Title
Cognitive Load In Speed-Accuracy Tradeoff: Theoretical and Empirical Evidence Based on Resource-Rational Analyses

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 46(0)

Authors
Shirasuna, Masaru
Kagawa, Rina
Honda, Hidehito

Publication Date
2024

Peer reviewed

Cognitive Load In Speed-Accuracy Tradeoff: Theoretical and Empirical Evidence Based on Resource-Rational Analyses

Masaru Shirasuna (m.shirasuna1392@gmail.com)
Department of Psychology, Otemon Gakuin University
2-1-15, Nishiai, Ibaraki-shi, Osaka, 567-8502, Japan

Rina Kagawa† (sonata.skazka@gmail.com)
Faculty of Medicine, University of Tsukuba
1-1-1, Tennoudai, Tsukuba-shi, Ibaraki, 305-8575, Japan

Hidehito Honda (hitohonda.02@gmail.com)
Department of Psychology, Otemon Gakuin University
2-1-15, Nishiai, Ibaraki-shi, Osaka, 567-8502, Japan

Abstract

In simple judgment tasks, it is generally assumed that thinking for longer leads to more accurate judgments, providing better benefits as suggested by the speed-accuracy tradeoff framework. However, human cognitive resources are limited, and longer thinking induces cognitive costs such as subjective workload. Therefore, a total benefit should be considered under the tradeoff between thinking benefits (i.e., improving accuracy) and thinking costs (i.e., increasing cognitive load) as suggested by the resource rationality framework. We examined this issue using computer simulations and behavioral experiments. Our simulations showed that, if a thinking cost was introduced based on resource-rational approaches, there was an optimal length of time for maximizing a total benefit and the total benefit gradually decreased there. In addition, our experiments demonstrated that judgment accuracy did not always improve even if participants were provided a longer thinking time; conversely, longer thinking time was likely to increase their subjective workload. These results are consistent with resource rationality rather than speed-accuracy tradeoff. The importance of considering cognitive load is suggested to further understand human intelligence in the context of a speed-accuracy tradeoff.

Keywords: speed-accuracy tradeoff; resource rationality; cognitive resources; computer simulation; behavioral experiment

Introduction

Speed-Accuracy Tradeoff Framework

Humans process information and make judgments in the real world. Human cognitive resources, such as computational capacity and information storage, are limited and cannot handle large amounts of data. If humans make judgments hastily, they will likely be wrong. In contrast, if they take a long time to think of tasks, judgment accuracy is likely to improve.

Considering the relationship between thinking time and judgment accuracy, a *speed-accuracy tradeoff* (SAT) is a well-known framework in information processing and judgments such as perceptual decision-making, recognition memory, and inference tasks (e.g., Heitz, 2014; Henmon, 1911; Karşılar, Simen, Papadakis, & Balci, 2014; Reed, 1973; Wickelgren, 1977). SAT assumes that faster responses collect less accumulated evidence and thus less accurate decisions, and vice versa. SAT is regarded as a ubiquitous effect that is closely related to an organism's judgment processes (e.g., Chittka, Skorupski, & Raine, 2009; Heitz, 2014). SAT process is often intuitively understood in terms of sequential sampling models (e.g., Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Heitz & Schall, 2012; Ratcliff & Smith, 2004; as review, Ratcliff, Smith, Brown, & McKoon, 2016). Sequential sampling models argue that individuals set a decision threshold and make a final choice when the amount of evidence, starting from a certain baseline, reaches that threshold. They sometimes adjust their decision thresholds and baseline. Lowering the threshold or raising the baseline leads to faster responses but simultaneously to an increase in the error rate, and vice versa (e.g., Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010). Some studies also suggest that people sometimes terminate their accumulation of evidence before a decision deadline, by collapsing or converging the threshold dynamically during thinking (Frazier & Yu, 2008; Karşılar et al., 2014).

In many cases, the performance of SAT is typically depicted as a curved line (concave function). When the x- and y-axes denote the response time and accuracy, respectively, the accuracy remarkably increases at an early stage; however, the extent of the increasing accuracy gradually diminishes as time elapses (e.g., McElree & Carrasco, 1999). In SAT, a benefit obtained by thinking for a long time such as increasing accuracy (we call "thinking benefit") is regarded as equivalent to a benefit obtained through tasks (we call "total benefit") (see upper panels in Figure 1; as described later).

