Speed of Response does not Affect Feelings of Rightness in Reasoning

by

Kaiden Stewart

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirements for the degree of

Master of Arts

in

Psychology

Waterloo, Ontario, Canada, 2019

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

It has been argued (Thompson, Prowse Turner & Pennycook, 2011) that the experience of *ease* (i.e., the ability to quickly generate an initial response) during processing influences one's likelihood of engaging reflectively with a problem. Thompson and colleagues argued that the ease with which an answer comes to mind (i.e., *answer fluency*) is a critical determinant of Feelings of Rightness (FOR), which, in turn, determine one's likelihood of reflecting. However, the possibility remained that the critical determinant of FORs is the speed of the total response, given the nature of the evidence for this claim. The critical difference between these two accounts is the contribution of factors occurring after the point where an answer comes to mind. Across two Experiments, we manipulated the duration of the physical response in order to identify whether participants' confidence (FOR) judgments are at least partially based on factors occurring after the initial mental generation of an answer. We found no evidence that FORs nor reflection are influenced by a manipulation of response execution. Broadly, the present investigation provides evidence that the relation between speed of response and FORs is likely due to the speed with which an answer is generated internally. That is, events occurring after the generation of a response, at least as operationalized here, do not influence FORs. This is consistent with Thompson and colleagues' (2011) suggestion that answer fluency is the critical variable in determining FORs.

Acknowledgements

I would like to thank my supervisors, Dr. Evan F. Risko and Dr. Jonathan Fugelsang, for their guidance and support, not only through the research presented herein but throughout my graduate study thus far. I would like to thank Dr. Derek Besner for his helpful feedback and comments.

I'd like to thank my colleagues in both the Cognition and Natural Behaviours (CaNB) Lab and the Reasoning and Decision-Making (RDM) Lab, but especially Daev McLean, Megan Kelly, Payal Patel, Torin Young, and Chandeep Mander for their assistance in collecting the data presented herein.

Table of Contents

List of Figures

Introduction

Dual-Process theories of reasoning suggest that, when making decisions, individuals produce two qualitatively distinct types of response. Type I responses are quick, and based on intuitive, heuristic processes, and sometimes individuals override these responses with a slower, reflective, deliberative, Type II response (de Neys, 2006; Evans, 2003, 2007, 2008; Evans & Stanovich, 2013). Though the common conceptualization of Type I processes as 'always bad' and Type II processes as 'always good' is fallacious (Evans & Stanovich, 2013; see Evans, 2007 for examples), Type II reasoning is thought to be necessary to engage in cognitive decoupling (i.e., abstracting the problem so as to reduce influence from real-world beliefs; Stanovich, 2011, 2012). Cognitive decoupling, in turn, is necessary to solve many of the abstract, representational problems studied in the reasoning domain (Evans, 2010). As such, much attention has been given to the characteristics of problems (e.g., Pennycook, Fugelsang, & Koehler, 2012; De Neys & Glumicic, 2008), and those who are solving them (e.g., Frederick, 2005; Littrell, Fugelsang, & Risko, 2019; Pennycook & Rand, 2019) that are associated with an increased tendency for Type II reasoning to be engaged.

Thompson, Prowse Turner, and Pennycook (2011) argued that, in addition to characteristics of problems and problem-solvers, characteristics of the reasoning process itself can drive the ultimate decision to reflect. Specifically, the experience of *ease* (i.e., the ability to quickly generate a response) during Type I processing is held to influence one's likelihood of engaging Type II processing. That is, Type I outputs that are generated easily will be associated with less reflection, and Type I outputs that are difficult to generate with more reflection. This framework for understanding Type II engagement is referred to as Metacognitive Reasoning Theory (MRT). Thompson and colleagues provided evidence for the core claim of this account using a two-response paradigm, wherein participants were initially required to assess the logicality of a syllogism, but did so under instruction to answer quickly, with the first response that came to mind. This was done to reduce the likelihood that the initial response was influenced by Type II processes. After the initial response, participants gave a Feeling-of-Rightness (FOR) judgment about how likely they thought it was that they initially generated the correct response. Participants then responded to the syllogism again, but this time were given unlimited time to try to generate the correct answer. Thompson and colleagues observed higher FORs for initial responses that were generated quickly, suggesting that responses accompanied by a feeling of ease are associated with greater confidence. Additionally, trials with greater FORs were associated with a decreased tendency to engage Type II processes, and as such they argued that the ultimate decision to reflect is at least partially driven by the ease, or fluency, experienced when generating a Type I response.

