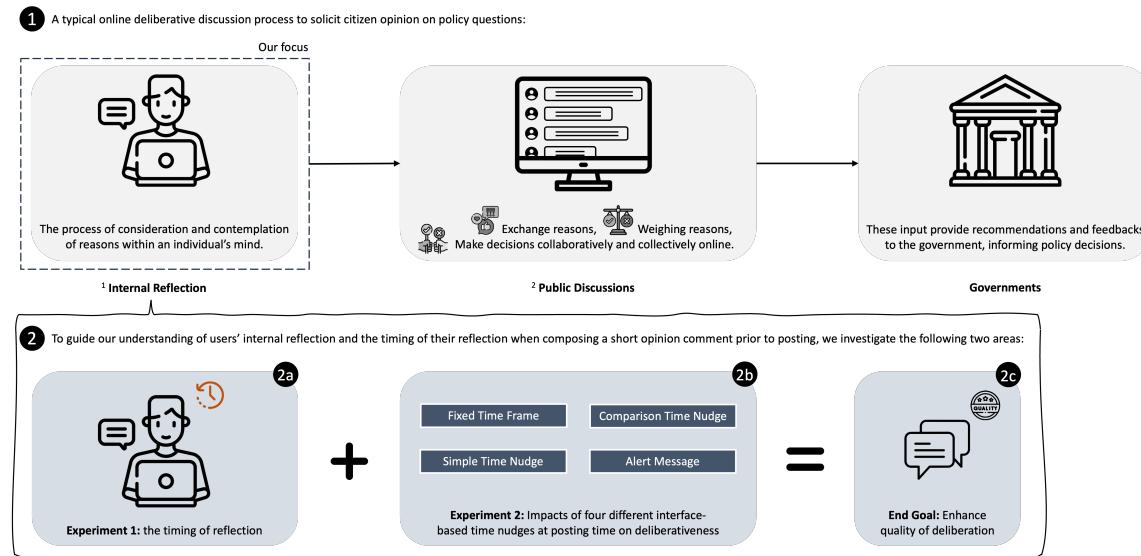


Not Too Long, Not Too Short: Goldilocks Principle of ‘Optimal’ Reflection Time on Online Deliberation Platforms

SHUNYI YEO, Singapore University of Technology and Design, Singapore

SIMON TANGI PERRAULT, Singapore University of Technology and Design, Singapore



¹Certain studies use the term “personal deliberation” to describe the concept of “internal reflection.” In this paper, we adopt the latter term.

²Certain studies utilize the terms “public deliberation”, “online deliberation”, “online discussions” or “public discussions”. In our work, we adopt the term “public discussions”.

Fig. 1. 1: A typical online deliberative discussion process involves individuals deliberating independently before collectively discussing and deciding on issues, often providing inputs to the government. As such, the quality of ensuing discussions downstream is intricately linked to the quality of internal reflection upstream; 2: In conventional online deliberation platforms, users share their opinions on a post. To enhance the public discussion quality downstream, we explore the relationship between internal reflection and reflection timing when crafting short opinion comments in minute-scale deliberations. Our investigation focuses on two primary areas: 2a: the influence of different reflection time lengths on deliberative quality and 2b: the effectiveness of four distinct interface-based time nudges: Fixed Time Frame, Simple Time Nudge, Comparison Time Nudge and Alert Message - in supporting the implementation of reflection time; 2c: Ultimately, we seek to augment the quality of deliberation on online deliberation platforms.

The deliberative potential of online platforms has been widely examined but the **impact of reflection time on the quality of deliberation** remains under-explored. This paper presents two user studies involving 100 and 72 participants respectively, to investigate the impact of reflection time on the quality of deliberation in **minute-scale deliberations**. In the first study, we identified an *optimal* reflection time for composing short opinion comments. In the second study, we introduced four distinct interface-based time nudges aimed at encouraging reflection near the *optimal* time. While these nudges may not improve the quality of deliberation,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

they effectively prolonged reflection periods. Additionally, we observed mixed effects on users' experience, influenced by the nature of the time nudges. Our findings suggest that reflection time is crucial, particularly for users who typically deliberate below the *optimal* reflection threshold.

CCS Concepts: • Human-centered computing → Empirical studies in HCI.

Additional Key Words and Phrases: deliberation, internal reflection, reflection time, time nudges, deliberative quality, deliberativeness, online deliberation platforms, public discussions, nudges, reflection, civic engagement

ACM Reference Format:

ShunYi Yeo and Simon Tangi Perrault. 2024. *Not Too Long, Not Too Short: Goldilocks Principle of 'Optimal' Reflection Time on Online Deliberation Platforms*. In *OzCHI 2024, November 30–December 4, 2024, Brisbane, Australia*. ACM, New York, NY, USA, 38 pages. <https://doi.org/XXXXXX.XXXXXXXX>

1 INTRODUCTION

Advancements in information and communication technologies have enabled the emergence of online forums dedicated to deliberative democracy [7]. Encompassing various formats (e.g., online discourse and reviews, discussion forums like Reddit and social media like Facebook) [25], these forums, sometimes called online deliberation platforms [7, 25], serve as virtual democratic public spheres that foster political participation [56]. On these platforms, large and heterogeneous groups of people contribute their knowledge and opinions [34, 52]. Asynchronous exchanges facilitated through discussion threads further empower citizens' participation to collaboratively engage in thoughtful discussions on societal matters [46, 52] and informing policy decisions [28, 40]. This process is commonly known as online deliberative discussions [25, 83] (see Figure 1-1). In this way, online deliberation platforms facilitate the process by enabling well-informed and reasoned discussions about public-interest issues, with the potential involvement of all stakeholders [106].

In some philosophical discussions of deliberation, it is assumed that informed, rational and open-minded discussions would naturally occur on these online deliberation platforms without further assistance [32, 46, 59]. However, many of these platforms are not designed with deliberation in mind [38, 67], lacking in features that support high quality discussions [26]. It is thus not uncommon to see posts that are incendiary or irrelevant [97], as well as contributions lacking substantial information [104]. Such common occurrences disrupt the internal reflection process and undermine the quality of subsequent public discussions, leading to chaotic and disjointed discourse [9, 75, 97]. Fundamentally, the essence of public discussions is intricately tied to the process of internal reflection, wherein individuals formulate reflective opinions that contribute to the deliberative process [48, 107]. This is further reinforced by Goodin et al. [41, 42] who postulated that internal reflection is fundamental in the deliberation process: ‘... *deliberation of the more internal-reflective sort precede formal group deliberations of the more discursive sort*’. Recognizing the primacy of internal reflection [10, 32, 42], scholarly attention has emphasized that fostering reflection is crucial for deliberative democracy to flourish [17, 29, 41, 74, 101]. Central to this discussion is the recognition that providing adequate reflection time is essential to enhance the quality of deliberation [36, 63, 70]. As Moon succinctly argues [70]: ‘*reflection takes time; therefore creating or allowing time for reflection is essential*’. Yet, the examination of reflection time materialized in interface design remains largely unexplored.

Motivated by Moon's insights, we thus seek to explore two aspects of reflection time on the quality of deliberation within the internal reflection process: (1) how long it takes and (2) how it can be supported on online deliberation platforms (see Figure 1-2). Our examination is facilitated by designing and implementing a simple interface design intended to encourage reflection time at minute-scale deliberations. To support our examination, we pose the following

research questions: **RQ1: What is the impact of reflection time on the quality of deliberation?** and **RQ2: How do different interface-based time nudges affect the quality of deliberation?** We conducted two user studies, each addresses the aforementioned questions respectively.

To answer RQ1, we conducted our first study with 100 participants, where participants deliberate on an issue at different reflection time intervals. Study results revealed that there exists an *optimal* reflection time for composing short opinion comments in minute-scale deliberations and that increasing reflection time is most favourable for users who deliberate below the optimal time. To answer RQ2, we conducted our second study with 72 participants to assess the effectiveness of different time nudges on the quality of deliberation. We designed and implemented four interface-based time nudges (i.e., Fixed Time Frame, Simple Time Nudge, Comparison Time Nudge and Alert Message), to facilitate the integration of nudges to existing online deliberation platforms. Study results revealed that all four time nudges effectively extended reflection time but they did not differentiate in their impact on deliberative quality. Moreover, the time nudges exhibited distinct traits, leading to varying levels of user experience when they craft their opinions online. We discuss later how these results provide important implications in the design of reflection time on online deliberation platforms.

The contribution of this work is twofold:

- The results of two studies investigating how the quality of deliberation is affected by the time length for reflection and four distinct interface-based time nudges in minute-scale deliberations.
- A set of design guidelines detailing the integration of the dimension of reflection time and the application of these interface-based time nudges in online deliberation platforms and in other similar contexts.

2 BACKGROUND AND RELATED WORK

We first begin with an overview of our main focus: reflection time within the deliberation process. Subsequently, we discuss its assessment methods and recent works on nudging to enhance deliberativeness.

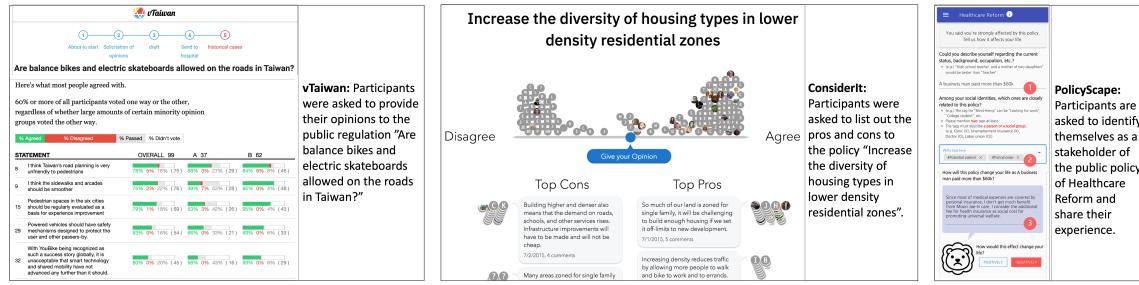


Fig. 2. Examples of existing online deliberation platforms (from left): **vTaiwan**¹ is a platform for opinion exchange to produce public regulations that align with the expectations and needs of stakeholders. The next two examples shows an illustration of online deliberation platforms that incorporate reflection approaches: **ConsiderIt**² is a platform to support users' reflection process by guiding them to reflect on trade-offs of policies through the creation of pros and cons points; **PolicyScape**³ is an online system designed to support users' reflection on public policies by helping them to explore diverse stakeholders' perspectives.

2.1 The Role of Reflection Time in Deliberation

Deliberation denotes the process of thoughtful consideration by an *individual* [6, 25, 42, 94]. It involves carefully weighing the reasons for and against a given measure [5, 25, 42, 94], empowering *individuals* to make informed and

rational choices [20, 107]. Deliberation theorists emphasize reflection as a fundamental element with its significance well-supported within deliberation studies [17, 29, 41, 74, 84]. As articulated by Goodin and Niemeyer [42]: ‘internal reflection inherently precedes public discussion’, laying the cognitive groundwork that shapes an individual’s stance and engagement in subsequent public discussions [17, 29, 41, 42, 74].

Research into enhancing deliberation quality through reflection has been widely examined on online deliberation platforms [3, 29, 58, 60, 78, 101, 107] (see Figure 2). However, ensuring effective reflection necessitates establishing appropriate conditions. Scholars [36, 70] postulate that reflection *takes time*, emphasizing that ‘creating or allowing time for reflection is essential’. Addressing this, designing for interactive systems should mitigate time as a barrier to reflection [63], acculturating users to take some time for attention and reflection before commenting [32]. Collectively, these studies underscore the significance of allocating time for reflection, supporting a shared conclusion: reflection time possess the potential to influence deliberativeness, a normatively desired effect for deliberation. We thus propose **H1: More reflection time increases deliberativeness.**

While reflection time is crucial for reflection, how reflection time influences deliberativeness remains unexplored even within the relevant CSCW and HCI fields of deliberation studies. For this study, we thus focus on examining the role of reflection time, particularly in **how long it takes** and **it how it can be supported** in the context of minute-scale deliberations, where users craft short opinion comments on online deliberation platforms. Given the prevalence of short-form communication on online deliberation platforms [92, 96, 102], which fosters rapid interactions and quick information dissemination [53, 103], focusing on minute-scale deliberations capture this fast-paced nature where users often respond swiftly to posts or comments of online discourse [53]. Understanding how reflection operates within these short time frames is crucial for enhancing the quality of deliberation in such contexts.

2.2 Measuring Deliberation: Deliberativeness

Deliberativeness denotes the quality of deliberation made by an individual [95], a fundamental aspect of public discussions [69]. In our study, we adopt the term ‘deliberativeness’ to characterize the quality of an individual’s opinion [10, 44, 78, 88, 89, 107] on online deliberation platforms.

Deliberativeness is a composite measure with multiple dimensions [95, 107]. Spatariu et al. [87] identified four methods to analyze deliberativeness, with three pertaining to public discussions among groups of people. The fourth method, argument structure analysis, applies even to individuals’ opinions. As we focus on internal reflection, we assess written opinions, emphasizing argumentation as our measure of deliberativeness and adopting methodologies from the literature that cover this aspect: Trénel [95] proposed a coding scheme for deliberativeness with eight measurements that point to the quantity, duality and diversity of opinions. Cappella et al. [13] propose *argument repertoire* which counts the number of non-redundant arguments for and against an issue, a common measure used in deliberative research [13, 57, 69, 101, 107]. Menon et al. [69] further include *response word count* as they find it to be an ‘overall interesting indicator of the evolution of a conversation on online platforms’. Other papers also emphasize the significance of *argument diversity* [2, 39, 80, 101], which measures the diversity of perspectives in an opinion, accounting for varied interpretations and viewpoints of an issue.

For this study, we follow the practice of prior work, assessing deliberativeness with three measures: argument repertoire, argument diversity and word count.

¹vTaiwan

²ConsiderIt. Access the complete paper at [60].

³Access the complete paper at [58].

2.3 Nudging to Augment Deliberativeness

Nudges involving simple interface changes have been shown to effectively change users’ behaviour [93]. Nudges seek to influence choices and decision process by effectively intervening and moderating an individual’s usual behaviour, redirecting to shift from their automatic mode of thinking towards the value introduced by the nudge [93]. The idea of nudge was eagerly adopted in HCI [14, 49, 77, 98]. Caraban et al. [14] presented a systematic review of the use of nudging in HCI research, of which they found 23 distinct mechanisms of nudging that can be grouped into 6 categories. Their work discussed the factors shaping the nudge’s effectiveness.

Within the realm of deliberation, there has been some interest in using simple interface designs to enhance the deliberativeness of online discourse [100, 105]. For instance, Menon et al.[69] demonstrated that interface nudges such as partitioning text fields, could lead to a significant increase in response length by up to 35% and the generation of more arguments by up to 25%. Likewise, Yeo et al. [101] found a significant positive impact of textual based reflective nudges on deliberative quality and how different reflective nudges influence the dynamics of online deliberations. Additionally, Zhang et al.[107] investigated the use of reflection nudges through simple question prompts on users’ perceived issue knowledge and their expression of attitudes and opinions. Their findings showed that prompting participants with questions like ‘What are your opinions on this issue?’ enhanced the quality of participants’ opinions.

Drawing from existing efforts, we utilize the concept of nudges in the dimension of time. Our goal is to enhance deliberativeness through various interface-based time nudges. We anticipate that these time nudges will enhance deliberativeness.