Resource Rationality Framework

Although many behavioral science studies have discussed SAT, the aspect of cognitive resources has been ignored. The

† The second author's current affiliation is National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

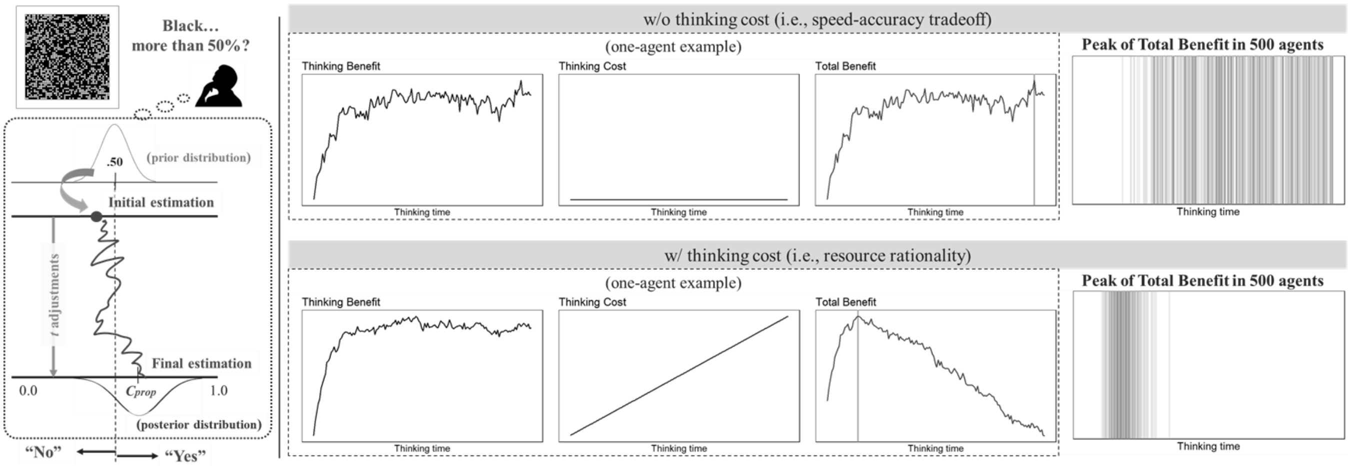

Figure 1 *Left picture*: Schematics of computer simulations of a grid task. An agent was asked to judge whether the proportion of black grids was over or under 50%. The agent's first estimation was sampled from a normal distribution (mean = .50); and after *t*-time adjustments, a final estimation was provided. Our simulations iterated this procedure 1,000 times per agent. *Right panels*: Results of computer simulations. The three left panels inside dotted frames show examples of a one-agent simulation, and vertical lines denote the times when a total benefit peaked. If a thinking cost was not assumed, a total benefit would be equivalent to a thinking benefit and did not decrease (as predicted by SAT; upper panels). If a thinking cost was introduced, the thinking cost continued to linearly increase because human cognitive resources are limited, and a total benefit gradually decreased after a peak (as predicted by resource rationality; lower panels). Right-end panels show the peak times of a total benefit (vertical lines shown in the "Total Benefit" panels) observed in 500-agents simulation. Peaks of total benefit in the resource rationality were observed in narrower ranges and at earlier times than those in SAT.

SAT has been evaluated mainly based on judgment accuracy. However, human cognitive resources are limited and thinking is considered costly (e.g., Lieder, Griffiths, Quentin, & Goodman, 2018; Shugan, 1980). Therefore, a longer thinking time may enhance more accurate judgments, but simultaneously will generate a larger subjective workload and irritation (we call "thinking cost").

Recent cognitive science studies have proposed a *resource rationality* framework (e.g., Griffiths, Lieder, & Goodman, 2015; Lieder & Griffiths, 2020). This framework helps us understand "how we think given that our time and minds are finite?" (Lieder et al., 2018). The resource-rational approach is key to modeling human minds considering their limited cognitive resources. Previous studies could better explain humans rational and accurate behaviors in various tasks such as action planning (e.g., Ho, Cohen, & Griffiths, 2023), goal pursuit (e.g., Prystawski, Mohnert, Tošic, & Lieder, 2021), and anchoring effects (e.g., Lieder et al., 2018).

