items were assessed on a 7-point scale (1–7). We used a translated Japanese version of the scale. Both the original English scale (that was used in Experiment 4) and the translated Japanese scale are available on the OSF (https://osf.io/a8bwm/).

As all subscales were highly correlated with each other, we averaged the scores from all items (after reverse coding the boredom items) to create a single index of "task motivation" (Cronbach's α = .92 for the predicted intrinsic motivation scale and .95 for the experienced intrinsic motivation scale). Exploratory factor analysis of the scale with a single factor solution showed that all the items loaded on the factor well (see online supplemental materials). Here we operationalized task motivation as an amalgam of (a) positive emotions (enjoyment, interest, and lack of boredom) and (b) task engagement. Although the basic pattern was similar across all of the subscales for all experiments, due to the theoretical distinction between positive emotional feelings and task engagement, we also report the results with subcategories.

Procedure

When participants arrived in the experiment, participant received the following instructions in a room next to the seminar room where the main experiment was conducted:

> The purpose of this experiment is to examine people's cognitive processing during waiting. You will be required to leave your belongings in this room and stay in the next room (i.e., experimental room) for 20 minutes. You have nothing else to do other than sit in a chair. While waiting, you may think about anything, but you may not do any of the following things: (a) sleep, (b) walk/exercise, (c) look at/use a smartphone, (d) eat, or (e) look at a watch.

The original instructions were delivered in Japanese and are available on the OSF (https://osf.io/a8bwm/). After the instructions, participants were asked to make a prediction about their task motivation at the end of the 20-min waiting task ("Please make a prediction about how you will feel about the task in 20 minutes"). Then participants were asked to remove their belongings (including mobile phone and wristwatch), were led to and seated in the seminar room (i.e., main experimental room; Figure 1A), and left alone. All the window blinds were closed. After 20 min, participants were collected from the seminar room and asked to rate their current (experienced) task motivation for the waiting task.

Results

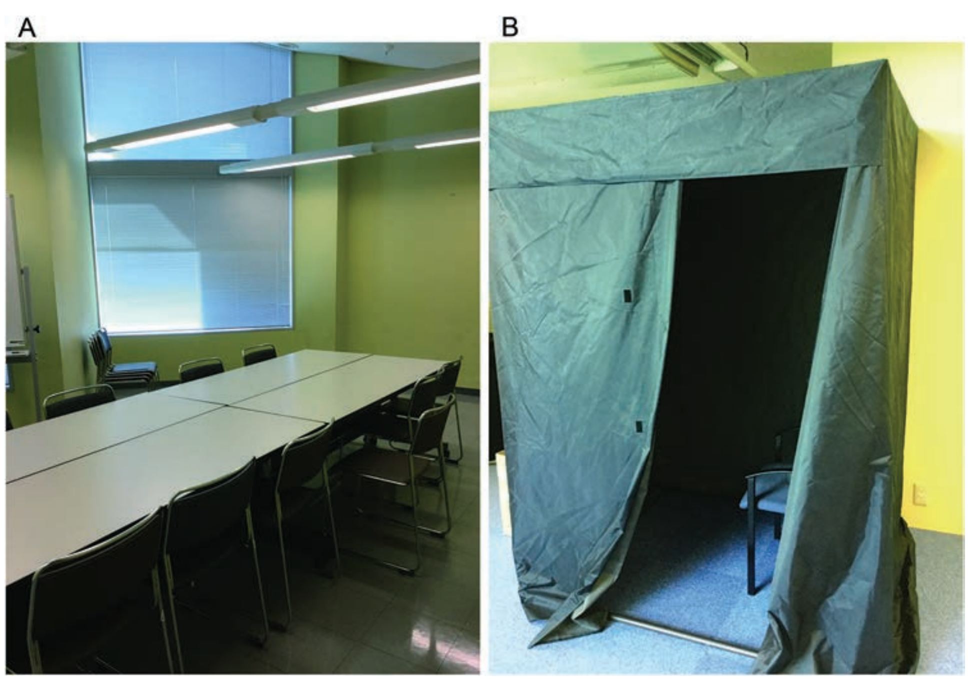
The comparison between predicted and experienced task motivation supported our prediction: The results revealed that participants experienced task motivation (i.e., enjoyment and task engagement) for the waiting task ($M = 3.34$, $SD = 1.54$) more than they had predicted ($M = 2.67$, $SD = 1.16$), $t(28) = 3.92$, $p < .001$, $dz = .73$. Figure 2 shows results with individual data points.

Table 1 reports the comparison of the conditions when analyzing the subcategories separately (i.e., enjoyment, engagement, interest, and boredom). The results were consistent across the subcategories: Participants experienced more enjoyment, more engagement, more interest, and less boredom than they had predicted, $t(29) = 2.76$, 3.34, $dz = .50$, .61, for enjoyment and engagement, respectively; $t(28) = 3.39$, 3.78, $dz = .63$, .70, for interest and boredom, respectively.

Experiment 1B

In Experiment 1A, the same participants provided the ratings for both predicted and experienced task motivation. With this within-participant design, it is possible that their ratings for experienced motivation were somewhat influenced by their prediction itself. In Experiment 1B, we had participants perform the same

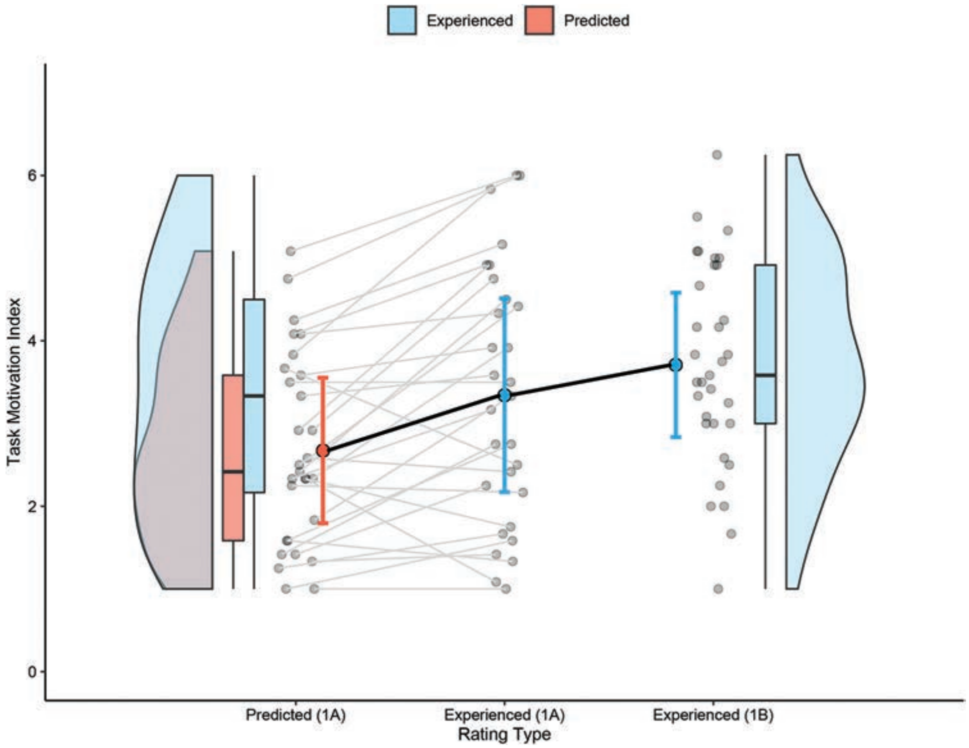
Figure 1
Waiting Room Used for Experiment 1 (A) and Experiment 2 (B)



Note. See the online article for the color version of this figure.

Figure 2

Comparison Between Predicted and Experienced Task Motivation in Experiments 1A and 1B



Note. Distributions are shown in raincloud plots and data points have been jittered to enhance visual clarity. Bold dots indicate means and bold bars indicate standard errors. See the online article for the color version of this figure.

waiting task and rate their experienced task motivation without asking them to make a prediction. Experienced task motivation scores in this no-prediction control were then compared with the predicted motivation scores obtained in Experiment 1A.

Method

Participants

Thirty-three Japanese university students participated in the experiment (female = 9, M_{age} = 19.91, SD = 1.57). This is the same cohort of students from the same university and department in Experiment 1A. We had further pilot data from a no-prediction control (N = 20) showing a significant underestimation effect in comparison with the pilot data used in Experiment 1A. Based on the estimated effect size from the pilot results (Cohen's d = .72), we set the minimal sample size that achieved at least 80% statistical power and collected data as much as possible within a term period. The final sample size gives us the statistical power 82%.

Measures

We used the same 12-item scale as Experiment 1A to assess task motivation (i.e., enjoyment and engagement; Cronbach's α = .92).

Procedure

The procedure was identical with Experiment 1A, except that participants did not make any predictions about their task motivation before commencing the waiting task.

Results

The results confirmed the robustness of the previous findings: Even when an independent sample evaluated their experienced

Table 1

Comparison Between Predicted and Experienced Subcategories of Task Motivation in Experiment 1A

Subcategories	Mean prediction (*SD*)	Mean experience (*SD*)	*t*	*p*	Effect size (Cohen's *dz*)
Enjoyment	2.77 (1.30)	3.42 (1.78)	2.76	.010	0.50
Engagement	2.57 (0.99)	3.21 (1.41)	3.34	.002	0.61
Interest	2.38 (1.19)	2.93 (1.75)	3.39	.002	0.63
Boredom	4.99 (1.76)	4.21 (1.84)	3.78	<.001	0.70

Note. Standard deviations of the group means are reported in parentheses.

motivation without prediction, participants' experienced task motivation (M = 3.71, SD = 1.23) was higher than the predicted task motivation reported in Experiment 1A (M = 2.67, SD = 1.16), $t(59.74) = 3.42$, $p = .001$, $d = .87$ (Welch's test). Figure 2 shows results with individual data points. The experienced task motivation was not significantly different between Experiment 1A and Experiment 1B, $t(53.54) = 1.03$, $p = .309$, $d = .27$ (Welch's test). Given that within- and between-subjects designs did not show significantly different results (with between-subjects design showing numerically larger effects), in the following experiments, we used within-subject design to ensure sufficient statistical power.