What is Fluency?

Cognitive processes vary in terms of the effort they require to complete. It has been argued that processes requiring little effort are accompanied by a generalized feeling of ease or goodness (Alter & Oppenheimer, 2009), which has been suggested to influence various judgments about a task, including confidence (Alter, Oppenheimer, Epley, & Eyre, 2007; Hertzog, Dunlosky, Robinson, & Kidder, 2003; Simmons & Nelson, 2006). *Processing fluency* is the broad name given to this subjective feeling of ease or goodness, and considerable work has gone into determining aspects of a task that are responsible for this experience (for a review, see Alter & Oppenheimer, 2009). Similarly, much work has gone into studying the consequences of subjectively experiencing fluency. Indeed, effects of facilitating or disrupting processing on confidence or other metacognitive judgments are relatively well-documented (e.g., Castel, McCabe, Roediger, 2007; Koriat, 1993; Simmons & Nelson, 2006).

Understanding what aspects of processing influence FORs is central to better understanding the tendency to reflect. In Thompson and colleagues' (2011) work, FORs were argued to arise from what they referred to as *answer fluency*. Answer fluency refers to the ease with which an individual can generate an answer; that is, answers that come to mind more easily are higher in answer fluency. This idea leaves out of the fluency experience, or at least the one influencing FORs, processes or events that occur after the generation of the answer (e.g., the execution of a response). In the present investigation we examine the potential influence of the latter on FORs and reflection.

The idea that the fluency experience might be affected by different stages of processing has been explored outside the reasoning context. Reber, Wurtz, and Zimmermann (2004) demonstrated in a word-identification task that both figure-ground contrast and font influence subjective judgments about fluency, despite hindering the perception of stimuli at different stages of processing. Though low figure-ground contrast hindered only the detection of stimuli and not their identification, and disfluent font hindered only the identification of stimuli and not their detection, both independently reduced subjective fluency judgments. This finding provided evidence that: 1) individuals are sensitive to the experience of ease at multiple stages of processing, and 2) that subjective judgments about ease are influenced by these experiences at different stages of processing.

The idea that post-answer processes/events might influence the experience of ease or difficulty also draws support from extant work in other domains. For example, Susser and Mulligan (2015; also Susser, Panitz, Buchin, & Mulligan, 2017) required participants in a memory task to respond to each word in a list by writing it down. This was done either with one's dominant or non-dominant hand. Writing words with one's non-dominant hand produced decreased Judgments of Learning (JOLs) for those words, despite showing no detriment in recall. This finding suggests that elements of one's response to a stimulus can act as a cue from which to make metacognitive judgments about the processing of that stimulus. Put another way, the actual physical execution of a response factors into a subjective feeling of difficulty.

In a similar vein, Undorf and Erdfelder (2011) provided evidence that individuals use time, in and of itself, as a metacognitive cue. In their studies, participants were asked to provide JOLs for another participant. However, in one condition, the only information participants had by which to make this judgment was the duration of the yoked participant's self-paced study of the items. Perhaps unsurprisingly, participants provided greater JOLs for short study periods, presumably inferring ease on the part of the participant to whom they were yoked. However, even when they had access to the items themselves (and thus, presumably, their difficulty), short study periods were still given greater JOLs than longer study periods. That is, participants use time *per se* as a metacognitive cue, even in the light of access to subjective difficulty information. Combining the results of Undorf and Erdfelder (2011) with those of Susser and Mulligan (2015; Susser et al., 2017), one can posit that 1) individuals infer ease or difficulty from the amount of time elapsed in a given task, and 2) elapsed time includes processes up until the completion of a response. Thus, it is possible that post-answer processes/events might influence FORs by virtue of the fact that they necessarily affect the time required to complete a response.

While the view that the relation between speed and FORs is a result of answer fluency is consistent with the data Thompson and colleagues reported, the nature of the evidence (i.e., a correlation between RT and FOR), of course, leaves open the possibility of post-answer processes or events influencing one's FOR. This is because an individual's Response Time, what Thompson and colleagues used as an index of answer fluency, necessarily includes both pre- and post-answer processes. Thus, it is entirely possible that what predicts FORs is not the speed that an answer comes to mind, but rather the speed with which a response is completed which would include post-answer processes (e.g., response execution). We test this idea here.