In the context of the relationship between thinking time and judgment accuracy, accuracy will be maximized at a certain time point by thinking slowly, as suggested by SAT. In other words, a thinking benefit (i.e., increasing accuracy) can be regarded as a total benefit. However, the resource rational framework assumes that cognitive costs will gradually increase as a person thinks of a task for a long time. Once the accuracy reaches the peak, it is unlikely to improve further; instead, a thinking cost will become larger. Thus, the total benefit gradually decreases after the peak accuracy is observed. That is, researchers should consider the total benefit under a tradeoff between thinking benefits and thinking costs (i.e., increasing workload).

Resource Rationality in Numerical Estimations

Using computational modeling and simulations, Lieder et al. (2018) showed that anchoring bias in numerical estimations reflects people's rational use of cognitive resources. The assumptions are briefly summarized as follows: In a numerical estimation task, individuals make the first estimation based on a presented anchor (i.e., an initially presented value). Then, they repeatedly adjust their estimation, and finally provide the final estimation. In their model, the number of steps, t^*, was selected to minimize the expected value of the time cost (or thinking cost) of adjustments plus the error cost of the final estimate:

$$t^* = arg\min_t\left[E_{Q(\widehat{x_t})}\{cost(x,\hat{x}) + \gamma * t\}\right]$$

where $\hat{x}$ and $Q(\widehat{x_t})$ were the estimation and the distribution of estimations after *t*-times adjustments, respectively; x was the (unknown) true value; $cost(x,\hat{x})$ was the error cost; and γ was a thinking cost per adjustment. As described in this equation, Lieder et al. (2018) introduced a time cost based on the resource-rational framework. Generally, an increasing number of adjustments (i.e., thinking for a longer time) might lead to a reduction in error cost (i.e., increasing judgment accuracy). However, the thinking cost should increase linearly as adjustments are repeated. Thus, the total benefit would decrease after t^*-time adjustments. Lieder et al. (2018) argued that there was a point that reflects an "optimal resource allocation" under the tradeoff between judgment accuracy and time cost.

Study Outline

In this study, we first theoretically examine the differences between the SAT and resource rational frameworks through computer simulations. We show that, because of humans limited cognitive resources, there should be an optimal point of total benefit under the tradeoff between thinking time and thinking cost. Next, we empirically examine the results of the simulations using behavioral experiments. We investigate actual human behaviors and confirm that an unnecessarily long thinking time (i) does not always improve judgment accuracy and (ii) induces a larger subjective workload. Data and R codes are available at
https://osf.io/6bkt5/?view_only=a4a870d2e088456bad1c3883176b5c6b

Simulation Study

Method

Tasks and Materials Consider the following a simple perceptual binary-choice task (called a "grid task"): A person is presented with a black-and-white grid stimulus and is asked to judge whether the black grids are more than half of the whole grids (i.e., a binary choice of yes or no).

Procedure We applied Lieder et al.'s framework of resource-rational analyses (Lieder et al., 2018) to our grid task (see Figure 1 left picture; hereafter, all normal distributions in our simulations were truncated for min = 0 and max = 1).

It was assumed that a person first estimated the proportion of black grids as approximately .50. The first value, v_1, was sampled from a normal distribution (prior distribution) with mean = 0.5 and *SD* = 0.1 because the person initially had no prior knowledge. Next, it was assumed that the person repeatedly adjusted the estimation by seeing the grid stimulus for a few seconds, based on the person's own belief (b) and the correct proportion of black grids (C_{prop}). We set C_{prop} as .55 [1] and assumed that adjustments were conducted 150 times (i.e., estimating v_t at each time t = 2, 3, …, 150) using the Metropolis-Hastings algorithm (Hastings, 1970), a Markov chain Monte Carlo method (according to Lieder et al., 2018).