Table 2 reports the comparison of the conditions when analyzing the subcategories separately (i.e., enjoyment, engagement, interest, and boredom). The results were consistent across the subcategories: Participants experienced more enjoyment, more engagement, more interest, and less boredom than the predicted task motivation rated by participants in Experiment 1A, $t(60.94) = 2.60$, $t(57.20) = 2.96$, $t(59.99) = 2.06$, and $t(57.78) = 3.76$, respectively, d = .65, .73, .52, and .96.

Experiment 2

Experiment 2 aimed to replicate the previous results by having participants wait in a dark room that blocked any visual and auditory input. This set-up completely prohibits participants from relying on external cues to make the waiting task engaging.

Method

Participants

Thirty Japanese university students participated in the experiment (female = 9, M_{age} = 19.67 years, SD = 1.45). Based on the estimated effect size from Experiment 1A (Cohen's dz = .73), we set the minimal sample size to achieve at least 80% statistical power and collected data as much as possible within a term period. The final sample size gives us a statistical power of 97%.

Measures

We used the same 12-item task motivation scale as Experiment 1A (Cronbach's α = .89 for the predicted task motivation scale and .94 for the experienced task motivation scale).

Procedure

The procedure was identical with Experiment 1A, except for the setting of the waiting task. Specifically, participants spent 20 min sitting on a chair in a small, dark booth built for vision experiments, with noise-proof headphones (Figure 1B). Participants were shown the headphones and the inside of the booth before making a prediction about their intrinsic motivation for the waiting task.

Results

Replicating the previous findings and consistent with our prediction, the comparison between the predicted and experienced task motivation showed a significant difference, $t(29) = 3.74$, $p < .001$, $dz = .68$, indicating that participants enjoyed and were engaged in the waiting task (M = 3.73, SD = 1.41) more than they had predicted (M = 2.80, SD = 1.02). Figure 3 shows results with individual data points.

Table 3 reports the comparison between conditions when analyzing the subcategories separately. The results were again consistent across the subscales: Participants experienced more enjoyment, more engagement, more interest, and less boredom than they predicted, $t(29)$ = 3.33, 2.87, 2.99, and 3.40, respectively, dz = .61, .52, .55, and .62.

Experiment 3

Experiment 3 aimed to examine the potential effects of time interval on the underestimation of task motivation by manipulating the time for the waiting task. We also examined the validity of the self-reported ratings of predicted task motivation by assessing actual behavioral choice.

Method

Participants

Sixty-three Japanese undergraduate students participated in this experiment (female = 19, M_{age} = 20.02 years, SD = 1.44). They were assigned randomly to the short- or long-wait condition. Because we did not have a strong hypothesis about the effect of duration, we simply aimed to collect a similar number of participants with the previous study (i.e., approximately 30 participants for each condition).

Measures

We used the same 12-item task motivation scale as Experiment 1A (Chronbach's αs = .91 ~ .94). For exploratory purpose, participants also rated two postexperiment questions: their feelings of wasting time ("I feel I wasted my time") and feelings of

Table 2

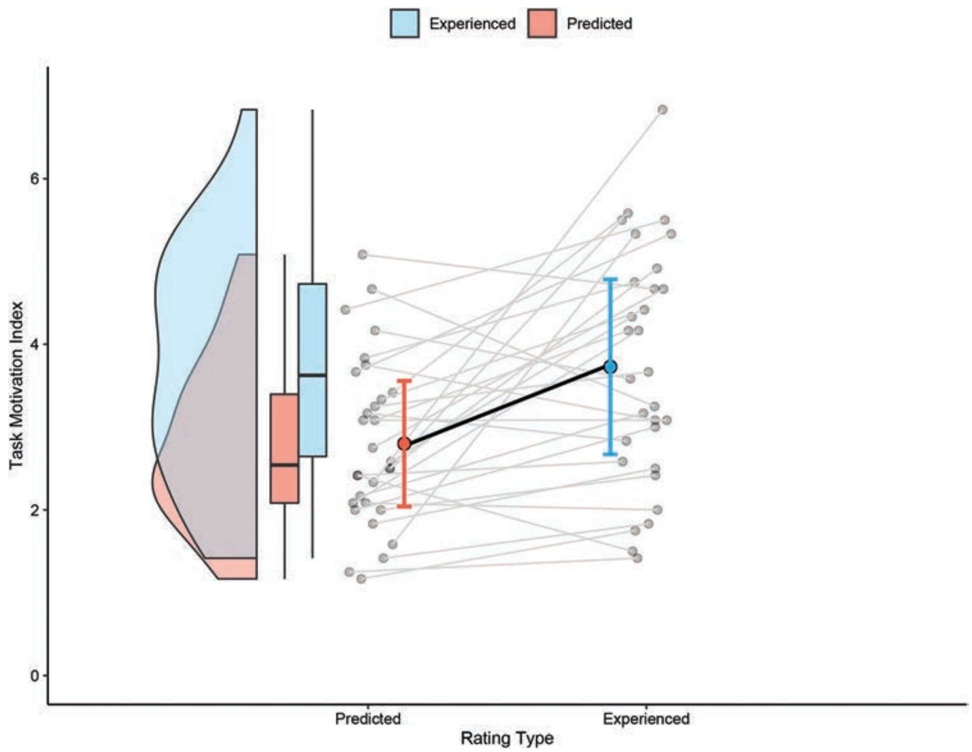
Comparison Between Predicted and Experienced Subcategories of Task Motivation in Experiment 1B

Subcategories	Mean prediction in Experiment 1A[a] (*SD*)	Mean experience (*SD*)	*t*	*p*	Effect size (Cohen's *d*)
Enjoyment	2.77 (1.30)	3.68 (1.48)	2.60	.012	0.65
Engagement	2.57 (0.99)	3.47 (1.43)	2.96	.005	0.73
Interest	2.38 (1.19)	3.04 (1.34)	2.06	.043	0.52
Boredom	4.99 (1.76)	3.35 (1.65)	3.76	<.001	0.96

Note. Standard deviations of the group means are reported in parentheses.
[a] Participants did not predict task motivation in Experiment 1B but the comparison was made with the predicted task motivation in Experiment 1A.

Figure 3

Comparison Between Predicted and Experienced Task Motivation in Experiment 2



Note. Distributions are shown in raincloud plots and data points have been jittered to enhance visual clarity. Bold dots indicate means and bold bars indicate standard errors. See the online article for the color version of this figure.

refreshment ("I feel refreshed") on a 7-point scale. Descriptive results on these exploratory items are reported in online supplemental materials.

Procedure

At the beginning of the experiment, participants were told that their task was to wait for 3 min (short-wait condition) or 20 min (long-wait condition) in a room by themselves. Participants were then asked to make a prediction about their task motivation (i.e., enjoyment and engagement) at the end of waiting task for both the short-wait and long-wait conditions.

After the prediction ratings, participants were told that they could choose one of the two lotteries that determine which condition the participants would be assigned to. The first lottery has 70% chance of being assigned to the short-wait condition and 30% chance of being assigned to the long-wait condition (short-wait dominant lottery); on the other hand, the second lottery has 70% chance of being assigned to the long-wait condition and 30% chance of being assigned to the short-wait condition (long-wait dominant lottery). After the participants' choice, they were informed about the condition that they were assigned to. Unbeknownst to participants, the outcome of the lottery was predetermined—half of participants were assigned to the short-wait condition whereas the other half were assigned to the long-wait condition, regardless of their choice. This procedure ensured the random assignment of participants while allowing us to assess their behavioral preference. To prevent participants from choosing the short-wait task just to finish the experiment earlier, we instructed that they would work on a completely separate task if they were assigned to the short-wait condition, to fill the time gap of the two conditions.

Table 3

Comparison Between Predicted and Experienced Subcategories of Task Motivation in Experiment 2

Subcategories	Mean prediction (*SD*)	Mean experience (*SD*)	*t*	*p*	Effect size (Cohen's *dz*)
Enjoyment	2.73 (1.31)	3.79 (1.60)	3.33	.002	0.61
Engagement	2.62 (1.10)	3.43 (1.42)	2.87	.008	0.52
Interest	2.66 (1.12)	3.48 (1.57)	2.99	.006	0.55
Boredom	4.81 (1.52)	3.79 (1.84)	3.40	.002	0.62

Note. Standard deviations of the group means are reported in parentheses.

Participants moved to a room after removing all of their belongings (including mobile phone and wristwatch) and spent 3 or 20 min alone in the room, depending on the condition they had been assigned to. After the waiting period, they rated their experienced task motivation.