Experiment 1

In Experiment 1, we manipulated the duration of the *physical* response in order to identify whether participants' confidence (FOR) judgments are at least partially based on factors occurring after the initial *mental* generation of an answer. We utilized the two-response paradigm introduced by Thompson and colleagues (2011), but with the important difference that the execution of participants' responses was slowed on half of trials. If it is true that participants evaluate their outputs in terms of ease or speed at the moment an answer is generated, then any manipulation beyond that point will have no effect on FORs nor reflection. If, however, participants evaluate Type I outputs based on the total time it takes to respond, a delay of said response might produce effects on FORs and reflection similar to outputs that are difficult to generate. That is, slower responses should feel less right.

Method

Participants. Forty undergraduate students from the University of Waterloo participated in exchange for partial course credit. This sample size was pre-registered and based on an *a priori* power analysis and gave us 80% power to detect a medium-sized effect equivalent to d = .45 with a two-tailed, paired-samples t-test. This effect size was based on a previous version of the experiment, but it has since been determined that the results of said experiment were due to a methodological artifact conceptually unrelated to the present investigation. As such, the prior experiment and any comparisons with present experiment are not reported.

Design. There was only one independent variable, which had two levels (Condition: Fast vs. Slow). Each condition had its own block, which were counterbalanced across participants. The key dependent variables of interest were FORs, Response Initiation Time (time to begin a response to the first iteration of each syllogism by moving the mouse toward one of the choices;

used instead of actual response time because response time necessarily includes our manipulation on half of trials), Rethinking Time (time to respond to the second iteration of each problem), and Answer Change (whether the response to the second iteration of the problem differed from that of the first iteration).

Stimuli and Apparatus. Stimulus presentation and response recording was controlled by MATLAB 2014b using Psychophysics Toolbox 3.0.12 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007). The stimuli consisted of 48 individual syllogisms, which varied systematically in terms of their type (categorical vs. conditional), validity, and believability of their premises and conclusions. Syllogisms were assigned randomly to Condition. Participants selected their response to each syllogism using the mouse.

Procedure. On each trial, participants were required to respond to the stimulus twice. Each time they responded, they were required to select "yes" if the provided conclusion followed logically from the premises, and "no" if the provided conclusion did not follow logically from the premises. On the first iteration, it was emphasized that they respond as fast as possible, and on the second, it was emphasized that they take as long as they need to get the correct answer. After the first iteration, but before the second iteration, participants indicated on a seven-point scale how likely they thought it was that their initial response was correct (their FOR). Before each individual response, the mouse reset to the vertical and horizontal centre of the screen. Response options were equidistant (~450 pixels) to the right and left of the horizontal centre of the screen, but centred vertically. In one block of trials, the mouse was slowed during the first, speeded response. This increased the duration of response by approximately three seconds, which also resulted in an overall increase in Response Time. The mouse did not deviate from vertical centre of the screen and also could not travel past the response options. This functioned to maintain fidelity of RT data, such that we know that the recorded RT involved only the time it took to make the decision plus the time it took for the mouse to travel to the chosen option.

Prior to the test trials, participants were given detailed instructions about the task, including that the mouse would be slowed on some trials. They were instructed to evaluate the conclusions as if the premises were true, regardless of whether they believed the premises to be true in the real world. All participants completed two practice trials, on one of which the mouse was slowed during the first response.

Results

All data analysis and visualization was completed in the R programming language (R Core Team, 2016) with the assistance of add-on packages (see: apa (Gromer, 2017), BayesFactor (Morey & Rouder, 2018), dplyr (Wickham, François, Henry, & Müller, 2018), ez (Lawrence, 2016), lsr (Navarro, 2015), outliers (Komsta, 2011), plyr (Wickham, 2011), stats (R Core Team, 2017)). All analyses were carried out with an α of .05, and all t-tests were two-tailed. Bayesian analyses were carried out using a default prior of $\sqrt{2}/2$ (Morey & Rouder, 2018). Outliers were addressed at the individual and at the group level for all analyses involving response times by removing any data points with a Z-score > 3. This resulted in the removal of 51 trials (2.65%).