At each time, a potential adjustment, δ, was proposed by sampling from a normal distribution with mean = 0 and *SD* = 0.05. This adjustment was either accepted (i.e., $v_t = v_{t-1} + \delta$) or rejected (i.e., $v_t = v_{t-1}$) as the following rule. If a proposed adjustment was likely to make an estimation more probable (i.e., $Prob(X = v_{t-1} + \delta) > Prob(X = v_{t-1})$), then it was always accepted. Even if the adjustment would be less probable (i.e., $Prob(X = v_{t-1} + \delta) < Prob(X = v_{t-1})$), it was also accepted with probability $Prob(X = v_{t-1} + \delta) / Prob(X = v_{t-1})$. Otherwise, the adjustment was rejected. As increasing t, the distribution of the final estimates was assumed to converge to a posterior distribution with mean = C_{prop} and *SD* = b. We regarded v_{150} as the final estimate. If v_{150} > .50, then we regarded it as correct, and vice versa.

According to resource rationality, we introduced a thinking cost at time t, $ThinkCost_t$. Based on Lieder et al. (2018), we assumed that thinking costs increased linearly with time t. The maximum cost 0.0 (i.e., $ThinkCost_{150}$= 0.0) meant that human cognitive resources were not assumed and longer thinking led to better benefits, which reflected a traditional SAT. In contrast, we considered it more natural to assume limited cognitive resources and increase thinking costs (i.e., $ThinkCost_{150}$ > 0.0), which reflected resource rationality. In our resource rationality simulations, we set $ThinkCost_{150}$ to .50.

The above procedure was iterated 1,000 times (i.e., assuming that the grid task with 150-time adjustments was iterated 1,000 times) per agent. We then defined the mean rates of correct judgments in the 1,000 iterations for each time t as $ThinkBenefit_t$ (note that $ThinkBenefit_t$was scaled by dividing each mean rate at time t by the maximum mean rate). Finally, we calculated the mean of the differences between benefit and cost (i.e., $ThinkBenefit_t$ – $ThinkCost_t$) in 1,000 iterations for each time t, and defined it as $TotalBenefit_t$. Our simulations assumed that 500 agents conducted this grid task. We investigated the differences in $TotalBenefit_t$ as time t progressed between cases where a thinking cost was not assumed (SAT; i.e., $ThinkCost_{150}$ = 0.0) and where it was assumed (resource rationality; i.e., $ThinkCost_{150}$ = .50).

Results and Discussion

First, we show an example of the estimation in one round (i.e., 1,000 iterations * 1 agent; panels in dotted frames in Figure 1). When the maximum thinking cost was 0.0, the total benefit became equivalent to the thinking benefit. This result could be captured by the traditional SAT; in fact, the curved line depicted in the "Thinking Benefit" and "Total Benefit" panels in Figure 1 (i.e., the extent of increasing benefit per time gradually declined, but the benefit did not decrease) is a typical shape explaining SAT (e.g., Heitz, 2014; Öztekin & McElree, 2010; Wickelgren, 1977). However, when the thinking cost was introduced, the total benefit peaked at a certain time point (vertical line) and then gradually decreased. This result can be captured by resource rationality, which assumes limited human cognitive resources.

Next, we repeated this estimation for 500 agents (i.e., 1,000 iterations * 500 agents) and investigated when the peak of the total benefit was observed for each round. As shown in the right-end panels in Figure 1 (darkness of color denotes frequency), the peak times in SAT ranged widely, especially at later times (mean = 95.17, min = 23, max = 150). In contrast, peaks in resource rationality were observed in narrow ranges, especially at earlier times (mean = 23.16, min = 10, max = 51). These results of resource rationality framework

[1] In this case, the response "yes" was regarded as correct because C_{prop} = .55 was over .50. We additionally simulated the symmetrical case with respect to .50 (i.e., C_{prop} = .45) and confirmed the same tendencies described in the main text. Furthermore, we simulated other proportion cases (e.g., C_{prop} = .65, .75, .85, etc.) and obtained the same results.