Results

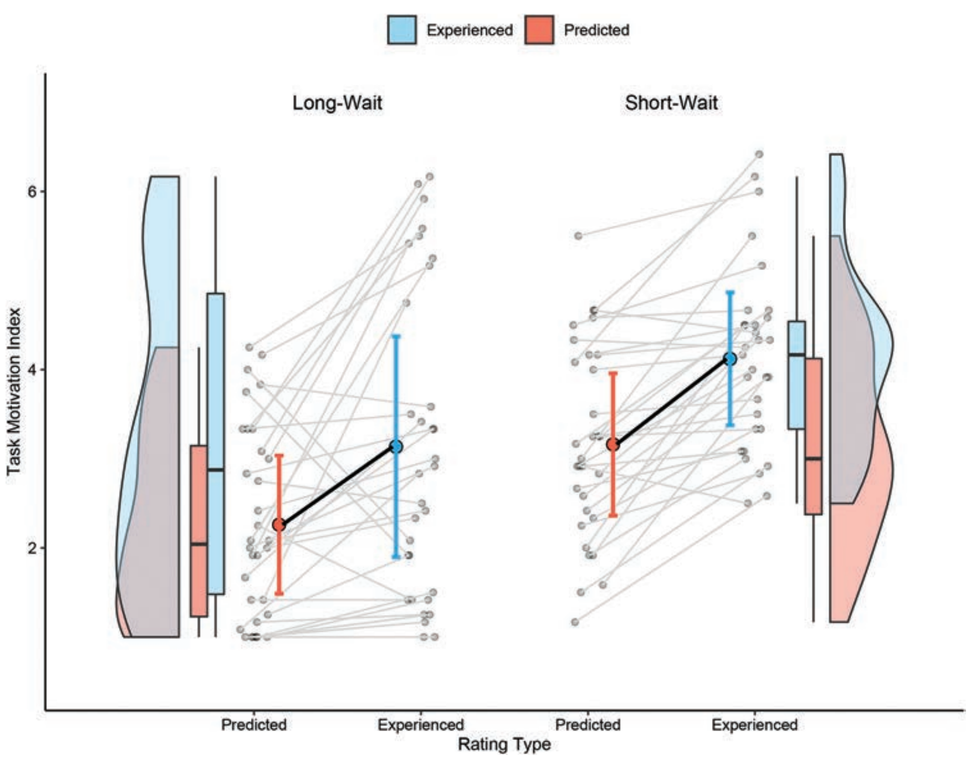
To examine whether the discrepancy between predicted and actual task motivation is different depending on the duration of the wait task, we conducted a 2 (Waiting condition: short-wait vs. long-wait) × 2 (Rating type: predicted motivation vs. experienced motivation) mixed analysis of variance (ANOVA) on task motivation. Note that, although all participants rated predicted motivation for both short- and long-wait tasks, this analysis focused on their predictive ratings for the task to which they were actually assigned (e.g., if a participant was assigned to a short-wait condition, predicted task motivation for the short-wait task was used in the primary analysis). We found a main effect of rating type, $F(1, 61) = 38.43$, $p < .001$, $\eta_G^2 = .12$, indicating that experienced task motivation ($M = 3.62$, $SD = 1.49$) was significantly higher than the predicted task motivation ($M = 2.70$, $SD = 1.17$). A main effect of the waiting condition was also statistically significant, $F(1, 61) = 11.26$, $p = .001$, $\eta_G^2 = .13$, showing that short-wait condition exhibited higher task motivation ($M = 3.64$, $SD = 1.15$) for both predicted and experienced motivations (long-wait condition: $M = 2.70$, $SD = 1.49$). However, the interaction effect was not significant, $F(1, 61) = .08$, $p = .776$, $\eta_G^2 = .00$.

Figure 4 shows results with individual data points. These results have two important implications: First, the underestimation of task motivation, which was consistently observed in the previous experiments, may not be influenced by the duration of the task. Second, the short-wait condition was generally associated with higher ratings for both predicted and experienced task motivation, indicating that participants were sensitive to the duration of the task when rating their predicted and experienced motivation. We also conducted additional analysis to confirm that the results cannot be explained by cognitive dissonance (i.e., incongruence between participants' choice and actual assignment of the condition; see online supplemental materials).

Did participants' choices reflect their predicted task motivation? Overall, predicted task motivation for the short-wait task ($M = 3.20$, $SD = 1.24$) was larger than the predicted task motivation for the long-wait task ($M = 2.49$, $SD = 1.09$), $t(62) = 5.57$, $p < .001$, $dz = .70$. Analysis on behavioral choice indicated that, consistent with the higher predicted task motivation for the short-wait task, 70% of participants preferred the short-wait dominant lottery over the long-wait dominant lottery, $p = .002$, binomial test. To further examine whether behavioral choice is consistent with participants' self-reported rating, we first computed the difference scores of predicted task motivation for the short-wait and long-wait tasks and

Figure 4

Comparison Between Predicted and Experienced Task Motivation as a Function of Short- and Long-Wait Conditions in Experiment 3



Note. Distributions are shown in raincloud plots and data points have been jittered to enhance visual clarity. Bold dots indicate means and bold bars indicate standard errors. See the online article for the color version of this figure.

correlated the scores with the behavioral choice. Here we coded the data in a way that positive correlation indicates the consistency between the self-reported preference and choice. The estimated point biserial correlation was positive and statistically significant, $r = .39$, $p = .001$, suggesting the validity of the self-reported predictive motivation as predictors of behavioral choice.

We also conducted the same 2 (Waiting condition: short-wait vs. long-wait) × 2 (Rating type: predicted motivation vs. experienced motivation) mixed ANOVA on subcategories separately. The primary underestimation effect was observed across all subcategories: There was a significant main effect of rating type for task enjoyment, $F(1, 61) = 16.27$, $p < .001$, $\eta^2_G = .06$, task engagement, $F(1, 61) = 44.77$, $p < .001$, $\eta^2_G = .14$, task interest, $F(1, 61) = 17.56$, $p < .001$, $\eta^2_G = .06$, and task boredom, $F(1, 61) = 21.33$, $p < .001$, $\eta^2_G = .09$. The interaction effect of rating type and waiting condition was not significant for any of the subcategories, indicating that the underestimation effect is not statistically different between the short- and long-wait conditions: enjoyment, $F(1, 61) = .24$, $p = .63$, $\eta^2_G = .00$, engagement, $F(1, 61) = .25$, $p = .62$, $\eta^2_G = .00$, interest, $F(1, 61) = .79$, $p = .38$, $\eta^2_G = .00$, and boredom, $F(1, 61) = .001$, $p = .98$, $\eta^2_G = .00$. The main effect of waiting condition was statistically significant for task enjoyment, $F(1, 61) = 14.08$, $p < .001$, $\eta^2_G = .15$, and task boredom, $F(1, 61) = 18.14$, $p < .001$, $\eta^2_G = .18$. The main effect of waiting condition was not statistically significant for task engagement, $F(1, 61) = 2.59$, $p = .11$, $\eta^2_G = .03$, and task interest, $F(1, 61) = 1.85$, $p = .18$, $\eta^2_G = .02$.

To provide a more intuitive insight into effect sizes, Table 4 reports the comparison of the conditions when analyzing the subcategories separately (i.e., enjoyment, engagement, interest, and boredom) for short- and long-wait conditions, respectively. The results were consistent across the subcategories: Participants in the Long-waiting condition experienced more enjoyment, more task engagement, more interest, and less boredom than they predicted, $t(31) = 2.09$, 4.99, 2.16, and 2.85, respectively, $dz = .37$, .88, .38, and .50. Participants in the Short-waiting condition also experienced significantly more enjoyment, more task engagement, more interest, and less boredom than they predicted, $t(30) = 4.42$, 4.47, 3.98, and 3.98, respectively, $dz = .79$, .80, .71, and .72.

Experiment 4

One potential explanation of the prior results is that participants experienced general positive feelings merely because they had just finished a long and boring task. Experiment 4 examined the robustness of the findings by comparing the prediction and experience of enjoyment and engagement *during* the waiting task and *after* the waiting task. In addition, Experiment 4 used a sample from the United Kingdom, allowing us to examine if the underestimation effect is robust across different cultural contexts.

Method

Participants

Forty United Kingdom university students participated in the experiment (female = 31, $M_{age} = 25.50$ years, $SD = 11.90$). Based on the estimated effect size from Experiment 2 (Cohen's $dz = .68$), we set the minimal sample size that achieved at least 80% statistical power and collected data as much as possible within a term period. The final sample size provided a statistical power of 99%.

Measures

We used the same 12-item task motivation scale as Experiment 1A (Chronbach's αs = .88 ∼ .94). Note that we used an English version of the scale in this experiment, which is available on the OSF (https://osf.io/a8bwm/).

Procedure

The procedure was identical with Experiment 1A with the following two exceptions. First, experiment was run in English. The exact instructions were as follows [we have uploaded the manual of the experiment at OSF (https://osf.io/a8bwm/)].

> For the following task you will be required to sit and abstain from any activity for 20 minutes. It is important for the experiment that you should not engage in any activity. For example, walking around the room and rummaging through the drawers in the room are considered activities. Please have a seat and stay there until the experiment is over.

Second, participants were additionally asked to rate their predicted and experienced task motivation (enjoyment and engagement) during the waiting task. More specifically, before the waiting task, participants were asked to make a prediction about their task motivation not only upon completion of the 20-min task, but also after X minutes from the start of the task. Participants were

Table 4
Comparison Between Predicted and Experienced Subcategories of Task Motivation in Experiment 3

Subcategories	Mean prediction (*SD*)	Mean experience (*SD*)	*t*	*p*	Effect size (Cohen's *dz*)
Enjoyment					
Long-waiting	2.22 (1.37)	2.89 (1.82)	2.09	.045	0.37
Short-waiting	3.38 (1.56)	4.23 (1.25)	4.42	<.001	0.79
Engagement					
Long-waiting	2.32 (1.14)	3.52 (1.72)	4.99	<.001	0.88
Short-waiting	2.91 (1.29)	3.95 (1.46)	4.47	<.001	0.80
Interest					
Long-waiting	2.16 (1.22)	2.73 (1.76)	2.16	.039	0.38
Short-waiting	2.44 (1.28)	3.32 (1.47)	3.98	<.001	0.71
Boredom					
Long-waiting	5.66 (1.60)	4.59 (2.23)	2.85	.008	0.50
Short-waiting	4.09 (1.54)	3.01 (1.42)	3.98	<.001	0.72

Note. Standard deviations of the group means are reported in parentheses.

then asked to report their ongoing experienced task motivation after X minutes had passed from the start of the task. The interval X (integer, $2 \leq X \leq 18$) was randomly sampled with equal probability and assigned to individual participants. The timing of the interim rating was notified by an alarm clock. Regardless of the length of X, all participants performed the waiting task for 20 min and reported their experienced task motivation afterward.