3 STUDY 1: REFLECTION TIME FOR DELIBERATION

To address **RQ1** and validate **H1**, we conducted an experimental study with 100 participants to investigate the most appropriate reflection time needed to achieve the highest deliberativeness.

3.1 Independent Variable and Experimental Design

A between-subject experiment was conducted with the independent variable being Reflection Time with five levels that are modeled at one minute intervals: {One minute, Two minutes, Three minutes, Four minutes, Five minutes}. Since no prior in-depth investigation on reflection times exists (prior research that comes closest is on investigating the optimal delay time for pre-moderated content on public displays [45]) and our focus on minute-scale deliberation, we chose concrete values to best capture the impact on deliberativeness.

Reflection time was modeled following Alós-Ferrer and Buckenmaier [1], as the sum of the times resulting from a chain of steps of reasoning to arrive at an observed response. Thus, reflection time is the total time taken by an iterative process of reading, reflecting, typing and reviewing. As it is also difficult to disentangle the different steps in reflection, using this framework allows us to capture the entire journey from task onset to submission.

The interface design, inspired by literature on hyperbolic time discounting and dual process theory, seek to encourage users to pause and reflect during the specified reflection periods. Following Wang et al.’s design considerations for timer nudge [98], our interface includes three key features, depicted in Figure 3.

- *Delay Submit Button.* *Delay Submit Button* hides the ‘Submit’ button until a certain reflection time is reached, inspired by the nudging mechanism of hiding [14]. This ensures that participants stay on the task page to reflect for the duration specified in their condition. The ‘Submit’ button becomes visible only 10 seconds before the timer concludes.

The figure shows two screenshots of a survey interface. On the left, a timer at the top left corner displays '0300'. Below it is a text input area with placeholder text about a 4-day work week, a word count of 0, and a question about position. A row of radio buttons for a Likert scale follows. On the right, a similar interface is shown with a timer at the top left corner displaying '0010'. Below it is the same text input area, question, and Likert scale. A large black arrow points from the left interface to the right one. At the bottom right of the right interface is a blue 'Next' button.

Fig. 3. Key features of the interface design in study 1. Left: The timer is positioned on the top left corner and there is no submit button for participants to proceed even if they have finished typing their responses before the timer ends. Right: When the timer reaches 10 seconds, the submit button appears. When the timer ends, participants' responses will be automatically submitted. This figure shows a screenshot of the Three minutes experimental condition but is the same for all five reflection time lengths used in study 1.

- *Timer*. To give participants a visual delay of time passing, a count down *Timer* is displayed at the top left corner of the interface. The timer would start at 60, 120, 180, 240 and 300 seconds for the five experimental conditions respectively.
- *Auto Submission*. To constraint participants to only reflect up to the specific time interval, the *Auto Submission* feature would kick in when the timer reaches 0 and automatically submit the participant's response.

All three features will be automatically reset if the participant exits and reopens the survey link, or navigate to a different page and then return to the same page.

3.2 Dependent Variables

As deliberativeness is multi-dimensional, we operationalized it through three measurements: argument repertoire, argument diversity and word count as discussed in section 2.2.

Argument repertoire and argument diversity were derived from a content analysis of participants' responses by two coders (see Appendix section 9). Both coders were PhD students with respectively two and five years of experience using content analysis. Cohen's Kappa was used to determine the agreement between the two coders' judgments: Argument Repertoire ($\kappa = 0.808$) and Argument Diversity ($\kappa = 0.808$). Kappa scores for both metrics were above the satisfactory threshold of 0.80 [37, 61, 72].

The dependent variables are coded as follows:

- *Argument repertoire* is the number of non-redundant arguments regarding each position of the discussion topic. Ideas produced along the two positions were combined.

- *Argument diversity* was coded by counting the number of unique themes present in the entire response. A higher diversity count indicates more varied perspectives present in the participant’s responses [2, 39].
- *Word count* is a simple count of the number of words in the participants’ opinion to the discussion topic.

3.3 Power Analysis

We conducted a power calculation for a five-group ANOVA study seeking a medium effect size according to Cohen’s conventions [21], at 0.90 observed power with an alpha of 0.05, giving $N = 20$ per experimental condition, hence we recruited 100 participants.

3.4 Participants and Ethics

A total of 100 participants was recruited through Amazon Mechanical Turk (MTurk). Refer to Appendix Table 9 on the breakdown of the demographic profile in each experimental condition. We ensured that the demographic profiles across the five conditions were similar so as to control for any fixed effects resulting from the differences in demographic factors. We obtained ethics approval from our local Institutional Review Board (IRB) and reimbursed participants at an hourly rate of 7.25 USD, as per IRB guidelines.

3.5 Task and Material

For the task, we identified a discussion topic on four-day workweek. Participants had to write at least 30 words when expressing their take on the topic⁴. We selected this topic as it is readily accessible and politically relevant⁵, thereby promoting constructive and open debate. We chose to provide minimal deliberation materials, consisting mainly of the topic background to align with the context of minute-scale deliberations.

3.6 Procedure

The study contained three parts with Qualtrics and MTurk being used for data collection.

3.6.1 Pre-Task: Consent, Instructions manual. MTurk workers who met the worker’s requirements for our task⁶ were invited to complete our study. Specifically, as language fluency could affect both reading and response times, we set strict eligibility criteria by requiring all participants to be fluent in English. Participation consent was obtained before the study. Participants were informed that they had to express their opinions on a discussion topic, which was undisclosed at this point, under a specified reflection time frame. They then proceeded to read the instruction manual which outlined the key features of the interface. Afterwards, participants saw a screen informing them that the timer would start immediately after.

3.6.2 Main Task. Detailed in section 3.5.

⁴We conducted a pilot of this task with 22 participants on MTurk and found that a simple text box often led to responses with low informational content, such as ‘Four-day workweek is a good thing in my view - P4, Pilot’, which made it difficult to compare the results of the experimental conditions. Following lessons learnt from other studies [4, 69, 91], we decided to add a minimal word count requirement that seek to improve the quality of responses.

⁵The topic on four-day workweek has garnered attention in various countries, including governments and in corporate discussions: [The New York Times](#) and [Straits Times](#).

⁶To be eligible for participation, participants had to be: (1) 18 years old and above, (2) fluent in English, with (3) a HIT (i.e., human intelligence task) approval rating above 98% and (4) the number of HITs approved greater than 10,000. The fairly stringent cutoffs are to ensure that participating workers have consistently delivered reliable task quality. Qualifiers were also assigned to ensure unique worker participation in the experiment.

3.6.3 Post-Task: Post-Task Feedback, Familiarity, Demographics. After completing the main task, participants completed a short post-task survey on the duration of the reflection time. This was assessed on a five-point Likert scale (1 = Too Short, 2 = A Little Short, 3 = Just Right, 4 = A Little Long and 5 = Too Long).

To control for any fixed effects, participants were asked to rate their familiarity on the discussion topic from a scale of 0 to 100 with 0 being not familiar at all and 100 being very familiar. Demographic information (age, gender, country) was also collected.

3.7 Outliers

Before analyzing our data, we plotted a box plot (box and whisker plot) to visually show the dispersion of our data and to identify any potential outliers [85]. We adopted the approach by Dawson [27] to discard data points that were either below $Q1 - 3IQR$ or above $Q3 + 3IQR$ (where $Q1$ is the first quartile, $Q3$ the third, and IQR is the interquartile range). This led us to discard two data points from the Four minute condition. Hence, our final sample size was 98.

4 FINDINGS (STUDY 1)

4.1 Results

Unless otherwise specified, ANOVA was used to identify main effects and pairwise t-tests with Bonferroni correction for post-hoc comparisons [76] for the three measurements of deliberativeness.

4.1.1 Word Count. We found a statistically significant main effect of *Reflection Time* on Word Count ($F_{4,93} = 8.21$, $p < .0001$). There was a statistical difference between One minute ($M = 31.45$ words) and Three ($M = 71$ words), Four ($M = 73.78$ words) and Five minutes ($M = 76.55$ words) (all $p < .01$). We also found a significant difference between Two minutes ($M = 47.15$ words) and Five minutes ($p = .032$). Results are summarized in Figure 4a.

4.1.2 Argument Repertoire. We found a significant main effect of *Reflection Time* on Argument Repertoire ($F_{4,93} = 4.99$, $p < .001$). There was a statistical difference between One minute and Three, Four and Five minutes such that One minute ($M = 1.85$ arguments) has lower number of arguments compared to Three ($M = 3.3$ arguments, $p < .05$), Four ($M = 3.39$ arguments, $p < .01$) and Five minutes ($M = 3.35$ arguments, $p < .05$). We did not find any other pairwise differences. Results are summarized in Figure 4b.

4.1.3 Argument Diversity. We found a statistically significant main effect of *Reflection Time* on Argument Diversity ($F_{4,93} = 6.16$, $p < .001$). Similarly, we found significant differences between One minute ($M = 1.65$ diversity) and Three ($M = 2.8$ diversity, $p < .05$), Four ($M = 3.44$ diversity, $p < .01$) and Five minutes ($M = 3.25$ diversity, $p < .01$). Results are summarized in Figure 4c.

4.1.4 Summary of Results. Overall, deliberative quality by the three measurements is summarized in Table 1. A careful look at the data reveals that Four minutes achieves the highest overall deliberativeness in the lowest amount of time among all the other reflection times. It is also worth mentioning that the typical reflection time⁷ participants took fell short of Four minutes.

⁷Before conducting study 1, we measured participants' ($N = 34$) typical reflection time on the four-day workweek task without any experimental influence. The mean time was 3.07 minutes, with a median of 2.71 minutes (range = 42.3 seconds to 6.14 minutes, $SD = 1.73$ minutes, $Q1 = 1.63$ minutes, $Q3 = 4.86$ minutes). We then conducted study 1 to find the most appropriate reflection time and if the mean typical reflection time aligns with this appropriate reflection time. It is noteworthy to note that the mean typical reflection time is below the 'optimal' reflection time of Four minutes.

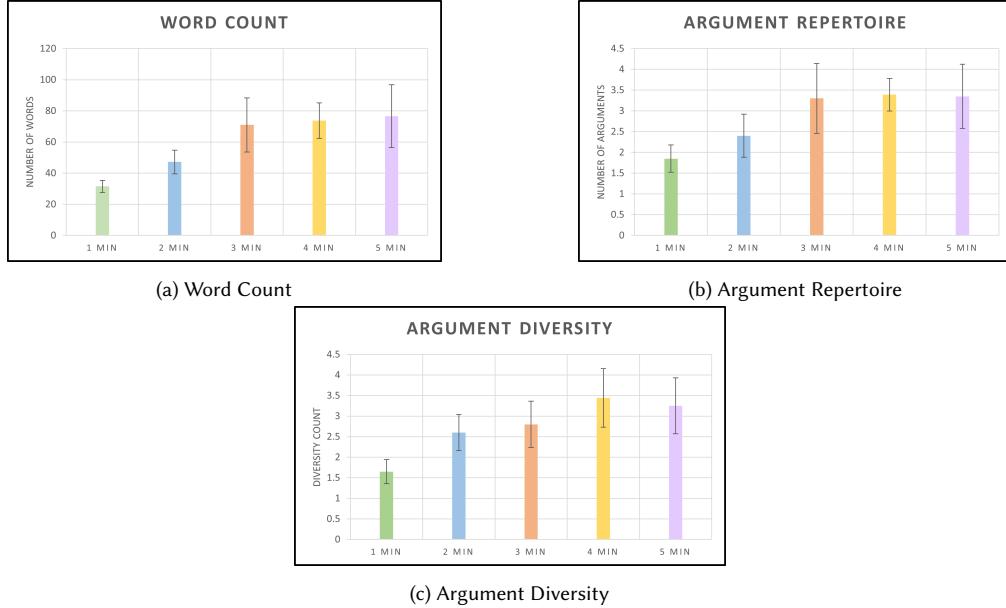


Fig. 4. Word count, argument repertoire and argument diversity for each reflection time. Error bars show 95% confidence intervals.

Table 1. Deliberative quality across the different reflection times. α , β , γ and μ show significant pairwise comparisons ($p < .05$).

Deliberativeness	Reflection Time				
	1 min ($M \pm S.D.$)	2 mins ($M \pm S.D.$)	3 mins ($M \pm S.D.$)	4 mins ($M \pm S.D.$)	5 mins ($M \pm S.D.$)
Word Count	31.45 ± 8.83 ^{α,β,γ}	47.15 ± 17.16 ^μ	71.00 ± 39.71 ^α	73.78 ± 24.60 ^β	76.55 ± 45.99 ^{γ,μ}
Argument Repertoire	1.85 ± 0.75 ^{α,β,γ}	2.40 ± 1.19	3.30 ± 1.92 ^α	3.39 ± 0.85 ^β	3.35 ± 1.76 ^γ
Argument Diversity	1.65 ± 0.67 ^{α,β,γ}	2.60 ± 0.99	2.80 ± 1.28 ^α	3.44 ± 1.54 ^β	3.25 ± 1.55 ^γ

4.1.5 Relationship between Reflection Time and Deliberativeness. Based on our results, we plotted three scatter plots to examine the relationship between reflection time and deliberativeness.

To determine which regression model (i.e., linear or logarithmic) provides a reasonable fit, we plotted both models in all three scatter plots (see Figure 5). We then ran regression analyses for both functional forms (see Appendix Table 11). The logarithmic regression model provided a better fit across all three scatter plots, with an increase of R^2 for Word Count (0.872 to 0.973), Argument Repertoire (0.782 to 0.980) and Argument Diversity (0.798 to 0.984). The main reason for preferring the logarithm model is that it makes more sense theoretically as we explain now.

Plateauing Effect. Firstly, by examining the data from all three scatter plots (see Figure 5), word count, argument repertoire and argument diversity increases with reflection time but plateau after three minutes. This hinted that the relationship between reflection time and deliberative quality might be nonlinear.

Intercept. Secondly, it does not make sense to have words or arguments present at time 0 minutes ($t = 0$). For instance, at $t = 0$, average word count is 10.52 as in the linear model (see Figure 5a), which is practically impossible. Hence, a logarithmic model may make more sense here as it accounts for the time lag participants took before responding.

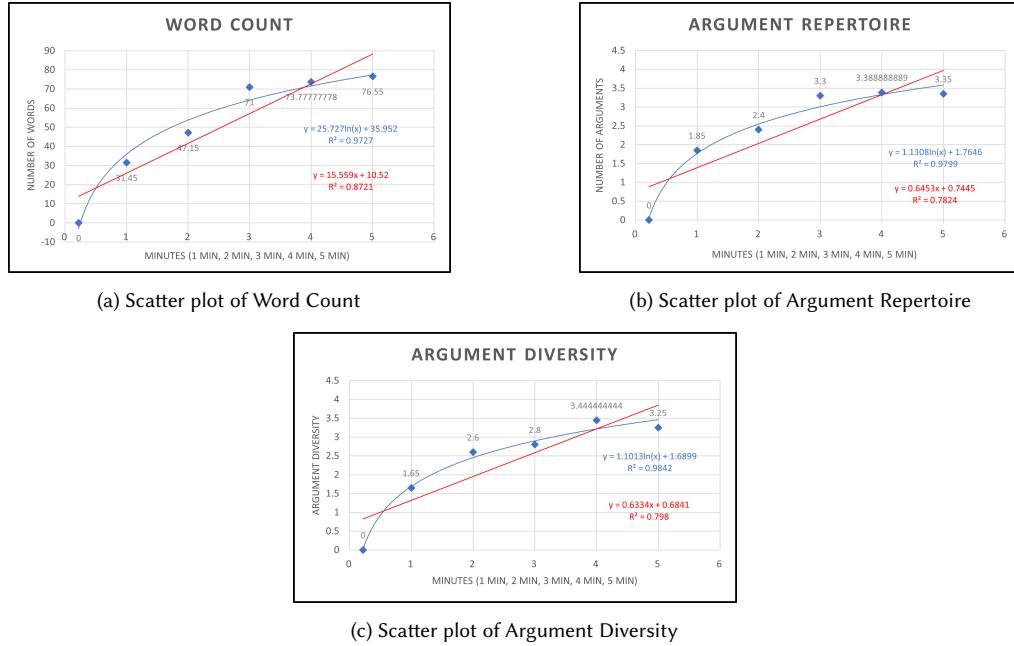


Fig. 5. Logarithm function (blue) vs linear function (red) for the three measurements of deliberativeness.