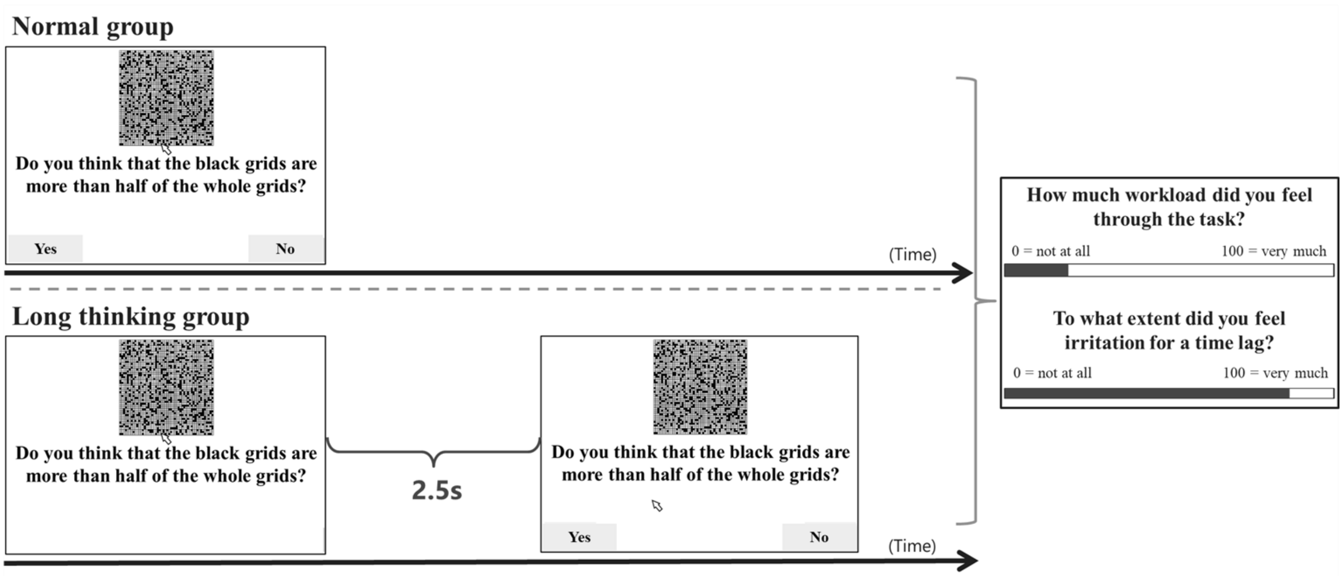

Figure 2 Schematics of behavioral experiments. In a grid task (80 questions), participants were asked to judge whether the proportion of black grids was over or under 50% (responded by clicking a "yes" or "no" button). Participants in the normal group could click a button immediately after a grid stimulus was presented. In contrast, in the long thinking group, a time lag was inserted for the first 2.5 seconds and participants could not click the buttons (i.e., a long thinking time was provided). After the grid task, participants evaluated their subjective workload through the task and irritation due to the time lag.

simulations suggest that thinking longer is not always a good strategy because a total benefit often peaks at early stages.

Behavioral Experiments

In this section, we empirically examine whether resource-rational analyses can accurately capture actual human behaviors. According to SAT, if people are provided with a longer thinking time, they will be able to make more accurate judgments. However, according to resource rationality, even if people are provided with a longer time, their judgment accuracy does not always improve; instead, their subjective workload will increase.

Based on these considerations, we manipulated the length of thinking time in behavioral experiments. In one group, participants responded to a task in a normal manner (we call "normal group"). Participants in the normal group could click a button immediately after a grid stimulus appeared. In another group, a sufficiently long time, 2.5 seconds, was inserted between a grid appearing and the two buttons appearing in every trial (we call "long thinking group"). Participants in the long thinking group could not click a button during the first 2.5s and were forced to think of the task.

We compared participants' judgment accuracy and thinking costs between the two groups. We predicted that participants' accuracy would not differ between them, but that their workload in the long thinking group would be greater than that in the normal group. If so, it is suggested that the total benefit sometimes decreases when a long thinking time is provided, which supports the resource rationality framework.

Method

Participants Fifty-five participants were assigned to normal group (n_{men} = 31, n_{women} = 24; M_{age} = 39.85, SD_{age} = 20.69), and 60 participants to long thinking group (n_{men} = 38, n_{women} = 21, n_{other} = 1; M_{age} = 41.91, SD_{age} = 8.74) via the Japanese crowdsourcing platform, Lancers (https://www.lancers.jp/). The total sample size was determined using G*Power (Faul, Erdfelder, Buchner, & Lang, 2009), assuming a medium effect size of 0.45–0.5 and a power of 0.8 for "*t*-tests" family and "Difference between two independent means" statistical test. The required sample size ranged from 102 (effect size 0.5) to 124 (effect size 0.45). Based on these estimates, approximately 60 participants were recruited per group.