Results

Analyses on Overall Task Motivation

We first sought to replicate the main findings that participants underestimated the task motivation for the 20-min wait task with the United Kingdom sample. Consistent with the previous findings, the comparison between the predicted and experienced motivation upon the completion of the 20-min task showed a significant difference, $t(39) = 9.09$, $p < .001$, $dz = 1.44$. Again, participants enjoyed and were engaged in the waiting task ($M = 4.66$, $SD = 1.11$) more than they had predicted ($M = 3.38$, $SD = .94$).

To examine whether the underestimation was observed during the task, we conducted a linear mixed-effects model predicting the interim rating of task motivation from interim rating time (a continuous variable centered on 10: range = −8 to 8), rating type (prediction vs actual experience; effect coded), and their interaction as fixed effects. The model included random participant intercepts (random participant slopes are inherently confounded with errors in this model). Results showed a significant main effect of rating type condition, $\beta = .81$, $SE = .14$, 95% confidence interval, CI [.54, 1.09], $t(38) = 5.78$, $p < .001$; indicating that the experienced task motivation ($M = 4.35$, $SD = 1.20$) was significantly larger than the predicted task motivation ($M = 3.52$, $SD = .92$) even if participants rated their motivation during the task. The main effect of interim rating time, $\beta = .02$, $SE = .04$, 95% CI [−.05, .09], $t(52.49) = .54$, $p = .595$, and its interaction effect with rating type, $\beta = -.03$, $SE = .03$, 95% CI [−.09, .03], $t(38) = -1.00$, $p = .322$, were not statistically significant. These results suggest that the difference between predicted and experienced task motivation did not significantly change with the interim rating time. In fact, with a simple linear regression analysis, neither the experienced task motivation, $\beta = -.01$, $SE = .04$, 95% CI [−.09, .07], $t(38) = -.25$, $p = .804$, nor the predicted task motivation, $\beta = .02$, $SE = .03$, 95% CI [−.04, .08], $t(38) = .63$, $p = .536$, was significantly predicted by the interim rating time.

The previous analyses showed that, regardless of whether participants rated their experienced motivation during or after the task, there was a significant underestimation of task motivation, which is consistent with the results from the previous experiments. But is the magnitude of the underestimation different between the ratings made during and after the waiting task? To test this possibility, a 2 (Rating type: predicted or experienced intrinsic motivation) × 2 (Rating timing: interim or after 20 mins period) repeated measures ANOVA was conducted. Because previous analyses showed no significant effect of interim rating time, we did not include this information to simplify the analysis. The ANOVA showed a significant main effect of Rating type, $F(1, 39) = 73.10$, $p < .001$, $\eta^2_G = .21$, which is consistent with the previous observations (i.e., participants experienced more intrinsic motivation than they had predicted, regardless of the timing of the rating). There was also a significant interaction effect of Rating Type × Rating Timing, $F(1, 39) = 11.86$, $p = .001$, $\eta^2_G = .01$. Participants' experienced task motivation was larger after 20 min ($M = 4.66$, $SD = 1.11$) compared with that they experienced during the waiting period ($M = 4.35$, $SD = 1.20$), $t(39) = 3.08$, $p = .004$, $dz = .49$. On the other hand, participants' predictions of task motivation after 20 min ($M = 3.38$, $SD = .94$) and during the waiting period ($M = 3.52$, $SD = .92$) did not significantly differ, $t(39) = 1.37$, $p = .178$, $dz = .22$. Although this is an exploratory analysis, these results may suggest that there is indeed an "added value" of completing the waiting task (that was not accurately accounted for in participants' predictions) but this effect still cannot fully explain the observed underestimation of task motivation for the wait task. The main effect of rating timing was not statistically significant, $F(1, 39) = 1.09$, $p = .302$, $\eta^2_G = .00$. Figure 5 shows the results with individual data points.

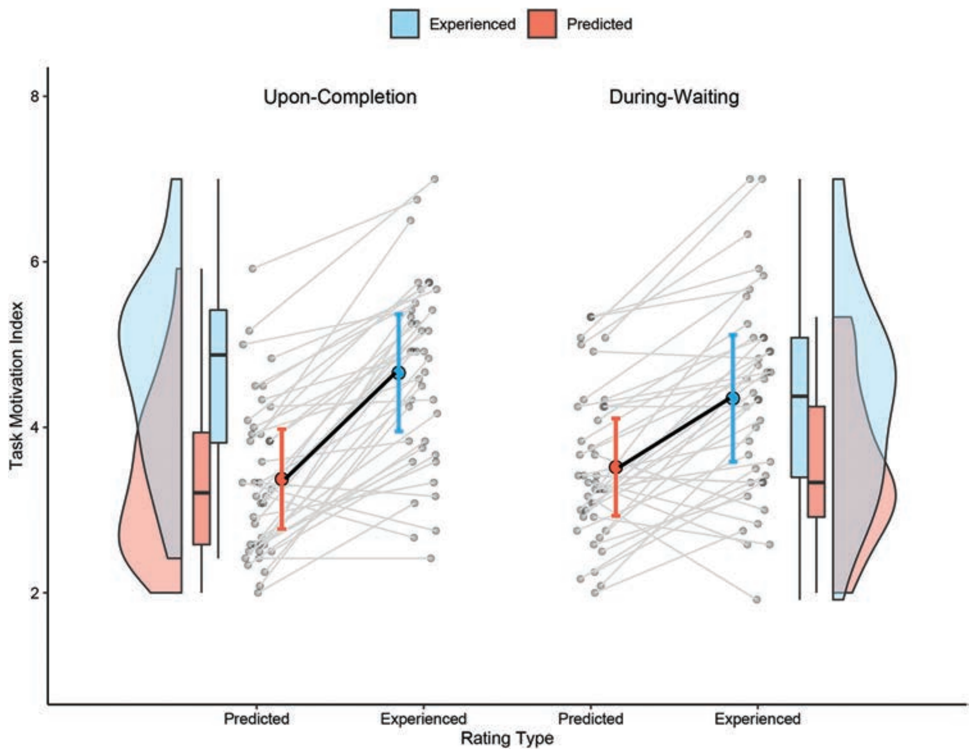
Analyses on Separate Subcategories

To examine whether results were consistent across subcategories (i.e., enjoyment, task engagement, interest, and boredom), we conducted the same linear mixed models predicting subcategory scores from interim rating time, rating type, and their interaction. As indicated in Table 5, the results were all consistent. A significant main effect of rating type was observed for enjoyment, $\beta = .56$, $SE = .18$, 95% CI [.22, .90], $t(38) = 3.19$, $p = .003$, task engagement, $\beta = .74$, $SE = .18$, 95% CI = [.39, 1.09], $t(38) = 4.12$, $p < .001$, interest, $\beta = .91$, $SE = .17$, 95% CI [.58, 1.23], $t(38) = 5.47$, $p < .001$, and boredom, $\beta = 1.04$, $SE = .22$, 95% CI [−1.47, −.61], $t(38) = 4.71$, $p < .001$. Regardless of the timing of the ratings, participants experienced greater enjoyment, engagement, and interest than they predicted, and they also experienced less boredom than they predicted. Critically, none of the interaction effects were significant, suggesting that the inaccurate prediction is not dependent on the time of the ratings (see Table 5).

Like the task motivation analysis, we also conducted a 2 (Rating type: predicted or experienced) × 2 (Rating timing: interim or after 20 mins period) repeated measures ANOVA for each subcategory separately. A significant main effect of rating type was observed for all subcategories: task enjoyment, $F(1, 39) = 34.01$, $p < .001$, $\eta^2_G = .13$, task engagement, $F(1, 39) = 38.51$, $p < .001$, $\eta^2_G = .17$, task interest, $F(1, 39) = 56.66$, $p < .001$, $\eta^2_G = .15$, and task boredom, $F(1, 39) = 40.47$, $p < .001$, $\eta^2_G = .18$. These results indicate that, regardless of the rating timing (whether the rating was done upon completion of the waiting task or during the waiting task), participants experienced greater enjoyment, greater engagement, greater interest, and less boredom than they predicted. The main effect of rating timing was not statistically significant for task enjoyment, $F(1, 39) = .001$, $p = .97$, $\eta^2_G = .00$, task interest, $F(1, 39) = 2.37$, $p = .13$, $\eta^2_G = .004$, and task boredom, $F(1, 39) = .05$, $p = .82$, $\eta^2_G = .00$. On the other hand, a significant main effect of rating timing was observed in task engagement, $F(1, 39) = 6.46$, $p < .05$, $\eta^2_G = .01$, showing that rating was generally higher (collapsing across experienced and predicted task engagement) after the waiting task than during the task.

Consistent with the overall task motivation analysis, the interaction effect of rating type and rating timing was statistically significant for task enjoyment, $F(1, 39) = 16.76$, $p < .001$, $\eta^2_G = .02$, task engagement, $F(1, 39) = 5.81$, $p < .05$, $\eta^2_G = .01$, and task boredom, $F(1, 39) = 6.17$, $p < .05$, $\eta^2_G = .01$. The only exception was task

Figure 5
Comparison Between Predicted and Experienced Task Motivation as a Function of the Timing of Assessment in Experiment 4



Note. Distributions are shown in raincloud plots and data points have been jittered to enhance visual clarity. Bold dots indicate means and bold bars indicate standard errors. See the online article for the color version of this figure.

interest, where the interaction effect was not statistically significant, $F(1, 39) = 1.44$, $p = .24$, $\eta^2_G = .002$. To clarify the nature of the interaction effect (and associated effect sizes), Table 6 reports the comparison of the predicted and experienced ratings when analyzing the subcategories separately (i.e., task enjoyment, task engagement, task interest, and task boredom). While predicted and experienced ratings were significantly different for all the subcategories regardless of the rating timing conditions, for task enjoyment, task engagement, and task boredom, the effect was stronger after the task completion ($dz = 1.08$, 1.04, and 1.04) in comparison with the effect during the waiting period ($dz = .54$, .67, and .77). This difference was not clear for task interest ($dz = 1.05$ after task completion and $dz = .90$ during the task).