Preregistered Analyses[1]. Before analyzing our key dependent variables, we tested whether there were differences in the initiation of response as a function of Condition. There was no statistically significant difference in Response Initiation between fast (M = 7.10 s, SD = 1.85 s) and slow (M = 7.21 s, SD = 2.21 s) trials, $t(37) = 0.45$, $p = .655$, $d = 0.07$, 95% CI of the difference [-0.62, 0.39], $BF_{01} = 4.11$. This analysis was not preregistered but confirms that participants were not responding with different speeds based on the block they were in. Paired-samples t-tests were conducted to test the effect of the response execution manipulation on FOR and, independently, our measures of reflection (Answer Change and Rethinking Time). There was no statistically significant difference in FOR between fast (M = 5.12, SD = 0.82) and slow (M = 5.10, SD = 0.96) trials, $t(39) = 0.19$, $p = .853$, $d = 0.03$, 95% CI = [-0.20, 0.24], $BF_{01} = 4.28$ (Figure 1). There was also no statistically significant difference in Answer Change between fast (M = 14.59%, SD = 8.73%) and slow (M = 18.86%, SD = 13.67%) trials, $t(39) = 1.85$, $p = .072$, $d = 0.29$, 95% CI = [-8.95, 0.40], $BF_{01} = 1.30$ (Figure 2). Note that, in the latter case, the Bayesian analysis suggests no substantial evidence in favour of either hypothesis, though we have marginally more evidence for the null hypothesis than the alternative hypothesis. There was no statistically significant difference in Rethinking Time between fast (M = 11.85 s, SD = 6.58 s) and slow (M = 11.93 s, SD = 6.43 s) trials, $t(38) = 0.09$, $p = .931$, $d = 0.01$, 95% CI = [-1.93, 1.77], $BF_{01} = 4.25$ (Figure 3). The latter analysis was not preregistered.

In order to examine how well our data fit with the predictions made by Metacognitive Reasoning Theory, average within-subjects correlations were computed between Response Initiation and FOR. We decided to correlate FOR and Response Initiation, and not Response Time, because slowed trials (i.e., half of trials) have had an approximately three-second delay added to every trial, making any correlational analysis using these times difficult to interpret. The average within-subjects correlation between Response Initiation and FOR was negative and significantly different from zero (-.17), $t(39) = 4.06$, $p < .001$, $d = 0.64$, 95% CI = [-.25, -.08], $BF_{10} = 114.16$. Similarly, the average within-subjects correlation between FOR and Rethinking Time was also negative and significantly different from zero (-.44), $t(39) = 15.20$, $p < .001$, $d = 2.40$, 95% CI = [-.50, -.38], $BF_{10} = 9.48 \times 10^{14}$. Trials on which participants changed their answer (M = 3.86, SD = 0.83) had a lower mean FOR than did trials on which participants did not change their answer (M = 5.35, SD = 0.82), $t(39) = 12.69$, $p < .001$, $d = 2.01$, 95% CI [-1.72,

-1.25], $BF_{10} = 7.94 \times 10^8$. Trials on which answers changed (*M* = 20.06 s, *SD* = 9.75 s) also had a greater mean Rethinking Time than did trials where answers did not change (*M* = 11.37 s, *SD* = 7.36 s), $t(39) = 8.25$, $p < .001$, $d = 1.30$, 95% CI [6.56, 10.83], $BF_{10} = 810.99$.

Figure 1. FOR as a function of condition. Points indicate individual participant means. Horizontal lines indicate the group mean.