Tasks, Materials, and Procedures Behavioral experiments were conducted online, using Qualtrics (Figure 2). We conducted a grid task, introduced in the computer simulations described above. First, a fixation cross was presented for 0.5s and then a grid stimulus was presented (the initial position of the mouse cursor was the center of the screen). Participants were asked to judge whether black grids were more than 50% in the stimulus and then click the "yes" or "no" button presented at the bottom of the screen. After clicking a button, a "next" button was presented. If they clicked it, a fixation cross appeared again, and the same procedure was repeated. Participants in both groups responded to 80 questions. We prepared four types of grid stimuli in terms of proportions of black grids: 35%, 45%, 55%, and 65%. Twenty stimuli (black-and-white arrangements differed among all stimuli) were used for each percentage per participant. The order of presenting grid stimuli was randomized.

After the grid task, participants were asked to subjectively evaluate their thinking costs. Specifically, they rated (a) how much subjective workload they felt throughout the task and (b) how much irritation they felt, using a visual analog scale (left: 0 [not at all] – right: 100 [very much]).

Results and Discussion

In the following analyses, we excluded trials in which the response time was longer than 6s as outliers (3SD of RT across all experiments was 5.90 and thus we defined the outlier criterion as 6s). We also excluded trials in which 35% and 65% black stimuli were presented because of an extreme ceiling effect (individual accuracy: .980).

Individuals' judgment accuracy and response time distributions are shown in the two left columns of Figure 3. We found that the grand means of individuals' accuracy (i.e., the mean of each participant's rate of correct judgments) did not differ between the two groups [2] (dotted lines; M_{normal} = .833; M_{long} = .826). To analyze accuracy in terms of the length of thinking time, we categorized times into three parts: Early (~2.5s), middle (2.5~4.0s), and late (4.0s~). We also found that, in both groups, the accuracy tended to be higher immediately after participants were allowed to click the buttons (i.e., "~2.5s" in normal group, and "2.5~4.0s" in long thinking group) and responses in later times were not always more accurate. In fact, as shown in the response time distributions, correct judgments (yellow bars) were likely to be observed frequently in the early stages and tended to decrease later.

In addition, we compared the subjective evaluations of thinking costs between the two groups (right panels inside a frame in Figure 3). We found that participants in long thinking group tended to experience greater costs than those in normal group in terms of both workload and irritation (workload: M_{normal} = 29.35, M_{long} = 37.37, W = 1317.5, p = .062, Cliff's delta = 0.202; irritation: M_{normal} = 20.29, M_{long} = 37.37, W = 780, p < .001, Cliff's delta = 0.527; Mann–Whitney U test). These results indicate that an unnecessarily long time is likely to induce thinking costs in participants.

In summary, thinking for a longer time provides a thinking benefit (i.e., more accurate judgments), but the benefit peaks at a certain time point. Simultaneously, the thinking cost (i.e., larger workload) gradually increases with time and then will exceed the benefit after the peak time of the thinking benefit. Therefore, the total benefit will gradually decrease. These behavioral data are supported and consistent with the resource rationality framework, rather than a speed-accuracy tradeoff.

General Discussion

This study focused on the relationship between thinking time and judgment accuracy. It is generally believed that a longer thinking time leads to improve accuracy (i.e., a thinking benefit can be regarded as a total benefit), as suggested by SAT. However, because human cognitive resources are limited, it is considered that thinking costs such as subjective workload will increase as a person thinks of a task for a longer time.