Experiment 5

What is the behavioral implication of this underestimation effect? In the real world, people make choices and decisions based on their subjective, predicted value. If people systematically underestimate the intrinsically motivating nature of free thinking, they may actively avoid such situations. In Experiment 5, we aimed to address this potential consequence of the underestimation effect by including a task choice phase (like Experiment 3) in which participants can choose between two different waiting tasks: (a) thinking-only condition, in which participants were only allowed to think (i.e., the condition identical with Experiment 1A); and (b) news-checking condition, in which participants were allowed to check Internet news during the waiting period.

Method

Participants

Sixty-three Japanese undergraduate students participated. They were assigned randomly to the thinking-only or news-checking condition. The data from three participants were excluded before data analysis because there was some interruption during the waiting period, leaving data from 60 participants for the main analysis (female = 17, $M_{age} = 18.92$ years, $SD = 1.15$). The sample size was predetermined in the following way: We expected that the participants in the Internet condition would show no difference between the predicted motivation and experienced motivation (i.e., $dz = .00$) whereas participants in the wait condition would show the same effect size as Experiment 1A (i.e., $dz = .73$). We then set the minimal sample size that achieved at least 80% statistical power to detect the difference of these effect sizes (i.e., interaction effect) and collected data as much as possible within a term period. The final sample size was just below the desirable number and gives us the statistical power of 79%.

Table 5

Linear Mixed Model With Rating Type (Predicted Motivation Versus Experienced Motivation: Effect Coded), Interim Rating Time (Continuous Variable: 2–18) and Their Interaction as Fixed Effects; Conducted Separately for Each Subcategory of Task Motivation in Experiment 4

Independent variables	β	*SE*	95% CI	*t*	*p*
Rating type					
Enjoyment	0.56	0.18	[0.22, 0.90]	3.19	.003
Engagement	0.74	0.18	[0.39, 1.09]	4.12	<.001
Interest	0.91	0.17	[0.58, 1.23]	5.47	<.001
Boredom	1.04	0.22	[−1.47, −0.61]	4.71	<.001
Interim rating time					
Enjoyment	0.03	0.04	[0.06, 0.11]	0.60	.550
Engagement	0.02	0.04	[0.09, 0.05]	0.50	.618
Interest	0.04	0.04	[0.05, 0.12]	0.83	.409
Boredom	0.03	0.05	[0.12, 0.06]	0.70	.490
Rating Type × Interim Rating Time					
Enjoyment	0.04	0.04	[0.11, 0.03]	1.21	.233
Engagement	0.01	0.04	[0.08, 0.06]	0.32	.749
Interest	0.03	0.03	[0.09, 0.04]	0.80	.429
Boredom	0.03	0.05	[0.06, 0.12]	0.73	.471

Note. CI = confidence interval.

Measures

We used the same 12-item intrinsic motivation scale as Experiment 1A (Chronbach's αs = .91 ∼ .94). Like Experiment 3, participants also rated the same two postexperiment questions for exploratory purposes: their feelings of wasting time and refreshment. Descriptive results on these exploratory items are reported in online supplemental materials.

Procedure

At the beginning of the experiment, participants were told that their task was to wait for 20 min in a room by themselves and were told about the thinking-only condition and news-checking condition. The thinking-only condition was identical with the one used in Experiment 1A. The news-checking condition allowed participants to browse some Internet news sites with a PC placed in the room (the computer did not display a clock). They were not allowed to check other web sites such as watching videos on the Internet. Participants were then asked to make a prediction about their task motivation at the end of the 20-min of waiting task, for both conditions separately.

After the prediction ratings, participants were told that they could choose one of the two lotteries that determine which condition the participants would be assigned to. The first lottery had a 70% chance of being assigned to the thinking-only condition and a 30% chance of being assigned to the news-checking condition (thinking-only dominant lottery); on the other hand, the second lottery had a 70% chance of being assigned to the news-checking condition and a 30% chance of being assigned to the thinking-only condition (news-checking dominant lottery). After participants' choice, they were informed about the condition that they were assigned to. Unbeknownst to participants, the outcome of the lottery was predetermined—half of the participants were assigned to the thinking-only condition whereas the other half were assigned to the news-checking condition, regardless of their choice. Like Experiment 3, this procedure ensured the random assignment of participants while enabling us to probe participants' preference.

Participants then moved to a room after removing all of their belongings (including mobile phone and wristwatch) and spent 20 min alone in the room with or without Internet news, depending on the assigned condition. After the waiting period, they rated their experienced task motivation.

Table 6

Comparison Between Predicted and Experienced Subcategories of Task Motivation in Experiment 4

Subcategories	Mean prediction (*SD*)	Mean experience (*SD*)	*t*	*p*	Effect size (Cohen's *dz*)
Enjoyment					
Upon completion	3.06 (1.16)	4.34 (1.25)	6.82	<.001	1.08
During waiting	3.41 (1.13)	4.00 (1.43)	3.38	.002	0.54
Engagement					
Upon completion	4.14 (1.00)	5.31 (1.14)	6.56	<.001	1.04
During waiting	4.19 (0.91)	4.94 (1.21)	4.26	<.001	0.67
Interest					
Upon completion	3.33 (1.12)	4.48 (1.35)	6.62	<.001	1.05
During waiting	3.29 (1.21)	4.22 (1.34)	5.68	<.001	0.90
Boredom					
Upon completion	5.03 (1.26)	3.50 (1.58)	6.60	<.001	1.04
During waiting	4.82 (1.2)	3.76 (1.56)	4.90	<.001	0.77

Note. Standard deviations of the group means are reported in parentheses.

Results

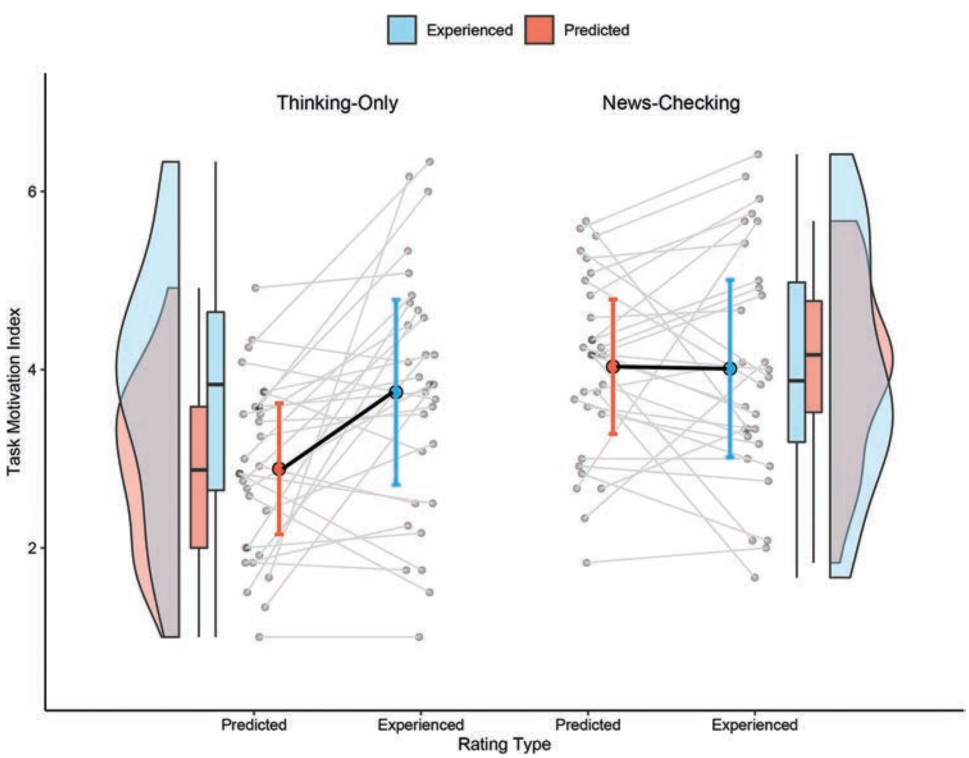
Consistent with our expectation, participants predicted that the news-checking condition (M = 4.24, SD = 1.08) would be more enjoyable and engaging than the thinking-only condition (M = 2.80, SD = 1.05), $t(59)$ = 10.95, $p < .001$, dz = 1.41, and preferred the news-checking dominant lottery (87%) more than the thinking-only dominant lottery (13%), $p < .001$ (binomial test). In the following analyses, we focused on participants' predictive ratings for the task to which they were actually assigned (e.g., if a participant was assigned to the news-checking condition, predicted intrinsic motivation for the news-checking condition was used).

A 2 (Activity condition: thinking-only vs. news-checking) × 2 (Rating type: predicted motivation vs. experienced motivation) mixed ANOVA on task motivation showed the main effect of Rating type, $F(1, 58) = 6.62$, $p = .013$, $\eta^2_G = .03$, indicating that participants generally underestimated the task motivation during the waiting period (predicted intrinsic motivation: M = 3.46, SD = 1.15; experienced task motivation: M = 3.88, SD = 1.36). The main effect of Activity condition was also statistically significant, $F(1, 58) = 7.27$, $p = .009$, $\eta^2_G = .08$. Participants gave overall higher ratings for the news-checking condition (M = 4.02, SD = 1.17) compared with the thinking-only condition (M = 3.32, SD = 1.27). Critically, these main effects were qualified by a significant Activity × Rating Type interaction, $F(1, 58) = 7.34$, $p = .009$, $\eta^2_G = .03$. As can be seen in Figure 6, although the predicted task motivation was significantly lower in the thinking-only condition (M = 2.89, SD = .98) than the news-checking condition (M = 4.03, SD = 1.01), $t(57.95) = 4.45$, $p < .001$, dz = 1.15 (Welch's test), the experienced task motivation did not show a significant difference between the thinking-only condition (M = 3.75, SD = 1.39) and the news-checking condition (M = 4.01, SD = 1.33), $t(57.89)$ = .75, p = .455, dz = .19 (Welch's test). We also conducted additional analyses to confirm that the results cannot be explained by cognitive dissonance (see online supplemental materials).