All three scatter plots for the logarithmic function (see Figure 5) showed that word count, argument repertoire and argument diversity remained at 0 until reflection time reaches 0.22 minutes (approximately 13.2 seconds). This reflects the average time participants took to respond to the task, aligning with our recorded data.

Summary. From this logarithmic law, and by observing our data, we can conclude that **increasing reflection time exhibit diminishing returns in terms of the increase for deliberative quality**.

4.1.6 Post-Task Feedback on the Duration for Reflection. One minute was widely considered inadequate, with 80% of participants indicating it was *Too Short* (40%) or *A little Short* (40%). At Two minutes, the proportion finding it too short decreased to 50% (25% *Too Short*, 25% *A little Short*), while 35% considered it *Just Right*. At Three minutes, none found it *Too Short*, though 40% still felt it was *A little Short*, and 40% thought it was *Just Right*. At Four minutes, 17% found it *A little Short*, and 61% considered it *Just Right*. At Five minutes, 60% of participants found it long (20% *Too Long*, 40% *A little Long*), while 10% found it *Just Right* - a significant drop from Four minutes. Surprisingly, 30% felt that Five minutes was *A little Short*, up from 17% at Four minutes. The results are summarized in Figure 6.

4.2 Discussion

In this section, we discussed our findings to answer **RQ1** and to validate **H1**. We also discussed how our results corroborate and enrich previous studies.

4.2.1 Logarithmic Relationship between Reflection Time and Deliberativeness. Our findings indicate a significant positive effect of reflection time on deliberativeness. This is inline with Alós-Ferrer and Buckenmaier [1] work, which showed that higher observed depth of reasoning is linked to longer reflection times. Our results would also mean that the

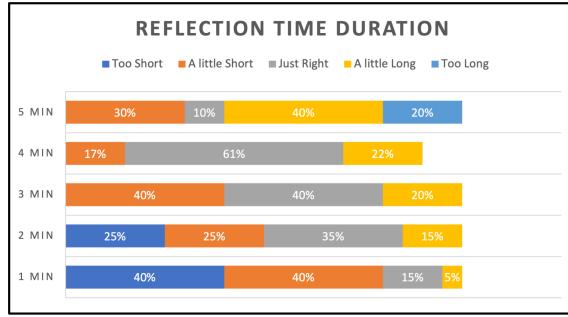


Fig. 6. Post-task feedback on the length of reflection time in each experimental condition. Note: the Four Minutes condition is missing two discarded data points but the total still amounts to 100%.

element of time in reflection is important in the deliberation process. This aligns with prior research that allowing for time in reflection is crucial and thus should be accounted for [33, 36, 63, 64, 70].

However, our study goes beyond to uncover the impact of increasing reflection time. We found that increasing reflection time does not necessarily increase deliberative quality indefinitely. There seems to exists a *limit* to how much time can enhance deliberativeness. Our results showed that all three measurements of deliberativeness obey the logarithmic law, indicating a non-linear relationship between reflection time and deliberativeness. Specifically, all three measurements provide converging evidence that as reflection time increases beyond a certain limit, deliberative quality plateaus or even drops. This suggests that increasing reflection time beyond the limit diminishes the rate of return for deliberativeness in minute-scale deliberations. Thus, **H1 is partially supported**.

4.2.2 Goldilocks Principle of ‘Optimal’ Time. The logarithmic relationship suggests that there exist a possibility of an ‘optimal time’ for reflection in minute-scale deliberations. The optimal time is the *limit* to how time can improve deliberative quality. It can also be interpreted as the point where deliberative quality is maximized most efficiently. From study 1, we deduce that the optimal reflection time is *Four minutes* because it has the best deliberative quality across the five experimental conditions in terms of argument repertoire and argument diversity. Moreover, participants’ subjective feedback for *Four minutes* is the most favourable (i.e., *Four minutes* garnered the highest percentage of 61% for *Just Right*).

The existence of an optimal reflection time in deliberation has a twofold effect when designing for minute-scale deliberations. First is the case of spending too little time which is below the optimal duration, leading to lower deliberative quality. In this case, increasing reflection time arbitrarily close to the optimum, may significantly enhance deliberativeness, signifying that designers should work towards enabling users to reflect at the optimal time. Second is the case of spending too much time that is well beyond the optimum. In this case, it may not necessarily improve deliberativeness any further and might even introduce trade-offs. We define this phenomenon as the Goldilocks Principle⁸ of ‘Optimal’ Reflection Time, too short is unsatisfactory but too long can be undesirable.

4.2.3 Below Optimal Time: Time-Induced Reflection. In the first case, we coined a term ‘time-induced reflection’ - a way of using time to induce participants to reflect more whenever they reflect below the optimal time. Study 1 illustrated that reflecting below the optimal time of Four minutes (e.g., One and Two minutes) results in lower deliberativeness.

⁸The Goldilocks principle is named after a children’s story of “The Three Bears”. It is used to illustrate the concept of having “just the right amount”.

Consequently, extending reflection time nearer to the optimum (e.g., Three - Four minutes) significantly enhance deliberativeness. Therefore, designers should aim not only to provide time for reflection but, more importantly, work towards encouraging users to reflect arbitrarily close or at the optimum. Time-induced reflection can be easily achieved through nudging techniques [93] such as prompting users to take more time to reflect or review before posting. In this aspect, time-induced reflection acts as an anchor influencing individuals' effort [69, 91]. Most task performers acting alone go about their usual behaviours [50], and coupled with the lack of incentives, they do not push themselves towards producing better quality of work [69]. Introducing an anchor to encourage more reflection shifts their previous performance towards the value introduced by the anchor [16, 69]. This is especially beneficial for two groups of users: (1) impulsive individuals and (2) underthinking individuals. Impulsive individuals often exhibit time-urgent personality, having a predisposition to rushing [79] and completing tasks hastily, all of which tended to correlate more with negative affect [30], resulting in poorer performance [71]. Time-induced reflection can effectively slow them down, compelling them to invest more time to consciously process their thoughts and be more thoughtful in their responses, consequentially moving them towards reflecting at the optimum. This intervention breaks their habitual behaviour and tendency of just progressing through, ultimately producing more reflective responses and enhancing deliberative quality. For underthinking individuals, time-induced reflection signals a cognitive awareness by reinforcing their engagement back to the task. It allows users to revisit and rethink their responses, ultimately enhancing deliberative quality.

4.2.4 Above Optimal Time: Time-Subduced Reflection. On the other hand, increasing reflection time beyond the optimum may not necessarily be beneficial. Our results showed that as time goes beyond the optimum, deliberative quality falls in terms of argument repertoire and argument diversity. We termed this phenomenon 'time-subduced reflection' - the point at which the drawbacks of extending reflection time beyond optimality outweigh the benefits. In other words, while reflection time supports the deliberation process [36, 63, 64, 70, 71], nudging individuals to take more time when they have already done so does not create additional value and meaning to their current thought-process. This could even demoralize and demotivate people, making them more impatient (see section 4.1.6), or overwhelms them, especially so when incentives are lacking. It is also harmful as longer reflection time without introducing new thinking prompts to support users may result in a decision paralysis [15, 71, 73]. This can happen to individuals who are constantly dwelling with the same thoughts, driving them into the same state of thinking. This causes unproductive thinking and may even result in worse performance [71] as individuals exhibit more randomness in their response [82]. This may explain why deliberative quality in the Five minutes condition deteriorates after the optimal time. In some cases, it may even lower their confidence levels [47, 65] and cause unfavourable feelings of self-doubts when participants could not add further dimensions to their thinking process [47]. This might explain why some participants find Five minutes limited (i.e., 30% indicated *A little Short*). All in all, additional reflection induced by increasing time beyond optimality adversely impacts deliberative quality.

Therefore, when designing for time on minute-scale deliberations, designers should be wary of the Goldilocks principle of 'optimal' reflection time.

5 STUDY 2: TIME NUDGES TO ENCOURAGE TIME-INDUCED REFLECTION

From study 1, we identified a Goldilocks principle of 'optimal' reflection time. However, a sizeable number of participants still considered One and Two minutes to be ideal (see Figure 6). Leveraging the concept of nudge (see section 2.3), we design time nudges to encourage users to deliberate longer. Our design considerations largely follows the nudging mechanisms by Caraban et al. [14]. In study 2, we thus address RQ2 by examining the effects of various interface-based

time nudges on deliberativeness and users' perceptions of the nudge to determine which time nudge best promotes time-induced reflection.

5.1 Independent Variable and Experimental Design

A between-subject experiment was conducted with the independent variable being Time Nudges with four levels: {Fixed Time Frame, Simple Time Nudge, Comparison Time Nudge and Alert Message}. The time nudges are described below:

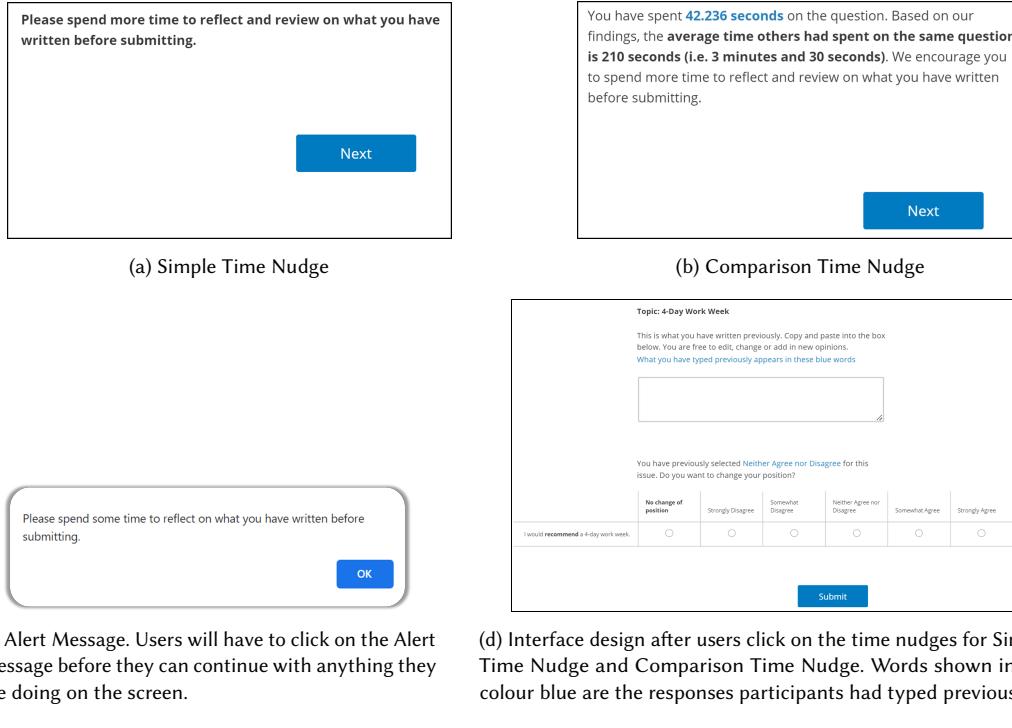


Fig. 7. Interface designs for the Time Nudges excluding Fixed Time Frame which was shown in Figure 3.

5.1.1 Fixed Time Frame. Fixed Time Frame follows the design principles by Wang et al. [98], retaining the same interface design as in study 1 (see Figure 3). We applied the nudge concept to encourage participants to spend a specific reflection time on the discussion topic. This was set to be 3.5 minutes (210 seconds), slightly below the optimal Four minutes (240 seconds) based on feedback from study 1, where 22% of participants found Four minutes a little long (see Figure 6).

5.1.2 Simple Time Nudge. Simple Time Nudge (see Figure 7a) follows the nudging mechanism of reminding [14]. The nudge appears after users have furnished their responses on the discussion topic. It prompts users to pause and review their responses before posting. Users can review their responses, shown in blue, and make edits before submitting their final opinions (see Figure 7d).

5.1.3 Comparison Time Nudge. Comparison Time Nudge follows the nudging mechanism of social influence through social comparisons [14] to nudge users towards a desirable behaviour by increasing self-awareness [31] and tapping

into *herd instinct bias* [14, 22, 35, 43]. The nudge works similar to *Simple Time Nudge* but displays a prompt revealing the user's reflection time compared to the average time spent by others (see Figure 7b). The average time is set at 3.5 minutes, serving as a goal-based anchor [107] to nudge users towards this reflection duration. Like the *Simple Time Nudge*, users can then review and edit their responses before posting (see Figure 7d).

5.1.4 Alert Message. Alert Message follows the nudging mechanism of reinforcement [14, 108]. It is a pop-up nudge that appears after 20 seconds, reminding users to take more time to reflect before submitting their responses (see Figure 7c).

5.2 Dependent Variables

To discern the impact of time nudges on deliberativeness and users' perceptions of the nudges, the study has three dependent variables: deliberativeness, reflection time and users' perceptions (users' experience, thinking process and perceived helpfulness).

5.2.1 Deliberativeness. Deliberativeness was measured in the same manner as in study 1 of word count, argument repertoire and argument diversity (see section 3.2).

5.2.2 Reflection Time. To assess the impact of time nudges on reflection time, we did not restrict reflection duration, except in the *Fixed Time Frame* condition. Participants could and were encouraged to take as much time as needed. Reflection time was measured as the total time users spent reflecting in the presence of the time nudge.

5.2.3 Users' Perceptions. Users' perceptions were measured using subjective measures (i.e., self-reported scores in questionnaires). These subjective measures encompassed eight question items on a 1-5 Likert scale (1 = strongly disagree to 5 = strongly agree), split into three categories: *Users' Experience*, *Thinking Process* and *Perceived Helpfulness* (see Table 2).

- *Users' experience* measures user interaction and engagement with time nudges, adapted from prior studies that studies haste [79].
- *Thinking process* evaluates the perceived impact of time nudges on users' cognitive processes.
- *Perceived helpfulness* assesses the overall perceived utility of time nudges in aiding participants.

5.3 Power Analysis

We conducted a power calculation for a four-group ANOVA study seeking a medium effect size according to Cohen's conventions [21], at 0.80 observed power with an alpha of 0.05, giving $N = 18$ per experimental condition, hence we recruited 72 participants.

5.4 Participants and Ethics

A total of 72 participants were recruited through MTurk. Refer to Appendix Table 10 for the complete demographic profiles. We excluded participants from study 1 as the task was the same. Similar to study 1, we ensured that the demographic profiles across the four conditions were similar to control for any fixed effects resulting from the differences in demographic factors. Likewise, we got ethics approval from our local IRB and reimbursed participants at an hourly rate of 7.25 USD.