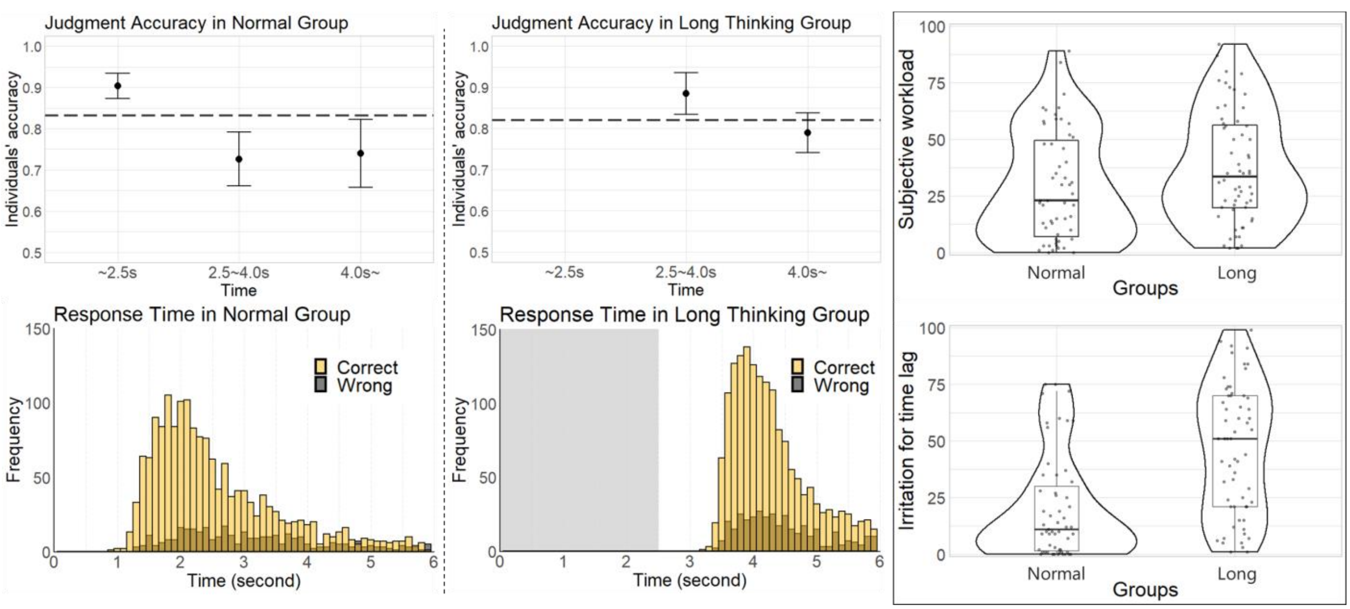

Figure 3 Results of behavioral experiments. *Left two columns*: The judgment accuracy (upper row; dots and error bars denote the means and 95% confidence intervals, respectively) and response time distributions (lower row; yellow and black denote trials where correct and wrong judgments were made, respectively, with 0.1s bin). Grand means of individual accuracy (dotted lines) did not differ between the two groups, and thinking longer did not always improve accuracy. Many correct judgments were often made in early times. In late times, correct judgments tended to decrease. *Right panels inside a frame*: Subjective ratings for thinking costs after the grid task. Participants in the long thinking group tended to feel larger workload and irritation than those in the normal group. *Remark*: Behavioral data suggested that actual human behaviors were consistent with the resource rationality framework, rather than SAT.

[2] We additionally estimated 95%CIs of participants' accuracy for each group by using R package *brms*, and confirmed that 95%CIs overlapped each other (normal group 95%CI = [.798, .869]; long thinking group 95%CI = [.794, .859]).

Thus, the total benefit is predicted to decrease after a certain time point (i.e., the total benefit should be considered under a tradeoff between thinking benefit and thinking cost), as suggested by the resource rationality framework. This study first conducted computer simulations to theoretically explain the differences between SAT and resource rationality. First, we conducted computer simulations and showed that if a thinking cost was introduced, there should be a peak in the total benefit and it gradually decreased. Next, we conducted behavioral experiments to empirically examine the findings using computer simulations. We found that even if participants were provided with a long thinking time, their accuracy did not always increase. We also found that an unnecessarily long thinking time induced a larger subjective workload. These behavioral data suggest that a total benefit will gradually decrease after the peak of thinking benefit (i.e., accuracy) because a thinking cost (i.e., workload) will continue to increase, which is consistent with the resource rationality framework.