We repeated the analysis for each subcategory of task motivation (i.e., enjoyment, engagement, interest, and boredom). The main effect of rating type was significant for task engagement, $F(1, 58) = 5.59$, $p = .021$, $\eta^2_G = .03$, task interest $F(1, 58) = 5.28$, $p = .025$, $\eta^2_G = .02$, and task boredom, $F(1, 58) = 5.61$, $p = .021$, $\eta^2_G = .03$, but not task enjoyment, $F(1, 58) = 1.44$, $p = .235$, η^2_G = .01. The main effect of activity condition was statistically significant for task enjoyment, $F(1, 58) = 7.52$, $p = .008$, $\eta^2_G = .08$, task engagement, $F(1, 58)$ =, 5.42, $p = .023$, $\eta^2_G = .06$ and task boredom, $F(1, 58) = 5.14$, $p = .027$, $\eta^2_G = .06$, but not significant for task interest, $F(1, 58) = 3.45$, $p = .068$, $\eta^2_G = .04$. Critically, the interaction effect (Activity Condition × Rating Type) was significant for task enjoyment, $F(1, 58) = 8.89$, $p = .004$, $\eta^2_G = .04$, task interest, $F(1, 58) = 4.71$, $p = .034$, $\eta^2_G = .02$, and task boredom, $F(1, 58) = 4.94$, $p = .030$, $\eta^2_G = .02$., but was not significant for

Figure 6

Comparison Between Predicted and Experienced Task Motivation as a Function of Activity Condition in Experiment 5



Note. Distributions are shown in raincloud plots and data points have been jittered to enhance visual clarity. Bold dots indicate means and bold bars indicate standard errors. See the online article for the color version of this figure.

task engagement, $F(1, 58) = 2.43$, $p = .124$, $\eta^2_G = .01$. These results indicate converging evidence for our main findings in all types of task motivation except task engagement.

To clarify the nature of interaction effect (and associated effect sizes), Table 7 reports the comparison between predicted and experienced ratings of each subcategory of task motivation, separately for thinking-only and news-checking conditions. The results were consistent across the subcategories: Participants in the Thinking-only condition experienced more enjoyment, more engagement, more interest, and less boredom than they predicted, $t(29) =$ 2.64, 2.81, 2.94, and 2.92, respectively, $dz = .48$, .51, .54, and .53. On the contrary, participants in the News-checking condition did not show significant differences between experienced and predicted enjoyment, engagement, interest, and boredom, and the effect sizes were all small (or in the opposite direction), $t(29) =$ 1.46, .56, .10, and .12, respectively, $dz = -.27$, .10, .02, and .02. The overall pattern was the same even for task engagement, which did not show a significant interaction effect between Activity condition and Rating type.

Meta-Analysis

In all experiments, we consistently found that people have more task motivation than they predicted. To provide more reliable estimate of the effect size of this phenomenon, we conducted a meta-analysis. In this meta-analysis, we discarded the news-checking condition in Experiment 5 as participants in this condition engaged in a task, not in free-thinking activities. We further discarded the data from Experiment 1B and from the during-waiting condition in Experiment 4, as the target effects were different. We computed the standard errors of each of the effect sizes and conducted a random effect meta-analysis using "metafor" package in R (Viechtbauer, 2010). Figure 7 showed the forest plot. The integrated effect size was $dz = -.87$, 95% CI [−1.14, −.60].

General Discussion

We consistently found that participants underestimated their task enjoyment for and task engagement in the waiting task. Although our data were mainly from university students, the effect was relatively large, and was reliably observed regardless of the cultural background of participants (Japan and United Kingdom), the independence of predictive rating (Experiment 1B), the amount of sensory input (Experiment 2), duration of the waiting task (Experiment 3), and timing of assessment (Experiment 4). We also found that participants proactively avoided just waiting in favor of checking the Internet, even though experienced task motivation was not statistically different between these two activities (Experiment 5). Just as previous research has found, thinking alone is not that enjoyable (Wilson, Reinhard, et al., 2014). However, it is not as bad as people naively expect.

The current findings have several theoretical and practical implications. For example, the current study sheds new light on recent studies claiming that affective experiences of mind wandering or just thinking are rather aversive. Previous studies have shown that people do not enjoy just thinking (Wilson, Reinhard, et al., 2014), but it has not been clear whether people expect to enjoy the experience or whether those expectations are accurate. Again, we are not arguing that the activity of thinking is hugely fun, thrilling, or enjoyable. In fact, although it is difficult to interpret the absolute level of Likert-scale ratings (see Blanton & Jaccard, 2006), consistent with previous studies (Buttrick et al., 2019; Wilson, Reinhard, et al., 2014; Wilson, Gilbert, et al., 2014), across all studies, the average rating of experienced task motivation was around the midpoint.[1] However, the fact that participants underestimated the potential task motivation during the thinking period may indicate that people avoid thinking not because it *is* aversive but because it is *expected* to be aversive.

Such an inaccurate expectation may lead people to unnecessarily avoid spending time thinking in their daily life. For example, the current widespread availability of the Internet and mobile phones makes it extremely easy for people to kill time when they have nothing to do, and our results suggest that people's continual engagement in electronic devices may in part reflect inaccurate metacognitive beliefs about the value of not doing anything. It is also worth noting that the avoidance of thinking may entail some opportunity cost, because previous studies have suggested a variety of benefits for different types of thinking activities. For example, Schacter (2012) argued that episodic future thinking has adaptive functioning by simulating mental futures, which promotes goal directed cognition (e.g., problem solving), learning, and psychological well-being. Sedikides and Wildschut (2018) provided a list of evidence that nostalgic thinking can help people find meaning of life, enhancing goal pursuit and vitality. Research on mind wandering indicated that self-generated thoughts during mind wandering facilitates creativity and relieves negative emotions (Baird et al., 2012; Smallwood & Schooler, 2015). By actively avoiding thinking activities, people may miss these important benefits.

Another interesting observation from the current study is that the underestimation effect was found both in the affective (i.e., enjoyment, interest, and lack of boredom) and motivational (i.e., engagement) components of task motivation. There is a great deal of work focusing on the metacognitive accuracy on affective components (i.e., affective forecasting). However, while recent years have seen increased interest in the field of metamotivation examining the metacognitive aspect of motivation (Miele & Scholer, 2018; Murayama et al., 2016; Scholer & Miele, 2016), the number of studies is still limited, and there has been little research that examined both affective and motivational components in metacognitive processes. One exception, Kuratomi et al. (2018) found that participants tend to underestimate task motivation for a relatively boring task, but this effect was more explicit in task engagement. Kuratomi et al.'s (2018) findings suggest the potential dissociation between metacognitive accuracy for affective and motivational components (see also Goldsmith & Dhar, 2013), somewhat showing a contrast with the current findings. One critical difference between the thinking activities and the tasks used in Kuratomi et

[1] One might argue that the observed underestimation effect could be explained by the regression to the mean: Because people had the general low ratings for predicted intrinsic motivation for waiting period, their ratings for experienced intrinsic motivation regressed toward the midpoint of the scale. However, this interpretation is based on the misunderstanding of regression to the mean. Regression to the mean happens for the data points that have *relatively* low (or high) observed values in the sample population. These *selected* observed values in the sample tend to regress toward the overall sample mean. However, it cannot explain the change of the overall mean of the *entire* sample toward the scale midpoint.

Table 7
Comparison Between Predicted and Experienced Subcategories of Task Motivation in Experiment 5

Subcategories	Mean prediction (*SD*)	Mean experience (*SD*)	*t*	*p*	Effect size (Cohen's *dz*)
Enjoyment					
Thinking-only	3.03 (1.23)	3.79 (1.37)	2.64	.013	0.48
News-checking	4.33 (1.13)	4.01 (1.39)	1.46	.156	−0.27
Engagement					
Thinking-only	2.69 (1.07)	3.56 (1.64)	2.81	.009	0.51
News-checking	3.73 (1.32)	3.91 (1.66)	0.56	.578	0.10
Interest					
Thinking-only	2.48 (1.01)	3.26 (1.60)	2.94	.006	0.54
News-checking	3.40 (1.27)	3.42 (1.33)	0.10	.922	0.02
Boredom					
Thinking-only	4.66 (1.70)	3.61 (1.88)	2.92	.007	0.53
News-checking	3.33 (1.36)	3.30 (1.62)	0.12	.907	0.02

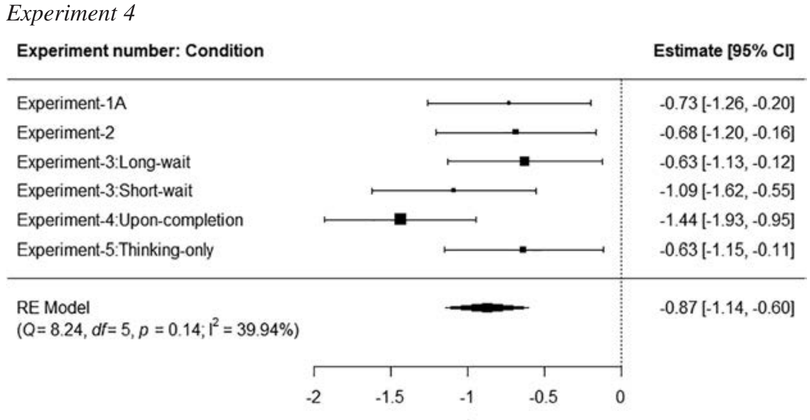
Note. Standard deviations of the group means are reported in parentheses.

al.'s (2018) work is that the latter has a clear task structure and feedback, making it relatively easy for people to engage in the task without the aid of positive affective feelings. Future studies should examine the potential similarities and differences between these two different types of metacognitive processes using tasks that can separate out these components. Integrative understanding of the cognitive, affective, and motivational aspects of metacognition has significant implications for our understanding of self-regulation in general (Inzlicht et al., 2021).