Table 2. The eight question items assessing users' perceptions on their experience, thinking process and perceived helpfulness of the time nudges. Participants have to rate each question on a five-point Likert scale (1 = strongly disagree to 5 = strongly agree).

Categories of Users' Perceptions	Question Items
Users' Experience	I felt in a hurry (time pressure) to complete the task.
	The timer/prompt (i.e., asking me to spend more time) is distractive .
Thinking Process	This process has helped me to understand my thinking process better.
	This process has helped me to be more mindful of my own thoughts.
	This process has helped me to be more aware of my own thoughts.
	This process has helped me to be more confident of my own opinions.
Perceived Helpfulness	This process allows me to reflect on my opinions before submitting.
	I believe this process is a good way to encourage people to take more time to review on what they have written before submitting.

5.5 Procedure and Task

The procedure for this study closely follows that of study 1 outlined in section 3.6. In the pre-task, participants went through the instructions for the study.

In the main task, participants type and post a response to the topic of four-day workweek under the influence of the time nudge. We maintain consistency in the topic across the two studies to draw robust comparisons between the different time nudges and their impact on deliberativeness within the same thematic domain [11].

In the post-task, participants completed a short survey comprising of the eight question items as well as to provide feedback on the time nudge they engaged with. Lastly, data on demographic and familiarity of the topic were collected.

6 FINDINGS (STUDY 2)

6.1 Quantitative Results

Similar to study 1, unless otherwise specified, ANOVA was used to identify main effects and pairwise t-tests with Bonferroni correction for post-hoc comparisons [76] for deliberativeness, reflection time and users' perceptions (users' experience, thinking process, perceived helpfulness).

6.1.1 Deliberativeness. No statistically significant effect of *Time Nudges* on word count, argument repertoire and argument diversity were found (all $p > .05$). This implies that there is no significant differences in the deliberative quality among the four time nudges. The results are summarized in Table 3. The values we measured for study 2 are overall aligned with the values observed in study 1 even when taking the reflection time per condition into account.

6.1.2 Reflection Time. We found a significant main effect of *Time Nudges* on Reflection Time ($F_{3,68} = 4.08, p < .01$). There was a statistical difference between *Fixed Time Frame* ($M = 3.5$ minutes) and *Simple Time Nudge* ($M = 5.8$ minutes, $p < .05$) which was explainable as users in *Fixed Time Frame* are bounded by the time limit of 3.5 minutes. We also found a significant difference between *Simple Time Nudge* and *Alert Message* ($M = 3.0$ minutes, $p < .05$). We did not find any other pairwise differences. The results are summarized in Figure 8.

Table 3. Deliberative quality of word count, argument repertoire and argument diversity across the four time nudges.

Deliberativeness	Time Nudges			
	Fixed Time Frame	Simple Time Nudge	Comparison Time Nudge	Alert Message
Word Count	66.2	61.7	66.7	57.1
Argument Repertoire	3.33	3.39	3.00	3.06
Argument Diversity	2.67	2.94	3.39	3.28

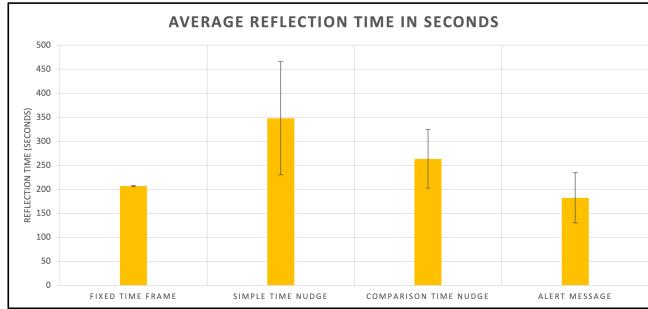


Fig. 8. Reflection Time in seconds for each time nudge. The Fixed Time Frame condition was the only one where participants had a fixed time to reflect. Error bars show 95% confidence intervals.

6.1.3 Users' Experience. Table 4 shows the average scores, main effects and pairwise differences. Overall, participants felt that the *Fixed Time Frame* gave them a significantly higher time pressure compared to the other three time nudges. In terms of distraction, *Comparison Time Nudge* was deemed more distractive compared to *Simple Time Nudge*.

Table 4. Users' Experience. α , β and γ show significant pairwise comparisons ($p < .05$) for each question item.

Users' Experience	$F_{3,68}$	p	Time Nudges			
			Fixed Time Frame ($M \pm S.D.$)	Simple Time Nudge ($M \pm S.D.$)	Comparison Time Nudge ($M \pm S.D.$)	Alert Message ($M \pm S.D.$)
I felt in a hurry (time pressure) to complete the task.	7.44	<.001	$2.78 \pm 1.35^{\alpha,\beta,\gamma}$	$1.56 \pm 0.62^\alpha$	$1.72 \pm 1.27^\beta$	$1.28 \pm 0.57^\gamma$
The prompt (i.e., asking me to spend more time) is distractive .	3.24	<.05	2.28 ± 1.36	$2.22 \pm 1.11^\alpha$	$3.39 \pm 1.14^\alpha$	2.56 ± 1.42

6.1.4 Thinking Process. Results for *Thinking Process* are summarized in Table 5. We found a main effect of *Time Nudges* for every question item. Users perceived *Simple Time Nudge* ($M = 4.22$) to help them better understand their thoughts compared to *Comparison Time Nudge* and *Alert Message*. We observed the same trend for the other three questions where *Simple Time Nudge* scores were significantly higher compared to other scores for *Comparison Time Nudge* (all questions) and *Alert Message* (all questions except **awareness**). In summary, *Simple Time Nudge* is perceived to better support users' thinking process when compared to *Comparison Time Nudge* and *Alert Message*.

6.1.5 Perceived Helpfulness. Results for *Perceived Helpfulness* is summarized in Table 6. Pairwise t-test with Benjamini-Hochberg correction was conducted only for the first question item. We found a significant main effect of *Time Nudges*

Table 5. Thinking Process. α , β and γ show significant pairwise comparisons ($p < .05$) for each question item.

Thinking Process	$F_{3,68}$	p	Time Nudges			
			Fixed Time Frame ($M \pm S.D.$)	Simple Time Nudge ($M \pm S.D.$)	Comparison Time Nudge ($M \pm S.D.$)	Alert Message ($M \pm S.D.$)
Going through this process has helped me to understand my thinking process better.	4.43	<.01	3.78 ± 0.88	$4.22 \pm 0.55^{\alpha,\beta}$	$3.11 \pm 1.41^{\alpha}$	$3.11 \pm 1.32^{\beta}$
Going through this process has helped me to be more mindful of my own thoughts.	3.61	<.05	3.72 ± 0.96	$4.39 \pm 0.61^{\alpha,\beta}$	$3.22 \pm 1.56^{\alpha}$	$3.28 \pm 1.45^{\beta}$
Going through this process has helped me to be more aware of my own thoughts.	3.88	<.05	3.72 ± 0.96	$4.33 \pm 0.59^{\alpha}$	$3.06 \pm 1.47^{\alpha}$	3.28 ± 1.56
Going through this process has helped me to be more confident of my own opinions.	4.69	<.01	3.61 ± 0.85	$4.39 \pm 0.98^{\alpha,\beta}$	$3.28 \pm 1.45^{\alpha}$	$3.00 \pm 1.33^{\beta}$

for both question items. Participants perceive *Simple Time Nudge* ($M = 4.61$) to help them reflect better compared to the other three nudges. Participants also perceived *Simple Time Nudge* to be a better way in encouraging people to take more time to review before submitting when compared to *Fixed Time Frame*.

Table 6. Perceived Helpfulness. α , β and γ show significant pairwise comparisons ($p < .05$) for each question item.

Perceived Helpfulness	$F_{3,68}$	p	Time Nudges			
			Fixed Time Frame ($M \pm S.D.$)	Simple Time Nudge ($M \pm S.D.$)	Comparison Time Nudge ($M \pm S.D.$)	Alert Message ($M \pm S.D.$)
This process allows me to reflect on my opinions before submitting.	2.91	<.05	$3.72 \pm 1.13^{\alpha}$	$4.61 \pm 0.50^{\alpha,\beta,\gamma}$	$3.83 \pm 1.15^{\beta}$	$3.83 \pm 1.15^{\gamma}$
I believe this process is a good way to encourage people to take more time to review on what they have written before submitting.	2.92	<.05	$3.44 \pm 1.04^{\alpha}$	$4.33 \pm 0.59^{\alpha}$	3.72 ± 0.96	3.94 ± 1.06

6.1.6 Summary of Quantitative Results. Overall, there is no significant difference in terms of deliberativeness across the four time nudges, which was explainable because all the reflection times are arbitrary close or beyond the optimal time of Four minutes (see section 6.1.2) - *Fixed Time Frame* ($M = 3.5$ minutes), *Simple Time Nudge* ($M = 5.8$ minutes), *Comparison Time Nudge* ($M = 4.4$ minutes) and *Alert Message* ($M = 3.0$ minutes). This is inline with study 1 results where we observed deliberativeness to have a plateauing effect after the Three minutes mark.

Although there is no significant difference in terms of deliberativeness across the four time nudges, they successfully convince users to spend more time to reflect. However, users' experience for each time nudge differs greatly. Significant effects of the *Time Nudges* were found to be subjectively helpful to users in terms of their thinking process and perceived helpfulness. One condition that stands out from the others seems to be *Simple Time Nudge*, as participants felt less pressure, did not find it distracting, found it helpful in their thinking process and generally found it useful.

6.2 Qualitative Results

In this section, we present our findings derived from the post-task feedback. We first detail the benefits of the time nudges followed by their drawbacks.

6.2.1 Common Benefit: Self-evaluate, Rethink and Reflect. Overall, users reported two key benefits in all of the time nudges. First, the time nudges allow users to self-evaluate, rethink and reflect. Participants appreciated the additional time provided by the time nudges as it offered a brief pause for reflection (Alert Message, P12), allow them to rethink their previous responses (Fixed Time Frame, P12), consider what they want to communicate in their writing and make adjustments and additions accordingly (Fixed Time Frame, P12; Simple Time Nudge, P4), and even enhanced the organization of their thoughts (Alert Message, P12). All of these are succinctly summarized by P17 from Fixed Time Frame condition: *'I think the timer [...] makes you want to go back and adjust what you wrote especially for those people who completed their thoughts very quickly. It allows a bit of time to finish, reflect and add more if needed.'*

6.2.2 Common Benefit: Review, Re-read and Checks. Secondly, users reported that the time nudges allow them to review, re-read and to check for mistakes. They appreciated the opportunity for revisions and error avoidance from posting anything they might regret typing (Fixed Time Frame, P18) and to correct grammatical or thought errors caused by momentary distractions (Comparison Time Nudge, P17). Others appreciated the opportunity for double checking what they have written to see if they could fix up their thoughts or strengthen them (Simple Time Nudge, P10). These benefits are succinctly put forth by P18 from Comparison Time Nudge condition: *'I like how the reminders made me check for silly typos. I always make them no matter how slow my original formulation/typing is. And a reminder to check and make sure you have missed anything can be useful for jogging the brain and refining initial thoughts.'* The above two benefits were present in the top key benefits for all of the time nudges.

6.2.3 Fixed Time Frame: More Focused. Unique benefits pertaining to the specific time nudges are described below. For Fixed Time Frame, users reported that it allows them to be more focused at the task at hand. *'It allows me to concentrate as best [as I could] so that I can answer the question within the time limit.'* (P2).

6.2.4 Simple Time Nudge: Confirmation. For Simple Time Nudge, specific benefits include serving as a confirmation for users to be sure of their opinions. *'I like how it actually makes me want to check over what I had written and to really take in the ideas and thoughts I have and make sure they're the ones that I want to [say].'* (P3). Another participant echoed the same: *[...] having the opportunity to review, reflect and confirm that this is really your opinion. This is good so people don't make a hasty opinion.'* (P16)

6.2.5 Simple Time Nudge & Alert Message: Prevent Impulsive Comments. Both Simple Time Nudge and Alert Message prevent impulsivity. Participants were quoted saying that the time nudges help them realize that: *'we can't be anxious and should pay more attention; not to be hasty, and be sure that our writing reflects our true opinion.'* (Simple Time Nudge, P14). As such, it encourages thoughtful contemplation on the issue rather than providing spontaneous reactions (Alert Message, P13). This could also help prevent negative online phenomena, as participants noted that both time nudges provided an opportunity to cool down before posting a hot-headed comment in a rage (Simple Time Nudge, P12) and may compel individuals to change their minds before posting toxic comments (Alert Message, P18).

6.2.6 Alert Message: Encourage Sincere Opinion. Moreover, Alert Message was reported to encourage users to provide a more sincere opinion. *'It helps me give the best answers I can provide knowing that I had the time to do so.'* (P3). Another

participant also echoed that it prompts them to give a more truthful response: '*I think what I like most is you do find yourself thinking deeper about what you really feel and think, and I think you give a more truthful, well-thought out response.*' (P7).

6.2.7 Common Drawbacks: Time-consuming/Extra steps. A key dislike present in all time nudges is that they can be time-consuming. In Fixed Time Frame, a user responded: '*I did not like that it takes a very long time for the submit button to appear, and I had to spend the extra time either on filler or rambling.*' (P9). For Simple Time Nudge, users noted: '*I like least that I need to spend additional time to reflect on my communication which isn't always what I do.*' (P4). Overall participants complained about the extra steps required to validate their messages in every condition.

6.2.8 Fixed Time Frame: Time Pressure & Being Forced. Unique drawback for Fixed Time Frame is that it induces time pressure on users. '*I hate the way it causes the likelihood of panicking and hurriedness.*' (P16), and '*I don't think working against a running clock is the best way. Some people who are very indecisive and anxious would need to write their thoughts down.*' (P17). Additionally, users feel forced when adhering to the strict time rules. '*I hate how it arbitrarily stops you from advancing. People don't like being forced to do things and I am one of them.*' (P18).

6.2.9 Comparison Time Nudge: Self-doubt and Discouraging & Induce Pressure. Specific drawback of the Comparison Time Nudge is that it induces self-doubt and discourages users which has a spillover effect of inducing pressure. '*It sounds kind of condescending. Like they're telling you your work is probably wrong or subpar just because you took less time.*' (P4). Another user also noted that: '*I dislike being compared to others, especially since I had spent a lot more time than others, but was still being told to slow down. It felt more like an admonishment overall, rather than an opportunity or encouragement.*' (P18). Another user felt discouraged: '*This prompt suggests that I did not spend enough time thinking about the topic although I feel that I had. This is discouraging.*' (P6). One user even felt self-conscious and felt the need to change his/her way of doing: '*I felt like I didn't write enough. I felt self-conscious on what I was writing and I feel the need to change.*' (P11). Another lost confidence in his/her responses: '*The prompt took away some of the confidence I had that my answer was "correct".*' (P14). As such, these negative implications pressurized users: '*It makes me feel pressured knowing that I had spent lesser time than others.*' (P15).

6.2.10 Alert Message: Distractive & Annoying. Unique drawback for Alert Message is being distractive, leading to some users indicating that the time nudge is annoying. '*I was actually kind of annoyed because the reminder came up when I only typed 3-4 words. It interfered with my thought pattern[s] because I was halfway typing a sentence. It stopped my train of thought, right when I was at the mid-sentence. It stopped me and I was kind of put out. It might help someone rephrase or re-frame their written text, but mostly it make me kind of mad.*' (P2). Another was quoted saying: '[...] That pop-up box was really annoying.' (P17).