We point out future directions of this study. First, "thinking costs" should be more clarified, such as annoyance, cognitive conflicts, etc. Participants' cognitive load during choice behaviors can be investigated through, for example, mouse tracking approaches (e.g., deviation of mouse trajectories from an ideal path can be an index for conflicts; Stillman, et al., 2018). By focusing on such behavioral indices, we may be able to empirically discuss some aspects of the theoretical findings which are not predicted by the simulation (e.g., regarding the decrease in accuracy with additional time). Second, we should clarify the reasons why the accuracy at 2.5-4.0s was higher for long thinking group than the normal group, even though they used same amount of time to think. We now do not have clear evidence for this issue, but we speculate the differences in cognitive load between considering and waiting. For the first 2.5s in a trial, participants in the normal group constantly considered the task without a waiting time, and larger cognitive costs might be induced. On the other hand, participants in the long thinking group did not have to deeply consider the task because of a waiting time, thus their cognitive costs might be smaller. In any case, this fact may be a key to shed light of a more complex dynamics of thinking time and cognitive cost than initially proposed.

Regarding practical implications, our results can be a key to considering interventions that improve individuals' judgment accuracy. As shown in our simulations and experiments, the total benefit should peak at an early stage. Thus, if an appropriate time (i.e., not too short but not too long, such as approximately 1s) is inserted at the beginning of a trial, individuals' accuracy may be maximized with a minimum workload. According to the resource-rational analyses (e.g., Griffiths et al., 2015; Lieder et al., 2018), the peak of total benefit can be regarded as the point at which individuals achieve the optimal allocation of their cognitive resources. Designing an intervention to enhance accurate judgments under limited cognitive resources will be strongly related to a *boost*, which aims at behavioral changes by fostering people's cognitive competence (e.g., Hertwig & Grüne-Yanoff, 2017). Taken together, this study emphasizes the importance of considering limited cognitive resources (i.e., thinking not only about benefits but also about costs) in the context of SAT.

References

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological Review*, *113*(4), 700–765.

Bogacz, R., Wagenmakers, E. J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences*, *33*(1), 10–16.

Chittka, L., Skorupski, P., & Raine, N. E. (2009). Speed-accuracy tradeoffs in animal decision making. *Trends in Ecology and Evolution*, *24*(7), 400–407.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.

Frazier, P., & Yu, A. J. (2008). Sequential hypothesis testing under stochastic deadlines. *Advances in neural information processing systems*, *20*.

Griffiths, T. L., Lieder, F., & Goodman, N. D. (2015). Rational use of cognitive resources: Levels of analysis between the computational and the algorithmic. *Topics in Cognitive Science*, *7*(2), 217–229.

Hastings, W.K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, 57, 97-109.

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*(150).

Heitz, R. P., & Schall, J. D. (2012). Neural mechanisms of speed-accuracy tradeoff. *Neuron*, *76*(3), 616–628.

Henmon, V. A. C. (1911). The relation of the time of a judgement to its accuracy. *Psychological Review*, *18*(3), 186–201.

Hertwig, R., & Grüne-Yanoff, T. (2017). Nudging and boosting: Steering or empowering good decisions. *Perspectives on Psychological Science*, *12*(6), 973–986.

Ho, M. K., Cohen, J. D., & Griffiths, T. L. (2023). Rational simplification and rigidity in human planning. *Psychological Science*, *34*(11), 1281–1292.

Karşılar, H., Simen, P., Papadakis, S., & Balci, F. (2014). Speed accuracy trade-off under response deadlines. *Frontiers in Neuroscience*, *8*(248).

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*(e1).

Lieder, F., Griffiths, T. L., Quentin, Q. J., & Goodman, N. D. (2018). The anchoring bias reflects rational use of cognitive resources. *Psychonomic Bulletin and Review*, *25*(1), 322–349.

McElree, B., & Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(6), 1517–1539.

Öztekin, I., & McElree, B. (2010). Relationship between measures of working memory capacity and the time course of short-term memory retrieval and interference resolution. *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(2), 383–397.

Prystawski, B., Mohnert, F., Toši´c, M. T., & Lieder, F. (2021). Resource-rational models of human goal pursuit. *Topics in Cognitive Science*, Advance online publication.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*(2), 333–367.

Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences*, *20*(4), 260–281.

Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science*, *181*(4099), 574–576.

Shugan, S. M. (1980). The cost of thinking. *Journal of Consumer Research*, *7*(2), 99–111.

Stillman, P. E., Shen, X., & Ferguson, M. J. (2018). How mouse-tracking can advance social cognitive theory. *Trends in Cognitive Sciences*, *22*(6), 531–543.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67–85.