Our findings were robust across two different cultural populations (Japan and the United Kingdom). This is consistent with previous findings on the cross-cultural consistency on people's perceived enjoyment during waiting periods (Buttrick et al., 2019), but the current study extended its scope from actual experiences to metacognitive accuracy about just thinking. Our results also join recent emerging findings showing the cross-cultural consistency of metamotivational beliefs on other domains (Murayama et al., 2016; Nguyen et al., 2020). However, one notable observation is that, while we found a consistent relative underestimation effect across Japanese and United Kingdom samples, the Japanese sample showed lower absolute magnitude of ratings than the United Kingdom sample for both predicted and experienced task motivation (see Figures 2–4 and 6 for Japanese sample and Figure 5 for the United Kingdom sample). A similar pattern was also observed in the aforementioned study (Buttrick et al., 2019), where the Japanese sample showed lower experienced enjoyment for the waiting task compared with other Western and South American countries. Meanwhile, participants from all countries perceived the waiting period as relatively less enjoyable than doing entertainment activities (although the United Kingdom was not included in this study). These observations indicate the possibility that cross-cultural differences in motivational functioning emerges only at the intercept level (i.e., the default reaction to tasks and activities), but not at the function level (i.e., relative universality; Pekrun, 2006).

Future studies need to address more precise mechanisms underlying the underestimation effect. As noted in the introduction, the reward-learning framework of knowledge acquisition indicates that underestimation happens due to the inherent difficulty in assessing the value of abstract thoughts and knowledge that are produced during the thinking process (Murayama, 2022). But literature on affective forecasting provides more specific mechanisms that explain the source of the difficulty, which can help us

Figure 7
Forest Plot of Meta-Analysis of All Experiments Except for Experiment 1B, News-Checking Condition in Experiment 5, and During-Waiting Condition in Experiment 4



understand the findings in more depth. For example, previous work on affective forecasting showed that people overestimate the intensity and duration of positive and negative emotions (Ubel et al., 2005; Wilson et al., 2000; Wilson & Gilbert, 2005). This metacognitive inaccuracy has been explained by "focalism." This is a tendency to focus too much attention on the event to be forecasted, which prevents people from recognizing the degree to which one is prepared to understand future events or appreciating our capability to fight negative emotion (Wilson & Gilbert, 2005). In the present study, it is possible that participants focused too much on the *negative* aspect of just thinking when making a prediction, leading to the underestimation of positive affective feelings (and perhaps engagement).

In fact, whereas previous studies asked participants to think for pleasure explicitly by telling them to entertain themselves with their thoughts (Alahmadi et al., 2017; Buttrick et al., 2019; Westgate & Wilson, 2018; Westgate et al., 2017; Wilson, Reinhard, et al., 2014; Wilson, Gilbert, et al., 2014), the instructions for the current experiments were relatively neutral (see Method in Experiments 1a and 4; the original instructions are available on the OSF), potentially making it easy for participants to focus on negative aspects. This difference in instructions might also explain why a few previous studies did not observe the underestimation effect (Wilson, Reinhard, et al., 2014). Although the event of "just thinking" may not be negative enough (see Fox et al., 2014) to cause focalism in comparison with the negative events typically examined in the literature of affective forecasting (e.g., dissolution of a romantic relationship, the failure to achieve tenure, negative personality feedback, etc.; Gilbert et al., 1998), this possibility is worth examining in future studies.

Relatedly, another important avenue for future research is to examine the content of people's free thinking that makes the waiting time intrinsically motivating. As noted earlier, not all thinking is intrinsically rewarding. Thinking for pleasure requires some mental resources (Alahmadi et al., 2017; Westgate et al., 2017), and there are some individuals who are prone to a vicious cycle of negative thinking (e.g., rumination; Nolen-Hoeksema et al., 2008). Self-relevant thoughts should have a bigger impact on the emotional experience (Sedikides & Strube, 1997), and rewarding value of generative thoughts should also be dependent on several factors such as uncertainty reduction (Marvin & Shohamy, 2016). A systematic content analysis (e.g., using think-aloud method) could provide insights into the potential mechanisms underlying the underestimation effect.

Future studies would also do well to examine the durability of the underestimation effect. We found that the underestimation effect was observed even if the expected thinking time was relatively short (e.g., 3 min), but there were two limitations. First, all of the time manipulations were performed in a between-subjects design (i.e., Experiments 3 and 4); thus, the studies have relatively low statistical power to detect a potential time effect. Second, we did not examine the effect for the expected waiting period beyond 20 min. Although it may be true that thinking activities facilitate self-generation of internal reward (Murayama, 2022), this would not last forever, because just thinking does not involve any input of knowledge from the external environment. As classic work of sensory deprivation revealed (e.g., Bexton et al., 1954), people suffer from relentlessness and emotional lability when they are completely isolated from external input for an enduring amount of time (e.g., 2–3 days). We chose a 20-min maximum waiting period because thinking freely without external stimulation for this time-period is relatively uncommon. However, given that people's metacognitive prediction is generally insensitive to the length of events (Koriat et al., 2004; Wilson et al., 2000), there may be a point at which the underestimation dissipates.

We are constantly making an implicit decision between doing something and doing nothing in our daily life. However, researchers of decision making have focused almost exclusively on the comparison between concrete options (e.g., different amount or type of reward). Doing nothing, or just waiting has been used only as a cost (i.e., loss; e.g., Leclerc et al., 1995), and has not been given substantial meaning beyond that. This assumption is true in that waiting time is not generally perceived as intrinsically motivating, but at the same time, such metacognitive assessment is unlikely to accurately reflect the actual experience of waiting (Murayama et al., 2016; Scholer et al., 2018). The current findings confirmed the possibility. We believe that the discovery of this systematic metacognitive bias for just thinking provides an important first step toward establishing a more comprehensive and ecologically valid model of human decision making.

Context of the Research

This study was inspired by two strands of research by the authors. First, the authors have been interested in people's intrinsic motivation and curiosity, examining the psychological mechanisms of how people can sustain long-term commitment without extrinsic incentives (see Murayama, 2022). In this context, we were greatly inspired by Wilson, Reinhard, et al.'s (2014) work, which showed that people have moderately aversive feelings during just thinking. The results were particularly interesting because they ran somewhat contrary to our intuition—we are often engaged in thinking so deeply that we lose track of time. Second, some of the authors have a background in cognitive psychology and were interested in applying the idea of metacognitive accuracy (e.g., Kornell & Bjork, 2008) to the domain of motivation, which we called *metamotivation* (Scholer et al., 2018). Our first set of studies showed that people tend to underestimate their task motivation (Kuratomi et al., 2018), and these findings let us to consider whether a similar underestimation effect could happen for the thinking task used by Wilson, Reinhard, et al. (2014). Thus, the current study represents a synergy between research on intrinsic motivation or curiosity and metacognition.

References

Alahmadi, S., Buttrick, N. R., Gilbert, D. T., Hardin, A. M., Westgate, E. C., & Wilson, T. D. (2017). You can do it if you really try: The effects of motivation on thinking for pleasure. *Motivation and Emotion*, *41*(5), 545–561. https://doi.org/10.1007/s11031-017-9625-7

Baird, B., Smallwood, J., Mrazek, M. D., Kam, J. W. Y., Franklin, M. S., & Schooler, J. W. (2012). Inspired by distraction: Mind wandering facilitates creative incubation. *Psychological Science*, *23*(10), 1117–1122. https://doi.org/10.1177/0956797612446024

Baranes, A. F., Oudeyer, P.-Y., & Gottlieb, J. (2014). The effects of task difficulty, novelty and the size of the search space on intrinsically motivated exploration. *Frontiers in Neuroscience*, *8*, 317. https://doi.org/10.3389/fnins.2014.00317

Bexton, W. H., Heron, W., & Scott, T. H. (1954). Effects of decreased variation in the sensory environment. *Canadian Journal of Psychology*, *8*(2), 70–76. https://doi.org/10.1037/h0083596

Bishai, D. M., & Lang, H. C. (2000). The willingness to pay for wait reduction: The disutility of queues for cataract surgery in Canada, Denmark, and Spain. *Journal of Health Economics*, *19*(2), 219–230. https://doi.org/10.1016/S0167-6296(99)00024-7

Bjork, R. A., Dunlosky, J., & Kornell, N. (2013). Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*, *64*, 417–444. https://doi.org/10.1146/annurev-psych-113011-143823

Blanton, H., & Jaccard, J. (2006). Arbitrary metrics in psychology. *The American Psychologist*, *61*(1), 27–41. https://doi.org/10.1037/0003-066X.61.1.27

Buttrick, N., Choi, H., Wilson, T. D., Oishi, S., Boker, S. M., Gilbert, D. T., Alper, S., Aveyard, M., Cheong, W., Čolić, M. V., Dalgar, I., Doğulu, C., Karabati, S., Kim, E., Knežević, G., Komiya, A., Laclé, C. O., Ambrosio Lage, C., Lazarević, L. B., . . . Wilks, D. C. (2019). Cross-cultural consistency and relativity in the enjoyment of thinking versus doing. *Journal of Personality and Social Psychology*, *117*(5), e71–e83. https://doi.org/10.1037/pspp0000198

Chater, N., & Loewenstein, G. (2016). The under-appreciated drive for sense-making. *Journal of Economic Behavior & Organization*, *126*, 137–154. https://doi.org/10.1016/j.jebo.2015.10.016

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review*, *107*(2), 261–288. https://doi.org/10.1037/0033-295X.107.2.261

Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology*, *70*(3), 461–475. https://doi.org/10.1037/0022-3514.70.3.461

Fox, K. C., Thompson, E., & Christoff, K. (2014). Is thinking really aversive? A commentary on Wilson et al.'s "Just think: The challenges of the disengaged mind". *Frontiers in Psychology*, *5*, 1427.