Key benefits and drawbacks for the respective time nudges are summarized in Table 7. For a comprehensive review of the feedback associated with the four time nudges, please refer to Appendix Figures 11 - 18.

6.3 Discussion

In this section, we discuss our findings to answer RQ2. We synthesized both quantitative data (see section 6.1) and qualitative data (see section 6.2) as well as findings from our initial study to provide a comprehensive assessment of the time nudges on deliberativeness.

6.3.1 Time Length is Key. Overall, our results from studies 1 and 2 showed that the length of time to reflect is key.

Table 7. Key Benefits and Drawbacks for each Time Nudge. Note that this table presents only the most significant benefits and drawbacks. Benefits and drawbacks are ranked with the top being the most key for the respective time nudge. For benefits and drawbacks that are ranked on the same footing, they are in the colour blue (benefits) and pink (drawbacks). Note that Simple Time Nudge only has one key drawback.

Time Nudge Interface	Key Benefits	Key Drawbacks
Fixed Time Frame	Self-evaluate, Rethink & Reflect Review, Re-read & Check More Focused	Time Pressure Time Consuming Being Forced
Simple Time Nudge	Self-evaluate, Rethink & Reflect Review, Re-read & Check Confirmation Prevent Impulsive Comments	Time Consuming
Comparison Time Nudge	Self-evaluate, Rethink & Reflect Review, Re-read & Check	Self-doubt, Discouraging Time Consuming Pressure
Alert Message	Self-evaluate, Rethink & Reflect Prevent Impulsive Comments Review, Re-read & Check Encourage Sincere Opinions	Distractive Time Consuming Annoying

Longer Time is Beneficial... Results from study 1 showed increase deliberativeness with longer reflection time length, especially between One/Two minutes conditions compared to Three/Four-Five minutes condition. This suggests that, for minute-scale deliberations, reflection time is key to increasing deliberativeness of online opinions.

...But has Diminishing Returns. We did however notice that the increase in deliberativeness was plateauing after the Three minutes mark. While this specific value may be topic-dependent, forcing users to wait too long may not be beneficial, as they would simply wait it out, or simply drop and not post opinions. Our modelling of deliberativeness over time shows a clear logarithmic relationship between the two quantities.

6.3.2 Time Nudges Trigger Time-Induced Reflection. Study 2 showed that each time nudge effectively extended participants' reflection time, approaching or surpassing the optimal reflection time determined in study 1.

Time Nudges may Not Improve Deliberativeness... From our findings, we found that the different time nudges may not improve overall deliberativeness (see section 6.1.1).

...But Trigger Time-Induced Reflection. However, each time nudge managed to convince participants to spend more time and reach around the optimal reflection time determined in study 1 (see section 6.1.2). Moreover, the key benefit these time nudges served was to allow users to self-evaluate, rethink and reflect better (see section 6.2). This would suggest that the time nudges at some point triggered time-induced reflection - using time to help participants to reflect more - but we could not sieve out which time nudge had done better in terms of deliberative quality. However, users' experience with each time nudge greatly differs as we discussed below.

6.3.3 Fixed Time Frame Induces Stress. Fixed Time Frame was intended to stop users from progressing through only after reflecting for a certain time duration. However, both quantitative and qualitative results showed that it induces a psychological stress of time pressure. While this stress could be an asset for some such as making users to be more focused at the task (see section 6.2.3), it may not be well-received by the majority. As seen from Table 7, users indicated

time pressure to be their top drawback. Thus, our findings suggest that the **usefulness of Fixed Time Frame is very much dependent on the users' tolerance to stress**. As the majority do not sit well with it, '*working against a running clock*' might not be the best way to support reflection. In this way, Fixed Time Frame could be seen as a **double-edge sword**. It does allow users to reflect and review before posting but the key drawback is having to endure the psychological stress. This time nudge was an interesting baseline that allowed us to understand the interplay between time, user experience and deliberativeness.

6.3.4 Simple Time Nudge Subjectively Better Overall. Our quantitative results showed a significant effect of Simple Time Nudge on *Thinking Process* and *Perceived Helpfulness*. This was also visibly seen from users' feedback wherein the top three key benefits were the ability to reflect, review, provide assurance and confirmation as well as preventing impulsivity.

A closer look at the reflection time induced by Simple Time Nudge amounts to 5.8 minutes (see section 6.1.2) which is beyond the optimal time deduced in study 1. Yet, no significant difference was reported in terms of deliberativeness which is inline with study 1 results. In this case, Simple Time Nudge may have increased deliberativeness by means of helping users to reflect better, depicted in both our quantitative and qualitative results. But **any additional reflection time beyond optimality helps users psychologically by means of making them 'feel better'** as they are able to look back at their previous response before submission. Providing confirmation and assurance was reflected as one of the key benefit of Simple Time Nudge. However, **this assurance comes at a cost as users reported that this process is time consuming** and not all users have the available time to do so (see Table 7 which depicts time consuming as the sole key drawback). Hence, additional reflection beyond optimality induced by the time nudge only gives users a **peace of mind** and does not seek to improve their actual deliberativeness.

6.3.5 Comparison Time Nudge - Don't Compare Users' Performance for Reflection. We design Comparison Time Nudge using a goal-based time anchor to influence users to move towards reflecting arbitrarily close to the optimal reflection time. We do this by comparing the actual reflection time users took versus the time they should work towards, in the hopes of increasing their deliberative quality. Instead, we found opposite effects. Users found that the comparison is discouraging them, with some saying that it is an '*admonishment*', leading them to feel pressurized. This is in contrast with previous studies that a judgemental and goal-based anchor bring change to individuals' performance [16, 69]. Furthermore, as individuals react negatively towards Comparison Time Nudge - having feelings of self-consciousness and lowering of their self-confidence (see section 6.2.9), naturally they perceive the nudge to be more distractive than Simple Time Nudge (see section 6.1.3). Distractiveness here could have implied that the time nudge was more disturbing to users as it caused them to have unfavourable feelings. This indicates that **reflection is a personal thing and therefore, should not be compared with others**.

6.3.6 Alert Message for Sincere Opinions but Needs Tweaking. Users' feedback that the Alert Message was distracting, causing them to lose their train of thoughts (see section 6.2.10). This may be because we set the Alert Message to appear at 20 seconds (see section 5.1.4) which might be too early. To mitigate this, the Alert Message could be delayed at a longer interval (e.g., 40 seconds) but still appear before users submit their opinions. This adjustment could help users provide more thoughtful and non-hasty opinions, as users found the Alert Message to be a good reminder (see Table 7). However, it is challenging to predict exactly when users will submit their opinions, making the timing of nudge inherently uncertain [14].

7 IMPLICATIONS FOR DESIGN

7.1 Design Implications

To our knowledge, this work is the first to explore the impact of reflection time on deliberativeness in minute-scale deliberation. Based on the outcomes of studies 1 and 2, we have derived guidelines to inform designers and researchers about the influence of reflection time in this context.

7.1.1 Consider the Optimal Timing for Reflection (Study 1). We found that there exists a Goldilocks principle of ‘optimal’ time in reflection. Specifically, there is a diminishing marginal rate of return for deliberative quality after the optimal time. Hence, it would be meaningless to just ask users to ‘*take more time to reflect*’. Instead, we propose encouraging users to reflect close to or at the optimal time. This is because deliberative quality significantly improves for users who reflect below optimal time. As it might be difficult to find the exact optimality point, based on our study, we propose a range of three minutes to four minutes for reflection in the context of minute-scale deliberation.

7.1.2 Time Factors to Consider to Enhance Reflection (Study 2). We found that each time nudge uniquely affected users’ experiences. Based on our results, we created a set of five design guidelines (G1-G5).

G1: Avoid Permanent Timers on the Interface. For Fixed Time Frame, restricting users to reflect against a countdown timer significantly increases their stress, which is not ideal. To mitigate this, we could remove the auto submission feature and have the countdown timer to be used only at the start to allow users to read and think. This would take away the effect of time pressure as users can submit their responses as and when they are done. Putting the timer at the start would also stimulate reflection by getting users to get into the momentum. In cases where working against a running clock proves to be effective, a possible mitigation could be using the concept of time blindness [8] by hiding the timer and only allowing the submit button to appear after a certain reflection time is reached. All this while, letting users know that the submit button will appear after a certain reflection time (e.g., 3.5 minutes). This would mean that although users innately know that there is a timer, they are unable to sense the passing of time, thus reducing their stress.

G2: Confirmation and Validation is Essential. Confirmation and validation proves to be essential in the reflection process, evident in the case of Simple Time Nudge. Though it may not necessarily translate to an increase in deliberativeness, taking in one’s own thoughts before posting is perceived to be helpful in users’ thinking process. Therefore, having an opportunity for users to check back before posting may help in their psychological thinking.

G3: Reflection is Personal - Don’t Compare Performance. From Comparison Time Nudge, we learnt that reflection is personal and thus should not be compared with others. Hence, we propose that designers should not design time nudges that seek to compare oneself with others.

G4: Careful use of Time Nudge in a Non-Distractive Way. Finally, time nudges could act as a helpful and non-distractive reminder if it appears at a later timing but just before users post their opinions, as in the case of Alert Message. However, this could be tricky as it can be hard to find the best time interval for the nudge to appear [14]. From our results, we suggest that the nudge could occur between two to three minutes, at the half past mark time interval just before the optimal time ends for the targeted behaviour to take place.

G5: Use of Modals for Time Nudges. The time nudges in this study were primarily delivered through modals, which are direct and attention-grabbing. As indicated by our qualitative findings, some of these nudges, particularly the

Comparison Time Nudge, may not be ideal for conveying subtle information such as optimal reflection time. The use of modals was intended to ensure participants clearly recognized the nudge, allowing an assessment on how reflection time could be influenced by explicit prompts. However, more implicit approaches, such as a subtle indicator near the submit button or a color-changing icon, could communicate this information in a less intrusive way, enabling users to engage with the reflection process more naturally.

Future research could explore the effects of more implicit nudges on user experience and deliberative quality. A comparative study between different formats (modal vs. implicit) could provide valuable insights into how different delivery methods of a nudge influence reflection time, user engagement, and overall perceptions of the interface.

7.1.3 Moving Beyond Time Nudges (Study 2). In study 2, we designed different time nudges to encourage users who reflected below the optimal threshold to take more time, aiming to guide them toward the optimal reflection period. Our findings show that while the time nudges successfully extended reflection time, they did not result in an improvement in deliberativeness. This suggests that simply increasing reflection time does not necessarily enhance deliberative quality. These results align with our findings from study 1, where we observed diminishing returns after the optimal reflection time. The results highlight nuanced dynamics, indicating that while time nudges can extend reflection, improving deliberative quality beyond the optimal time may require more than just additional time. For users who have already reflected for the optimal duration, we recommend employing non-time nudges to further support the reflection process. For instance, the use of thinking prompts [101] could complement the timing approach and encourage deeper engagement.

7.2 Aligning Reflection Time with Deliberation Goals

While the three measurements of deliberativeness in study 1 emphasize on the breadth of arguments, it is possible that longer reflection times promote more focused deliberation, enhancing other aspects of deliberativeness (see section 2.2 that deliberativeness is multifaceted). Participants might shift from generating a broad range of arguments to refining a smaller set of more carefully considered points. This narrowing in argument diversity could indicate deeper engagement.

This would suggest that the effect of extended reflection time on deliberativeness might depend on the specific goals of the deliberation. For instance, if the aim is to explore a wide range of perspectives, shorter reflection times might encourage more diverse contributions. On the other hand, longer reflection periods could facilitate the refinement of participants’ thoughts if the goal is to foster deeper, more focused deliberation.

Future work could assess reflection time on different dimensions of deliberativeness. Additionally, it would be interesting to explore the impacts of non-time-based nudges in encouraging participants to reflect on different aspects of a topic over time. This would reduce the likelihood of repetitive thinking, allowing for a richer, more balanced reflection process.

7.3 Implications to Other Areas

While the task performed in this study was specific and done in a minute-scale deliberation context, we believe that our results could pave the way towards leveraging reflection time in other contexts.

7.3.1 Online Discussions. Online discussions enable users to engage anonymously or pseudonymously, which can lead to toxic online disinhibition [54, 55, 62, 90, 97]. Trolling phenomena such as flame wars occurs on impulse [97] and derail constructive societal and political discourse [97]. Moreover, poor design of online discussions platforms can further exacerbate these issues by encouraging ill-intentioned behavior.

Integrating a reflection duration on platforms like Reddit could reduce the prevalence of such phenomena by preventing impulsivity and regrets associated with users' postings [99], evident from our findings. The idea of giving time for users to reflect before posting has been implemented on subreddit /r/discussion_patiente/⁹ (i.e., 'patient discussion' in French) where commenters need to wait 24 hours to fully reflect before posting. Our results suggest that even a short delay of three to four minutes can significantly enhance deliberativeness, without the need for an excessively long wait.

Moreover, specialized communities such as subreddit/r/AskHistorians/¹⁰, which enforce strict community rules and emphasize quality answers (i.e., '*Answers must be in-depth and comprehensive, or they will be removed*'), could benefit from integrating a time nudge. This could enhance response quality by encouraging users to invest more time in thoughtful reflection before posting, potentially reducing the need for moderators over time.

7.3.2 Online Review Systems. Online reviewing platforms such as Yelp or Rotten Tomatoes could benefit from adding a reflection delay before users post reviews. For instance, allowing users to post only after a two minutes reflection period. This can be easily implemented by displaying the submit button after two minutes. Such a delay could foster higher levels of empathy and responsible online behaviour. For instance, users unhappy with a restaurant service can cool down before posting, preventing impulsive comments – a key benefit reflected in our results. Additionally, it could reduce review-bombing¹¹, where products or services are unfairly judged based on biases rather than merits. As immediate posting is not crucial, leveraging time nudges could lead to more nuanced and balanced reviews.

7.3.3 Beyond Human-Generated Content Platforms. Time nudges designed for minute-scale deliberation can be adapted in other contexts such as decision-making processes of non-human agents. For instance, integrating reflection periods in autonomous systems to allow the system to consult its decision-making algorithms can enhance decision quality.

In hybrid human-AI collaborative environments, integrating a reflection period can improve the quality of joint decision-making and optimize the interaction between human and AI. For instance, after the AI presents information, a prompt can appear, suggesting the human to take a two-minute reflection period to reflect on the AI's suggestions and consider their own knowledge. This would ensure that decisions made are not hasty and allow for a more thorough deliberation process, combining the strengths of both human judgment and artificial intelligence.

8 LIMITATIONS

There are limitations in this work. Firstly, the tasks in both studies were specific, aiming to achieve depth and clarity in understanding the influences of reflection time within a well-defined context [86]. While this establishes internal validity, it may limit generalizability and ecological validity. Assessing ecological validity relies on first establishing internal validity [11, 12, 23, 86], thus our focused approach allows for a thorough examination of reflection time on deliberativeness. Future research could extend the application of reflection time on other contentious topics with varying complexity and nature.

Similarly, both studies investigate reflection times in the context of minute-scale deliberation. This focus allows us to capture the fast-paced nature and rapid interaction on online deliberation platforms [53, 53, 92, 96, 102, 103]. Future work could explore different deliberation contexts, including those with longer deliberation time frames.

⁹ subreddit Patient Discussion

¹⁰ subreddit Ask Historians

¹¹Hollywood Review-Bombing

Lastly, in study 1, we identified an optimal reflection time of four minutes using a public issue of four-day workweek. However, this optimal reflection time may vary with the complexity of the issue/theme discussed and the exact nature of the task in realistic contexts. Therefore, we do not claim the generalization of the optimal timing of four minutes.

9 CONCLUSION

In this work, we investigated how reflection time impacts deliberativeness in minute-scale deliberations. We provided the missing link between the two by conducting two studies. In the first study, our results suggest that there exists an optimal reflection time on the quality of minute-scale deliberation when writing a short opinion comment, and that increasing reflection time is most favourable when users reflect below the optimum. In the second study, we provide design considerations for enacting the dimension of time where it is warranted to account for better deliberative quality in minute-scale deliberations, through direct applications of nudging techniques. Our results expand current work on minute-scale deliberation and is a first step towards a more complete account of how time supports reflection in the deliberation process as well as how researchers and designers can better incorporate time in minute-scale deliberations.

REFERENCES

- [1] Carlos Alós-Ferrer and Johannes Buckenmaier. 2021. Cognitive sophistication and deliberation times. *Experimental Economics* 24, 2 (2021), 558–592.
- [2] Ross C Anderson, Meg Guerreiro, and Joanna Smith. 2016. Are all biases bad? Collaborative grounded theory in developmental evaluation of education policy. *Journal of Multidisciplinary Evaluation* 12, 27 (2016), 44–57.
- [3] Kevin Arceneaux and Ryan J Vander Wielen. 2017. *Taming intuition: How reflection minimizes partisan reasoning and promotes democratic accountability*. Cambridge University Press.
- [4] Albert Bandura. 1982. Self-efficacy mechanism in human agency. *American psychologist* 37, 2 (1982), 122.
- [5] Jonathan Barnes. 1984. *Complete works of Aristotle, volume 1: The revised Oxford translation*. Vol. 1. Princeton University Press.
- [6] Jonathan Barnes (Ed.). 1984. *Complete Works of Aristotle, Volume 2: The Revised Oxford Translation*. Princeton University Press.
- [7] John Carlo Bertot, Paul T. Jaeger, Sean Munson, and Tom Glaisyer. 2010. Social Media Technology and Government Transparency. *Computer* 43, 11 (2010), 53–59. <https://doi.org/10.1109/MC.2010.325>
- [8] Kevin Birth. 2017. *Time Blind*. Springer.
- [9] George F. Bishop, Robert W. Oldendick, Alfred J. Tuchfarber, and Stephen E. Bennett. 1980. Pseudo-opinions on public affairs. *Public Opinion Quarterly* 44 (1980), 198–209. <https://doi.org/10.1086/268584>
- [10] James Bohman. 2000. *Public deliberation: Pluralism, complexity, and democracy*. MIT press.
- [11] Kaya Cahit. 2015. Internal validity: A must in research designs. *Educational Research and Reviews* 10, 2 (2015), 111–118.
- [12] Donald T Campbell and Julian C Stanley. 2015. *Experimental and quasi-experimental designs for research*. Ravenio books.
- [13] Joseph N Cappella, Vincent Price, and Lilach Nir. 2002. Argument repertoire as a reliable and valid measure of opinion quality: Electronic dialogue during campaign 2000. *Political Communication* 19, 1 (2002), 73–93.
- [14] Ana Caraban, Evangelos Karapanos, Daniel Gonçalves, and Pedro Campos. 2019. 23 ways to nudge: A review of technology-mediated nudging in human-computer interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–15.
- [15] Dorwin Cartwright. 1941. Relation of decision-time to the categories of response. *The American Journal of Psychology* 54, 2 (1941), 174–196.
- [16] Daniel Cervone and Philip K Peake. 1986. Anchoring, efficacy, and action: The influence of judgmental heuristics on self-efficacy judgments and behavior. *Journal of Personality and social Psychology* 50, 3 (1986), 492.
- [17] Simone Chambers. 2003. Deliberative democratic theory. *Annual review of political science* 6, 1 (2003), 307–326.
- [18] Kathy Charmaz. 2014. *Constructing grounded theory*. sage.
- [19] Nan-chen Chen, Rafal Kocielnik, Margaret Drouhard, Vanessa Peña-Araya, Jina Suh, Keting Cen, Xiangyi Zheng, and Cecilia R Aragon. 2016. Challenges of applying machine learning to qualitative coding. In *ACM SIGCHI Workshop on Human-Centered Machine Learning*.
- [20] Thomas Christiano and John Christman. 2009. *Debates in Political Philosophy*. Wiley Online Library.
- [21] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. Routledge.
- [22] Lucas Colusso, Gary Hsieh, and Sean A Munson. 2016. Designing closeness to increase gamers’ performance. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 3020–3024.
- [23] Thomas D Cook, Donald Thomas Campbell, and Arles Day. 1979. *Quasi-experimentation: Design & analysis issues for field settings*. Vol. 351. Houghton Mifflin Boston.
- [24] Juliet Corbin and Anselm Strauss. 2014. *Basics of qualitative research: Techniques and procedures for developing grounded theory*. Sage publications.
- [25] Todd Davies and Reid Chandler. 2013. Online deliberation design: Choices, criteria, and evidence. *arXiv preprint arXiv:1302.5177* (2013).

- [26] Todd Davies and Seeta Peña Gangadharan. 2009. *Online Deliberation: Design, Research, and Practice*. CSLI Publications.
- [27] Robert Dawson. 2011. How significant is a boxplot outlier? *Journal of Statistics Education* 19, 2 (2011).
- [28] Rianne Dekker and Victor Bekkers. 2015. The contingency of governments' responsiveness to the virtual public sphere: A systematic literature review and meta-synthesis. *Government Information Quarterly* 32, 4 (2015), 496–505. <https://doi.org/10.1016/j.giq.2015.09.007>
- [29] John S. Dryzek. 2002. *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*. Oxford University Press. <https://doi.org/10.1093/019925043X.001.0001>
- [30] Robert A Emmons and Ed Diener. 1986. Influence of impulsivity and sociability on subjective well-being. *Journal of Personality and social psychology* 50, 6 (1986), 1211.
- [31] Thomas Erickson and Wendy A Kellogg. 2000. Social translucence: an approach to designing systems that support social processes. *ACM transactions on computer-human interaction (TOCHI)* 7, 1 (2000), 59–83.
- [32] Cynthia R. Farina, Dmitry Epstein, Josiah Heidt, and Mary J. Newhart. 2014. Designing an Online Civic Engagement Platform: Balancing "More" vs. "Better" Participation in Complex Public Policymaking. *Int. J. E-Polit.* 5, 1 (jan 2014), 16–40. <https://doi.org/10.4018/ijep.2014010102>
- [33] Cynthia R Farina, Mary Newhart, and Josiah Heidt. 2012. Rulemaking vs. democracy: Judging and nudging public participation that counts. *Mich. J. Envtl. & Admin. L.* 2 (2012), 123.
- [34] James D Fearon. 1998. Deliberation as. *Deliberative democracy* 1 (1998), 44.
- [35] Leon Festinger. 1954. A theory of social comparison processes. *Human relations* 7, 2 (1954), 117–140.
- [36] Rowanne Fleck and Geraldine Fitzpatrick. 2010. Reflecting on reflection: framing a design landscape. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*. 216–223.
- [37] Joseph L Fleiss, John C Nee, and J Richard Landis. 1979. Large sample variance of kappa in the case of different sets of raters. *Psychological bulletin* 86, 5 (1979), 974.
- [38] Dennis Friess and Christiane Eilders. 2015. A systematic review of online deliberation research. *Policy & Internet* 7, 3 (2015), 319–339.
- [39] Jie Gao, Kenny Tsu Wei Choo, Junming Cao, Roy Ka-Wei Lee, and Simon Perrault. 2023. CoAIcoder: Examining the Effectiveness of AI-assisted Human-to-Human Collaboration in Qualitative Analysis. *ACM Transactions on Computer-Human Interaction* (2023).
- [40] John Gastil, Robert Richards, and Katherine Knobloch. 2014. Vicarious Deliberation: How the Oregon Citizens' Initiative Review Influenced Deliberation in Mass Elections. *International Journal of Communication* 8, 0 (2014). <https://ijoc.org/index.php/ijoc/article/view/2235>
- [41] Robert E. Goodin. 2000. Democratic Deliberation within. *Philosophy & Public Affairs* 29, 1 (2000), 81–109. <http://www.jstor.org/stable/2672865>
- [42] Robert E. Goodin and Simon J. Niemeyer. 2003. When Does Deliberation Begin? Internal Reflection versus Public Discussion in Deliberative Democracy. *Political Studies* 51, 4 (2003), 627–649. <https://doi.org/10.1111/j.0032-3217.2003.00450.x> arXiv:<https://doi.org/10.1111/j.0032-3217.2003.00450.x>
- [43] Rúben Gouveia, Fábio Pereira, Evangelos Karapanos, Sean A Munson, and Marc Hassenzahl. 2016. Exploring the design space of glanceable feedback for physical activity trackers. In *Proceedings of the 2016 ACM international joint conference on pervasive and ubiquitous computing*. 144–155.
- [44] Todd Graham and Tamara Witschge. 2003. In search of online deliberation: Towards a new method for examining the quality of online discussions. (2003).
- [45] Miriam Greis, Florian Alt, Niels Henze, and Nemanja Memarovic. 2014. I can wait a minute: uncovering the optimal delay time for pre-moderated user-generated content on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1435–1438.
- [46] Jürgen Habermas. 1984. *The theory of communicative action*. Beacon Press.
- [47] May Hamblin and Philippe Crisp. 2022. Negative Focus, Self-Doubt, and Issues of 'Tool Proficiency': Beginner-Coaches' Reflections on Reflective Practice. *Physical Culture and Sport. Studies and Research* (2022).
- [48] Pelle Guldborg Hansen and Andreas Maaløe Jespersen. 2013. Nudge and the manipulation of choice: A framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation* 4, 1 (2013), 3–28.
- [49] Marian Harbach, Markus Hettig, Susanne Weber, and Matthew Smith. 2014. Using personal examples to improve risk communication for security & privacy decisions. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2647–2656.
- [50] Verlin B Hinsz, Lynn R Kalnbach, and Nichole R Lorentz. 1997. Using judgmental anchors to establish challenging self-set goals without jeopardizing commitment. *Organizational Behavior and Human Decision Processes* 71, 3 (1997), 287–308.
- [51] Daniel J Hruschka, Deborah Schwartz, Daphne Cobb St. John, Erin Picone-Decaro, Richard A Jenkins, and James W Carey. 2004. Reliability in coding open-ended data: Lessons learned from HIV behavioral research. *Field methods* 16, 3 (2004), 307–331.
- [52] Lawrence R. Jacobs, Fay Lomax Cook, and Michael X. Delli Carpini. 2009. *Talking Together: Public Deliberation and Political Participation in America*. The University of Chicago Press.
- [53] Sirkka L Jarvenpaa and Ann Majchrzak. 2010. Research commentary—Vigilant interaction in knowledge collaboration: Challenges of online user participation under ambivalence. *Information Systems Research* 21, 4 (2010), 773–784.
- [54] Adam N Joinson. 2003. Understanding the psychology of Internet behaviour: Virtual worlds, real lives. *Revista iberoamericana de educación a distancia* 6, 2 (2003), 190.
- [55] Adam N Joinson. 2007. Disinhibition and the Internet. In *Psychology and the Internet*. Elsevier, 75–92.
- [56] John Keane. 2000. Structural Transformations of the Public Sphere. In *Digital Democracy: Issues of Theory and Practice*, Kenneth L. Hacker and Jan van Dijk (Eds.). SAGE Publications Ltd, London, 70–89. <https://doi.org/10.4135/9781446218891>
- [57] Hyunwoo Kim, Haesoo Kim, Kyung Je Jo, and Juho Kim. 2021. StarryThoughts: Facilitating Diverse Opinion Exploration on Social Issues. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–29.

- [58] Hyunwoo Kim, Eun-Young Ko, Donghoon Han, Sung-Chul Lee, Simon T Perrault, Jihee Kim, and Juho Kim. 2019. Crowdsourcing perspectives on public policy from stakeholders. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6.
- [59] Juho Kim, Eun-Young Ko, Jonghyuk Jung, Chang Won Lee, Nam Wook Kim, and Jihee Kim. 2015. Factful: Engaging taxpayers in the public discussion of a government budget. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. 2843–2852.
- [60] Travis Kriplean, Jonathan Morgan, Deen Freelon, Alan Borning, and Lance Bennett. 2012. Supporting reflective public thought with considerit. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. 265–274.
- [61] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977), 159–174. <http://www.jstor.org/stable/2529310>
- [62] Noam Lapidot-Lefler and Azy Barak. 2012. Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. *Computers in human behavior* 28, 2 (2012), 434–443.
- [63] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 557–566.
- [64] Siân E Lindley, Richard Harper, and Abigail Sellen. 2009. Desiring to be in touch in a changing communications landscape: attitudes of older adults. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1693–1702.
- [65] Tanja Lischetzke and Michael Eid. 2003. Is attention to feelings beneficial or detrimental to affective well-being? Mood regulation as a moderator variable. *Emotion* 3, 4 (2003), 361.
- [66] Kathleen M MacQueen, Eleanor McLellan, Kelly Kay, and Bobby Milstein. 1998. Codebook development for team-based qualitative analysis. *Cam Journal* 10, 2 (1998), 31–36.
- [67] Bernard Manin. 2011. Comment promouvoir la délibération démocratique? Priorité du débat contradictoire sur la discussion. *Raisons politiques* 42, 02 (2011), 83–113.
- [68] Megh Marathe and Kentaro Toyama. 2018. Semi-automated coding for qualitative research: A user-centered inquiry and initial prototypes. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.
- [69] Sanju Menon, Weiyu Zhang, and Simon T Perrault. 2020. Nudge for deliberativeness: How interface features influence online discourse. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [70] Jennifer A Moon. 2013. *Reflection in learning and professional development: Theory and practice*. Routledge.
- [71] Brent Moritz, Enno Siemsen, and Mirko Kremer. 2014. Judgmental forecasting: Cognitive reflection and decision speed. *Production and Operations Management* 23, 7 (2014), 1146–1160.
- [72] Janice M Morse. 1997. "Perfectly healthy, but dead": the myth of inter-rater reliability. , 445–447 pages.
- [73] Robert S Moyer and Richard H Bayer. 1976. Mental comparison and the symbolic distance effect. *Cognitive Psychology* 8, 2 (1976), 228–246.
- [74] Lala Muradova. 2021. Seeing the other side? Perspective-taking and reflective political judgements in interpersonal deliberation. *Political Studies* 69, 3 (2021), 644–664.
- [75] W. Russell Neuman. 1986. *The Paradox of Mass Politics: Knowledge and Opinion in the American Electorate*. Harvard University Press.
- [76] Geoff Norman. 2010. Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education* 15, 5 (2010), 625–632.
- [77] Sounieil Park, Seungwoo Kang, Sangyoung Chung, and Junehwa Song. 2009. NewsCube: delivering multiple aspects of news to mitigate media bias. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 443–452.
- [78] Vincent Price, Joseph N Cappella, and Lilach Nir. 2002. Does disagreement contribute to more deliberative opinion? *Political communication* 19, 1 (2002), 95–112.
- [79] Elizabeth Rendon-Velez, Imre Horváth, and W Van der Vegt. 2012. A pilot study to investigate time pressure as a surrogate of being in haste. In *Proceedings of the ninth international symposium on tools and methods of competitive engineering*. 393–406.
- [80] K Andrew R Richards and Michael A Hemphill. 2018. A practical guide to collaborative qualitative data analysis. *Journal of Teaching in Physical education* 37, 2 (2018), 225–231.
- [81] Tim Rietz and Alexander Maedche. 2021. Cody: An AI-based system to semi-automate coding for qualitative research. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [82] Aldo Rustichini. 2009. Neuroeconomics: Formal Models of Decision Making and Cognitive Neuroscience. In *Neuroeconomics*. Elsevier, 33–46.
- [83] Jorge Saldivar, Cristian Parra, Marcelo Alcaraz, Rebeca Arteta, and Luca Cernuzzi. 2019. Civic technology for social innovation: A systematic literature review. *Computer Supported Cooperative Work (CSCW)* 28 (2019), 169–207.
- [84] Giovanni Sartori. 1987. *The Theory of Democracy Revisited*. Steven Bridges Press.
- [85] Neil C Schwertman, Margaret Ann Owens, and Robiah Adnan. 2004. A simple more general boxplot method for identifying outliers. *Computational statistics & data analysis* 47, 1 (2004), 165–174.
- [86] Marion K Slack and Jolaine R Draugalis Jr. 2001. Establishing the internal and external validity of experimental studies. *American journal of health-system pharmacy* 58, 22 (2001), 2173–2181.
- [87] Alexandru Spatariu, Kendall Hartley, and Lisa D. Bendixen. 2004. Defining and Measuring Quality in Online Discussions. *Journal of Interactive Online Learning* 2, 4 (2004). <https://www.ncolr.org/issues/jiol/v2/n4/defining-and-measuring-quality-in-online-discussions.html>
- [88] Marco R Steenbergen, André Bächtiger, Markus Spörndli, and Jürg Steiner. 2003. Measuring political deliberation: A discourse quality index. *Comparative European Politics* 1 (2003), 21–48.
- [89] Jennifer Stromer-Galley. 2007. Measuring deliberation's content: A coding scheme. *Journal of Deliberative Democracy* 3, 1 (2007).