Frederick, S., Loewenstein, G., & O'Donoghue, T. (2002). Time discounting and time Preference: A critical review. *Journal of Economic Literature*, *40*(2), 351–401. https://doi.org/10.1257/jel.40.2.351

Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *75*(3), 617–638. https://doi.org/10.1037/0022-3514.75.3.617

Goldsmith, K., & Dhar, R. (2013). Negativity bias and task motivation: Testing the effectiveness of positively versus negatively framed incentives. *Journal of Experimental Psychology: Applied*, *19*(4), 358–366. https://doi.org/10.1037/a0034415

Gruber, M. J., Gelman, B. D., & Ranganath, C. (2014). States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit. *Neuron*, *84*(2), 486–496. https://doi.org/10.1016/j.neuron.2014.08.060

Hsee, C. K., Yang, A. X., & Wang, L. (2010). Idleness aversion and the need for justifiable busyness. *Psychological Science*, *21*(7), 926–930. https://doi.org/10.1177/0956797610374738

Inzlicht, M., Werner, K. M., Briskin, J. L., & Roberts, B. W. (2021). Integrating models of self-regulation. *Annual Review of Psychology*, *72*, 319–345. https://doi.org/10.1146/annurev-psych-061020-105721

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, *10*(12), 1625–1633. https://doi.org/10.1038/nn2007

Koriat, A., Bjork, R. A., Sheffer, L., & Bar, S. K. (2004). Predicting one's own forgetting: The role of experience-based and theory-based processes. *Journal of Experimental Psychology: General*, *133*(4), 643–656. https://doi.org/10.1037/0096-3445.133.4.643

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science*, *19*(6), 585–592. https://doi.org/10.1111/j.1467-9280.2008.02127.x

Kuratomi, K., Johnsen, L., Kitagami, S., Hatano, A., & Murayama, K. (2018, July 24). *People underestimate their capability to motivate themselves without performance-based extrinsic incentives*. https://doi.org/10.31219/osf.io/5uvyf

Leclerc, F., Schmitt, B. H., & Dube, L. (1995). Waiting time and decision making: Is time like money? *Journal of Consumer Research*, *22*(1), 110. https://doi.org/10.1086/209439

Marvin, C. B., & Shohamy, D. (2016). Curiosity and reward: Valence predicts choice and information prediction errors enhance learning. *Journal of Experimental Psychology: General*, *145*(3), 266–272. https://doi.org/10.1037/xge0000140

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2007). Time discounting for primary rewards. *The Journal of Neuroscience*, *27*(21), 5796–5804. https://doi.org/10.1523/JNEUROSCI.4246-06.2007

Miele, D. B., & Scholer, A. A. (2018). The Role of Metamotivational monitoring in motivation regulation. *Educational Psychologist*, *53*(1), 1–21. https://doi.org/10.1080/00461520.2017.1371601

Murayama, K. (2022). A reward-learning framework of knowledge acquisition: An integrated account of curiosity, interest, and intrinsic-extrinsic rewards. *Psychological Review*, *129*(1), 175–198. https://doi.org/10.1037/rev0000349

Murayama, K., FitzGibbon, L., & Sakaki, M. (2019). Process account of curiosity and interest: A reward-learning perspective. *Educational Psychology Review*, *31*(4), 875–895. https://doi.org/10.1007/s10648-019-09499-9

Murayama, K., Kitagami, S., Tanaka, A., & Raw, J. A. L. (2016). People's naiveté about how extrinsic rewards influence intrinsic motivation. *Motivation Science*, *2*(3), 138–142. https://doi.org/10.1037/mot0000040

Nguyen, T., Togawa, T., Scholer, A. A., & Fujita, K. (2020). A cross-cultural investigation of metamotivational knowledge of construal level in the United States and Japan. *Motivation Science*, *6*(4), 386–400. https://doi.org/10.1037/mot0000178

Nolen-Hoeksema, S., Wisco, B. E., & Lyubomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, *3*(5), 400–424. https://doi.org/10.1111/j.1745-6924.2008.00088.x

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, *37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Ruan, B., Hsee, C. K., & Lu, Z. Y. (2018). The teasing effect: An underappreciated benefit of creating and resolving an uncertainty. *Journal of Marketing Research*, *55*(4), 556–570. https://doi.org/10.1509/jmr.15.0346

Sansone, C., Weir, C., Harpster, L., & Morgan, C. (1992). Once a boring task always a boring task? Interest as a self-regulatory mechanism. *Journal of Personality and Social Psychology*, *63*(3), 379–390. https://doi.org/10.1037/0022-3514.63.3.379

Schacter, D. L. (2012). Adaptive constructive processes and the future of memory. *American Psychologist*, *67*(8), 603–613. https://doi.org/10.1037/a0029869

Schacter, D. L., Benoit, R. G., & Szpunar, K. K. (2017). Episodic future thinking: Mechanisms and functions. *Current Opinion in Behavioral Sciences*, *17*, 41–50. https://doi.org/10.1016/j.cobeha.2017.06.002

Scholer, A. A., & Miele, D. B. (2016). The role of metamotivation in creating task-motivation fit. *Motivation Science*, *2*(3), 171–197. https://doi.org/10.1037/mot0000043

Scholer, A. A., Miele, D. B., Murayama, K., & Fujita, K. (2018). New directions in self-regulation: The role of metamotivational beliefs.

Current Directions in Psychological Science, *27*(6), 437–442. https://doi.org/10.1177/0963721418790549

Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be better. *Advances in Experimental Social Psychology*, *29*, 209–269. https://doi.org/10.1016/S0065-2601(08)60018-0

Sedikides, C., & Wildschut, T. (2018). Finding meaning in nostalgia. *Review of General Psychology*, *22*(1), 48–61. https://doi.org/10.1037/gpr0000109

Sedikides, C., Wildschut, T., Arndt, J., & Routledge, C. (2008). Nostalgia: Past, present, and future. *Current Directions in Psychological Science*, *17*(5), 304–307. https://doi.org/10.1111/j.1467-8721.2008.00595.x

Singer, J. L. (1975). *Daydreaming and fantasy*. Routledge.

Smallwood, J., & Schooler, J. W. (2015). The science of mind wandering: Empirically navigating the stream of consciousness. *Annual Review of Psychology*, *66*(1), 487–518. https://doi.org/10.1146/annurev-psych-010814-015331

Ubel, P. A., Loewenstein, G., & Jepson, C. (2005). Disability and sunshine: Can hedonic predictions be improved by drawing attention to focusing illusions or emotional adaptation? *Journal of Experimental Psychology. Applied*, *11*(2), 111–123. https://doi.org/10.1037/1076-898X.11.2.111

van Langen, J. (2020). *Open-visualizations in R and Python*. https://github.com/jorvlan/open-visualizations

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Westgate, E. C., & Wilson, T. D. (2018). Boring thoughts and bored minds: The MAC model of boredom and cognitive engagement. *Psychological Review*, *125*(5), 689–713. https://doi.org/10.1037/rev0000097

Westgate, E. C., Wilson, T. D., & Gilbert, D. T. (2017). With a little help for our thoughts: Making it easier to think for pleasure. *Emotion*, *17*(5), 828–839. https://doi.org/10.1037/emo0000278

Wigfield, A., & Eccles, J. S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wilson, T. D., & Gilbert, D. T. (2005). Affective forecasting: Knowing what to want. *Current Directions in Psychological Science*, *14*(3), 131–134. https://doi.org/10.1111/j.0963-7214.2005.00355.x

Wilson, T. D., Centerbar, D. B., Kermer, D. A., & Gilbert, D. T. (2005). The pleasures of uncertainty: Prolonging positive moods in ways people do not anticipate. *Journal of Personality and Social Psychology*, *88*(1), 5–21. https://doi.org/10.1037/0022-3514.88.1.5

Wilson, T. D., Gilbert, D. T., Reinhard, D. A., Westgate, E. C., & Brown, C. L. (2014). Would you fund this movie? A reply to Fox et al. (2014). *Frontiers in Psychology*, *5*, 1428. https://doi.org/10.3389/fpsyg.2014.01428

Wilson, T. D., Wheatley, T., Meyers, J. M., Gilbert, D. T., & Axsom, D. (2000). Focalism: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, *78*(5), 821–836. https://doi.org/10.1037/0022-3514.78.5.821

Wilson, T. D., Reinhard, D. A., Westgate, E. C., Gilbert, D. T., Ellerbeck, N., Hahn, C., Brown, C. L., & Shaked, A. (2014). Social psychology. Just think: The challenges of the disengaged mind. *Science*, *345*(6192), 75–77. https://doi.org/10.1126/science.1250830

Woolley, K., & Fishbach, A. (2015). The experience matters more than you think: People value intrinsic incentives more inside than outside an activity. *Journal of Personality and Social Psychology*, *109*(6), 968–982. https://doi.org/10.1037/pspa0000035

Yan, V. X., Thai, K.-P., & Bjork, R. A. (2014). Habits and beliefs that guide self-regulated learning: Do they vary with mindset? *Journal of Applied Research in Memory & Cognition*, *3*(3), 140–152. https://doi.org/10.1037/h0101799

Received September 21, 2021
Revision received March 29, 2022
Accepted April 13, 2022 ■