- [90] John Suler. 2004. The online disinhibition effect. *Cyberpsychology & behavior* 7, 3 (2004), 321–326.
- [91] Fred S Switzer III and Janet A Snieszek. 1991. Judgment processes in motivation: Anchoring and adjustment effects on judgment and behavior. *Organizational Behavior and Human Decision Processes* 49, 2 (1991), 208–229.
- [92] Pamela Takayoshi. 2015. Short-form writing: Studying process in the context of contemporary composing technologies. *Computers and Composition* 37 (2015), 1–13.
- [93] Richard Thaler and Cass Sunstein. 2008. Nudge: The gentle power of choice architecture. *New Haven, Conn.: Yale* (2008).
- [94] Hobbes Thomas. 1946. *Leviathan*. London: Basil Blackwell.
- [95] Matthias Trénel. 2004. Measuring the quality of online deliberation. Coding scheme 2.0. Retrieved December 1, 2023 from http://www.wz-berlin.de/~trenel/tools/qod_2_0.pdf
- [96] Natalia Umansky and Christian Pipal. 2023. Dances, Duets, and Debates: Analysing political communication and viewer engagement on TikTok. *OSF Preprints*. September 21 (2023).
- [97] Birgit J Voggesser, Ranjit K Singh, and Anja S Göritz. 2018. Self-control in online discussions: Disinhibited online behavior as a failure to recognize social cues. *Frontiers in psychology* 8 (2018), 316647.
- [98] Yang Wang, Pedro Giovanni Leon, Alessandro Acquisti, Lorrie Faith Cranor, Alain Forget, and Norman Sadeh. 2014. A field trial of privacy nudges for facebook. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2367–2376.
- [99] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. 2011. "I regretted the minute I pressed share" a qualitative study of regrets on Facebook. In *Proceedings of the seventh symposium on usable privacy and security*. 1–16.
- [100] Lu Xiao, Weiyu Zhang, Anna Przybylska, Anna De Liddo, Gregorio Convertino, Todd Davies, and Mark Klein. 2015. Design for online deliberative processes and technologies: Towards a multidisciplinary research agenda. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. 865–868.
- [101] ShunYi Yeo, Gionnieve Lim, Jie Gao, Weiyu Zhang, and Simon Tangi Perrault. 2024. Help Me Reflect: Leveraging Self-Reflection Interface Nudges to Enhance Deliberativeness on Online Deliberation Platforms. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 806, 32 pages. <https://doi.org/10.1145/3613904.3642530>
- [102] Amy X Zhang, Michael S Bernstein, David R Karger, and Mark S Ackerman. 2024. Form-From: A Design Space of Social Media Systems. *arXiv preprint arXiv:2402.05388* (2024).
- [103] Cevin Zhang, Hemingxi Zheng, and Qing Wang. 2022. Driving factors and moderating effects behind citizen engagement with mobile short-form videos. *IEEE Access* 10 (2022), 40999–41009.
- [104] Weiyu Zhang. 2005. Are online discussions deliberate? A case study of a Chinese online discussion board. *Trípodos* 1, 1 (2005), 119–134.
- [105] Weiyu Zhang, Xiaoxia Cao, and Minh Ngoc Tran. 2013. The structural features and the deliberative quality of online discussions. *Telematics and informatics* 30, 2 (2013), 74–86.
- [106] Weiyu Zhang and Leanne Chang. 2014. Perceived speech conditions and disagreement of everyday talk: A proceduralist perspective of citizen deliberation. *Communication Theory* 24, 2 (2014), 124–145.
- [107] Weiyu Zhang, Tian Yang, and Simon Tangi Perrault. 2021. Nudge for Reflection: More Than Just a Channel to Political Knowledge. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–10.
- [108] Fengyuan Zhu, Ke Fang, and Xiaojuan Ma. 2017. Exploring the effects of strategy and arousal of cueing in computer-human persuasion. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2276–2283.

APPENDIX

The appendix contains supplementary data, which, while not part of the analysis, may provide additional details that can help in replicating the study.

Methodology Appendix A1: Content Analysis for Analyzing Qualitative Data

Data Preparation. Deliberativeness was operationalized through three measurements: argument repertoire, argument diversity and word count as discussed in section 2.2. Of the three, argument repertoire and argument diversity were derived from a content analysis of participants’ responses, which we elucidate as follows:

Collected textual data to the four-day workweek task was subjected to qualitative research techniques following the Grounded Theory method [18, 24]. This involved open coding of participants’ responses by two coders. Both coders were PhD students with respectively two and five years of experience in content analysis. A single independent coder was used for reliability checks and to review the coding process.

To ensure the consistency and clarity of the coding process [19, 68, 81], coding was done sentence-by-sentence where one or more codes from the codebook could be applied to each sentence.

To generate the codebook, a subset of the data (10% of the data from study 1 and 30% of the data from study 2) were distributed to the coders. Both coders independently examined the participants’ responses and proposed a set of codes. They then met to compare and discuss similar codes and to resolve code conflicts while focusing on (1) how relevant the codes were to the dependent variable of deliberativeness and (2) whether the code appeared in the sentence. After agreeing on an initial master list of codes, the coders followed the practice by Hruschka et al. [51] and MacQueen et al. [66] to derive a set of rules. These rules were meant to ensure efficient coding between the two coders in which they decide on the criteria of whether a specific sentence fell under a particular code. More specifically, an inclusion list (see Appendix Table 8) was generated to clearly define the code and to clarify which sentence should or should not constitute an instance of a code.

After the initial draft of the codebook was developed, the coders checked for inter-coder reliability rates in an iterative process. A total of three coding rounds were conducted and Cohen’s Kappa was computed: Argument Repertoire ($\kappa = 0.808$); Argument Diversity ($\kappa = 0.808$), both above the satisfactory threshold of 0.80 [37, 61, 72].

Afterwards, the entire set of responses was coded according to the finalized codebook which comprised of 24 codes. As Hruschka et al. [51] found that a large number of codes consequentially decreases the likelihood of any real agreement between the coders as there are just too many codes to choose from, we restricted the number of possible codes to less than 30.

Content Analysis. For argument repertoire, a relevant sentence is counted if it has at least one code. Sentences stating one’s personal preference (e.g., *“I am hugely in favour of four-day workweek.”*) or emotional response (e.g., *“I felt very happy for the idea on four-day workweek.”*) were not counted. For argument diversity, a theme is a unique code that is present in the participant’s opinion. Codes that were repeated were only counted once. An illustration of the computation for argument repertoire and argument diversity is depicted in Figure 9.

Open coding was conducted in two instances. In the first instance, open coding was conducted in both studies to prepare the data for analyzing deliberativeness quantitatively. The second instance was conducted in study 2 in which open coding was conducted by the primary author on participants’ feedback towards the interface-based time nudges.

Demographics Appendix B1: Demographics Profile

We found no significant differences in terms of the demographics profile for any of the experimental conditions in study 1. Similarly, we found no significant differences in terms of the demographics profile for any of the experimental conditions in study 2.

Supplementary Statistical Details C1: Additional Statistics for Study 2

Although not included in the primary analysis, we collated participants' perspectives of reflection time in the post-task survey of study 2. We asked participants to rate their perceptions to a question that reads "*Do you think it is important to provide time for people to think and reflect before they post their opinions online?*". They have to rate this question on a five-point Likert scale (1 = definitely not, 2 = probably not, 3 = might or might not, 4 = probably yes and 5 = definitely yes). Majority (86.1%) of the participants indicated 'yes' to providing time for people to reflect during the deliberation process (see Figure 10).

Supplementary Statistical Details C2: Regression Analysis for Study 1

Regression analysis of both linear and logarithmic model for the three deliberative measures (word count, argument repertoire and argument diversity) are shown in Table 11.

Benefits & Drawbacks for Study 2

Figures 11 - 18 show a snapshot of the benefits and drawbacks (both key and non-key benefits/drawbacks) for each time nudge. During the coding process, one or more codes could be applied to each participant's feedback.

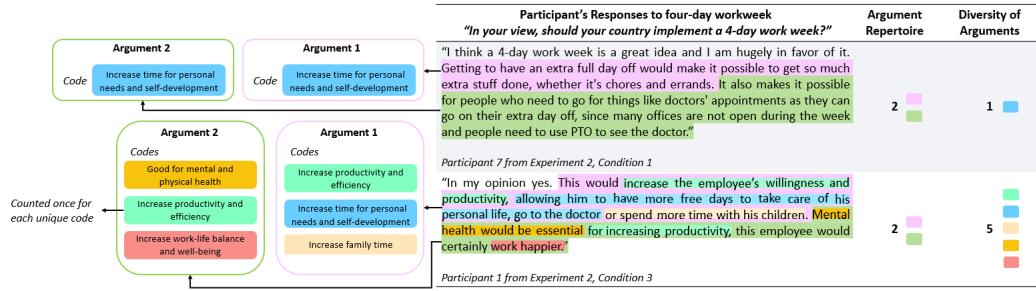


Fig. 9. Illustration of two participants' opinions. Both participants have the same number of argument repertoire but participant 1 has a higher diversity count. Note that even though P1 and P7 were subjected to different conditions in study 2, this does not impact the diversity of the arguments as participants were given the same task.

Table 8. A section of the finalized codebook for Study 1. The inclusion list is not exhaustive.

Codes	Inclusion List
4-day workweek is not generalizable	Depends on different types of work Depends on different industries Depends on cultures Depends on the country's customs Don't trust Japan's study
Good deal despite trade-offs	Good deal Gain and loss Comparison of trade-offs Merits of 4-day workweek outweighs longer working hours Additional 2 working hours is not a big deal compared to a whole day off
Increase time for personal needs and self development	More quality time for self Increase time for skills upgrading and/or learning More time for hobbies and/or pursuing other interests More time for personal business More time to attend to personal needs
Increase productivity and efficiency	Increase productivity/productive* Increase efficiency/efficient* Increase work output/able to accomplish more Good for working Increase concentration Increase motivation/able to focus better

* broadens the inclusion list to include various word endings and spellings

Table 9. Summary Statistics of the Demographic Profiles in Study 1

Demographics Profile		Reflection Time				
		1 min	2 mins	3 mins	4 mins	5 mins
Participants	Total number of participants (after excluding outliers)	20	20	20	18	20
Gender	Number of Males	13	11	13	10	13
	Number of Females	7	9	7	8	7
Country	USA	10	14	16	11	17
	India	5	4	3	6	1
	Europe (UK, Germany, Bulgaria, Italy)	2	2	0	0	0
	Canada	1	0	0	0	0
	Brazil	1	0	1	1	2
	Iran	1	0	0	0	0
Age	Mean Age	39.0	37.3	38.2	37.6	37.7
Familiarity	Mean Familiarity	60.6	65.0	61.5	61.6	53.5

Table 10. Summary Statistics of the Demographic Profiles in Study 2

Demographics Profile		Time Nudges			
		Fixed Time Frame	Simple Time Nudge	Comparison Time Nudge	Alert Message
Participants	Total number of participants	18	18	18	18
Gender	Number of Males	10	9	11	10
	Number of Females	8	9	7	7
	Prefer not to say	0	0	0	1
Country	USA	15	14	15	15
	India	1	1	0	3
	Europe (Italy, Brazil)	2	3	3	0
Age	Mean Age	36.2	39.7	35.6	44.3
Familiarity	Mean Familiarity	58.8	58.8	65.1	71.4

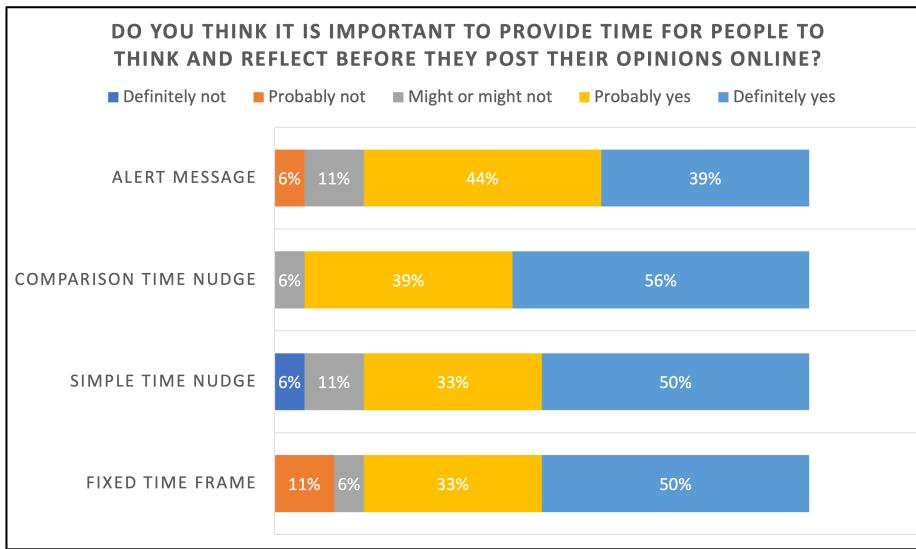


Fig. 10. Participants' perspectives on the importance of allowing time for reflection before posting. A significant majority 86.1% (62 out of 72 participants), expressed agreement by selecting 'yes' (probably yes and definitely yes).

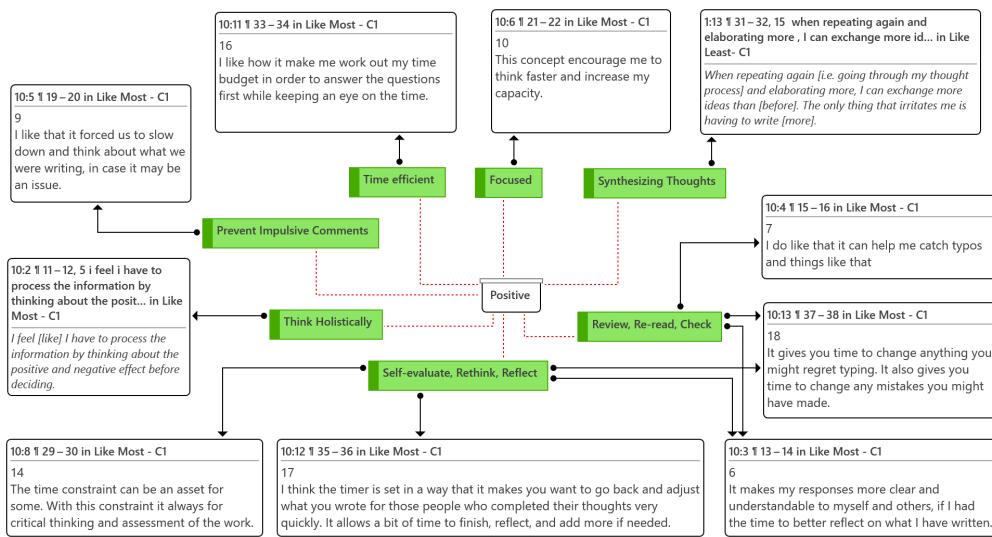


Fig. 11. Benefits: Fixed Time Frame

Table 11. Regression analysis of linear and logarithmic model for deliberativeness. Note: * $p<0.1$; ** $p<0.05$; *** $p<0.01$

<i>Dependent variable:</i>	
Word Count	
Time	15.559*** (2.979)
Constant	10.520 (9.024)
Observations	6
R ²	0.872
Adjusted R ²	0.840
Residual Std. Error	12.079 (df = 4)
F Statistic	27.277*** (df = 1; 4)

(a) Linear model for word count

<i>Dependent variable:</i>	
Argument Repertoire	
Time	0.645** (0.170)
Constant	0.744 (0.515)
Observations	6
R ²	0.782
Adjusted R ²	0.728
Residual Std. Error	0.690 (df = 4)
F Statistic	14.381** (df = 1; 4)

(c) Linear model for argument repertoire

<i>Dependent variable:</i>	
Argument Diversity	
Time	0.633** (0.159)
Constant	0.684 (0.483)
Observations	6
R ²	0.798
Adjusted R ²	0.748
Residual Std. Error	0.646 (df = 4)
F Statistic	15.804** (df = 1; 4)

(e) Linear model for argument diversity

<i>Dependent variable:</i>	
Word Count	
Time	25.727*** (2.156)
Constant	35.952*** (2.564)
Observations	6
R ²	0.973
Adjusted R ²	0.966
Residual Std. Error	5.582 (df = 4)
F Statistic	142.453*** (df = 1; 4)

(b) Logarithmic model for word count

<i>Dependent variable:</i>	
Argument Repertoire	
Time	1.131*** (0.081)
Constant	1.765*** (0.096)
Observations	6
R ²	0.980
Adjusted R ²	0.975
Residual Std. Error	0.210 (df = 4)
F Statistic	194.663*** (df = 1; 4)

(d) Logarithmic model for argument repertoire

<i>Dependent variable:</i>	
Argument Diversity	
Time	1.101*** (0.070)
Constant	1.690*** (0.083)
Observations	6
R ²	0.984
Adjusted R ²	0.980
Residual Std. Error	0.181 (df = 4)
F Statistic	249.136*** (df = 1; 4)

(f) Logarithmic model for argument diversity

Goldilocks Principle of 'Optimal' Reflection Time

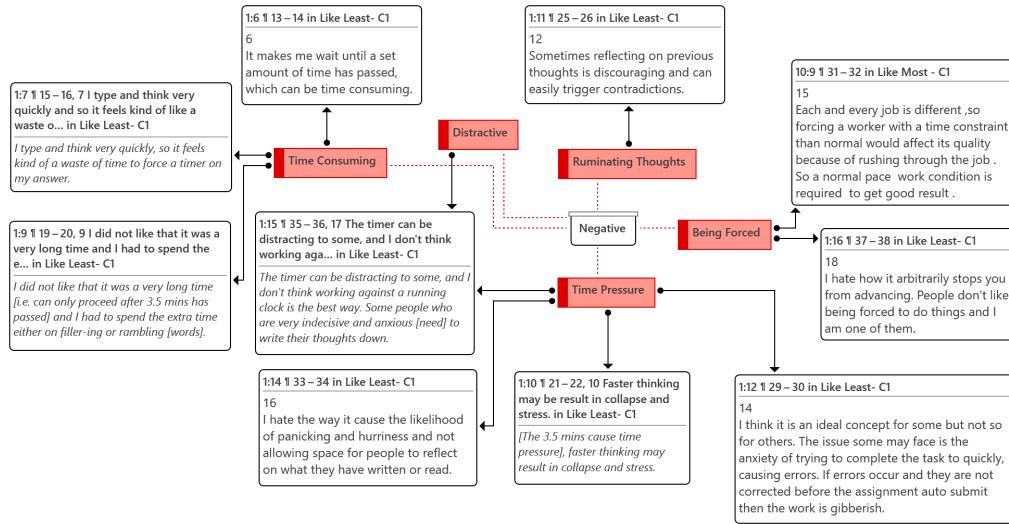


Fig. 12. Drawbacks: Fixed Time Frame

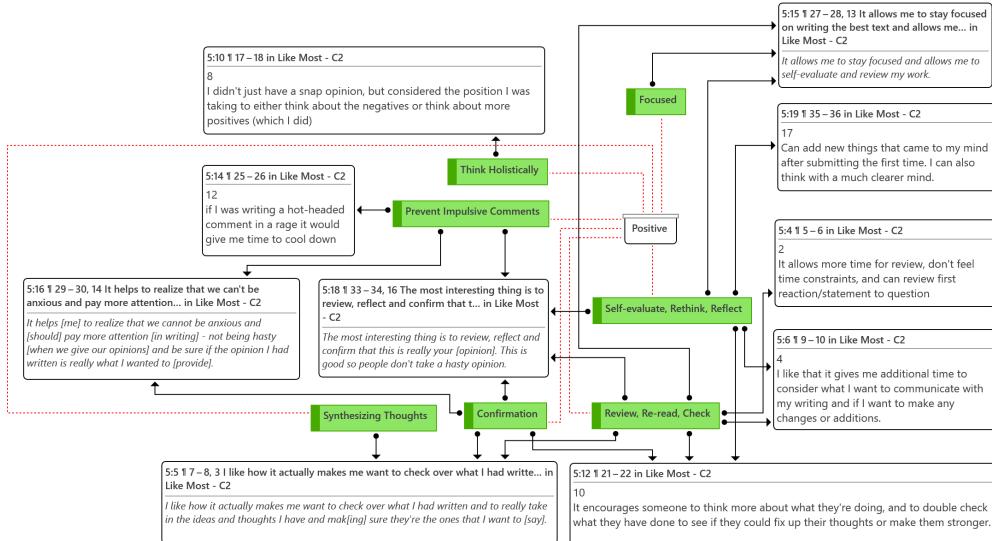


Fig. 13. Benefits: Simple Time Nudge

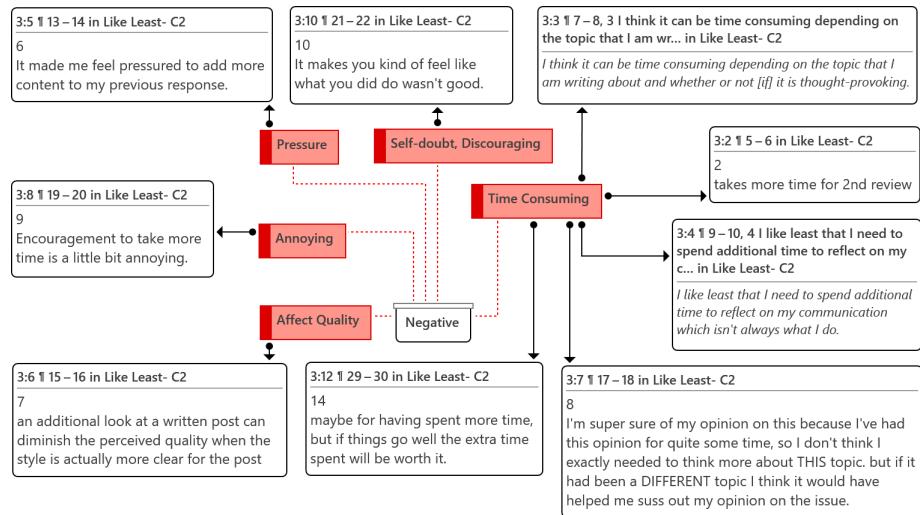


Fig. 14. Drawbacks: Simple Time Nudge

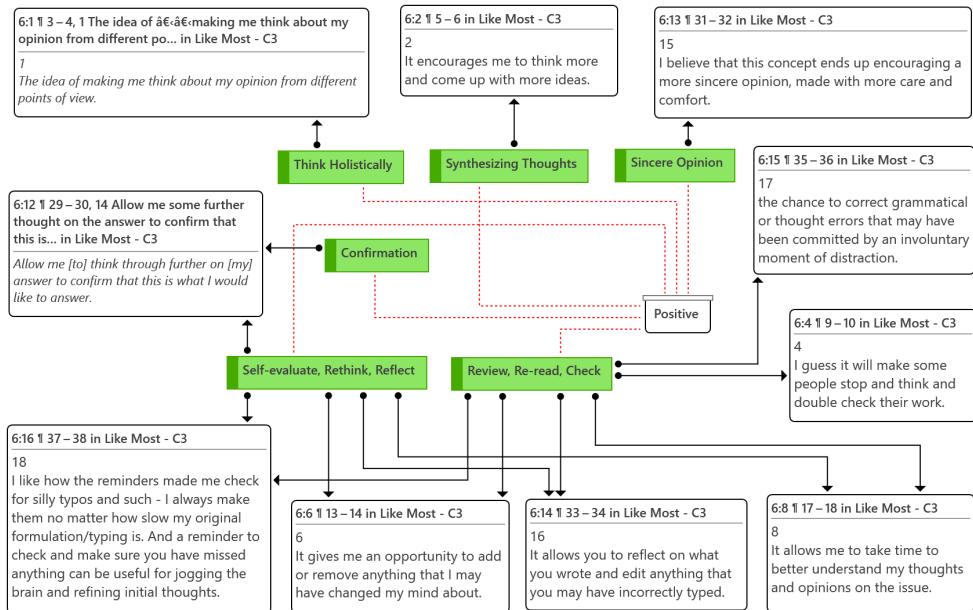


Fig. 15. Benefits: Comparison Time Nudge

Goldilocks Principle of 'Optimal' Reflection Time

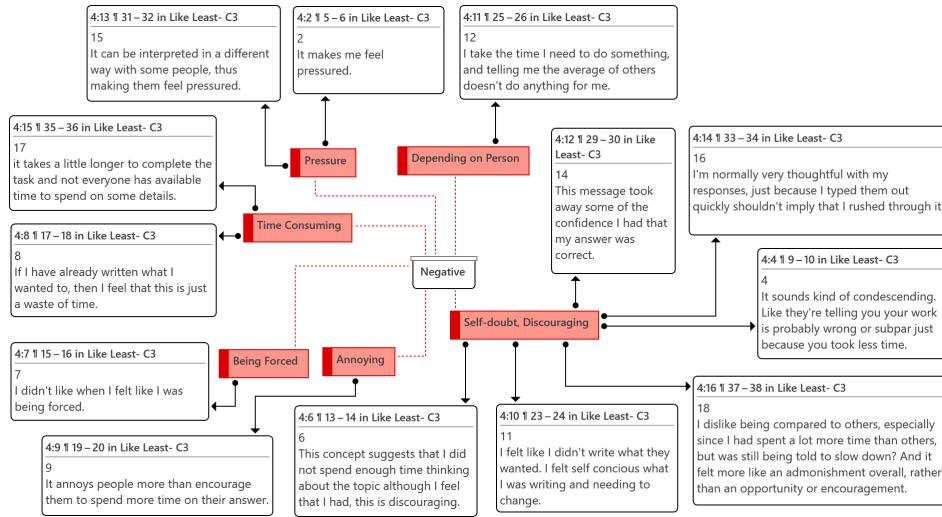


Fig. 16. Drawbacks: Comparison Time Nudge

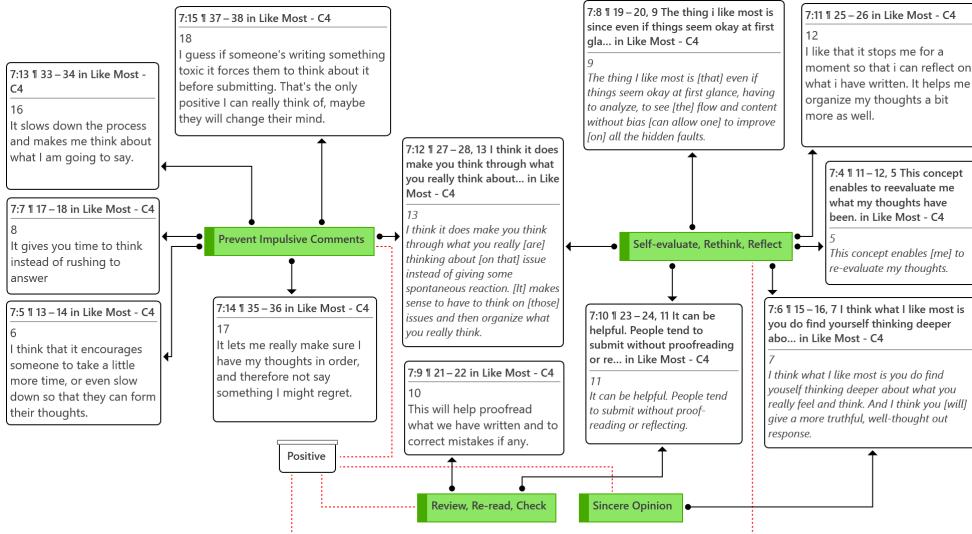


Fig. 17. Benefits: Alert Message

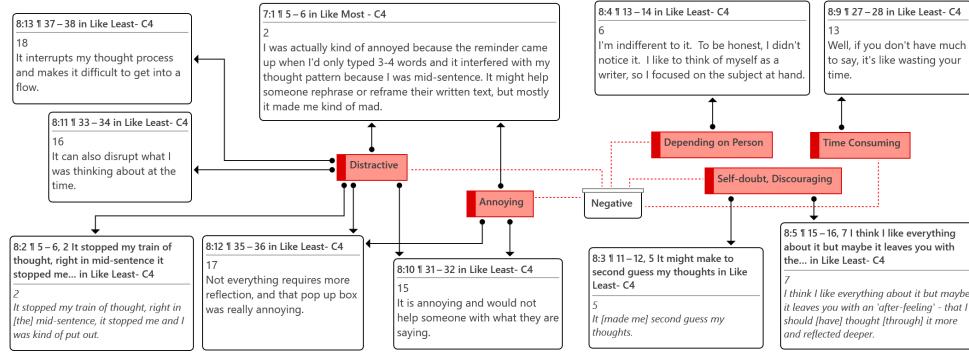


Fig. 18. Drawbacks: Alert Message