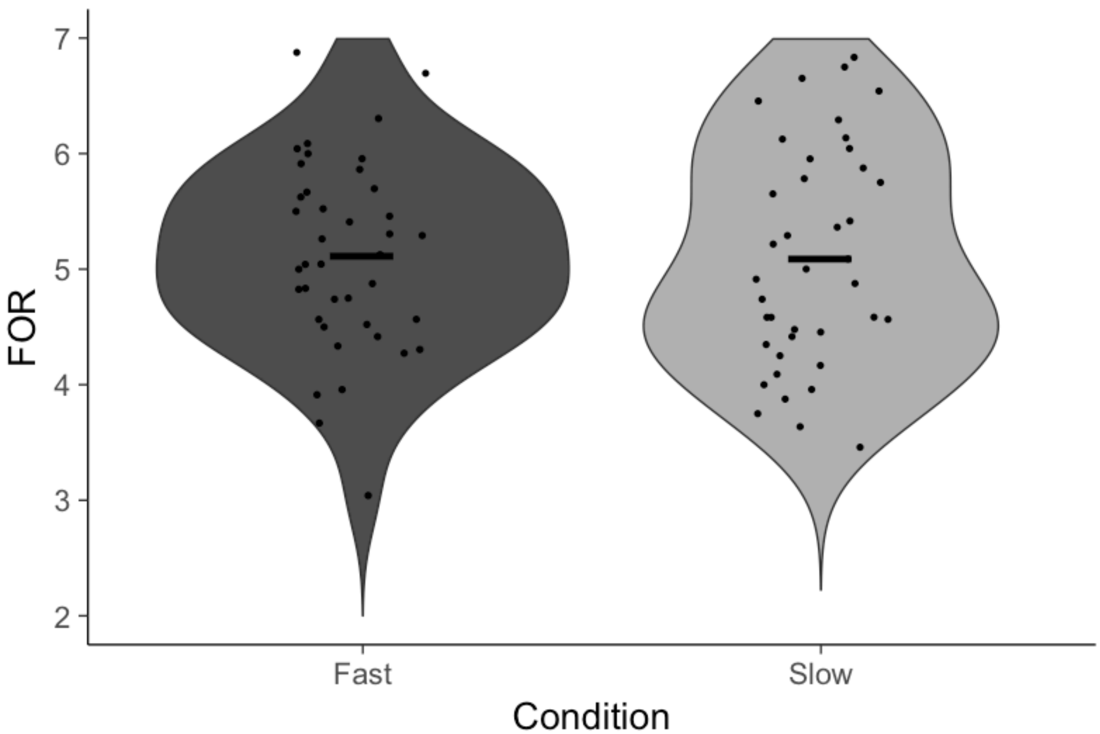


Figure 2. Answer Change as a function of condition. Points indicate individual participant means. Horizontal lines indicate the group mean.

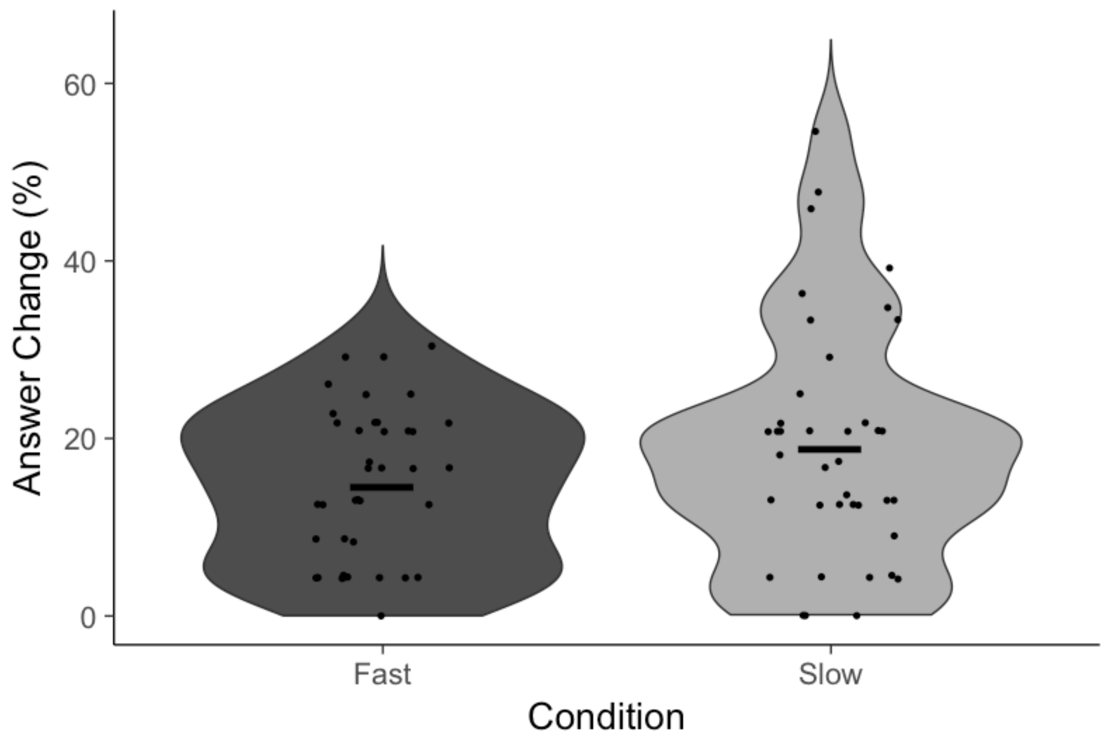
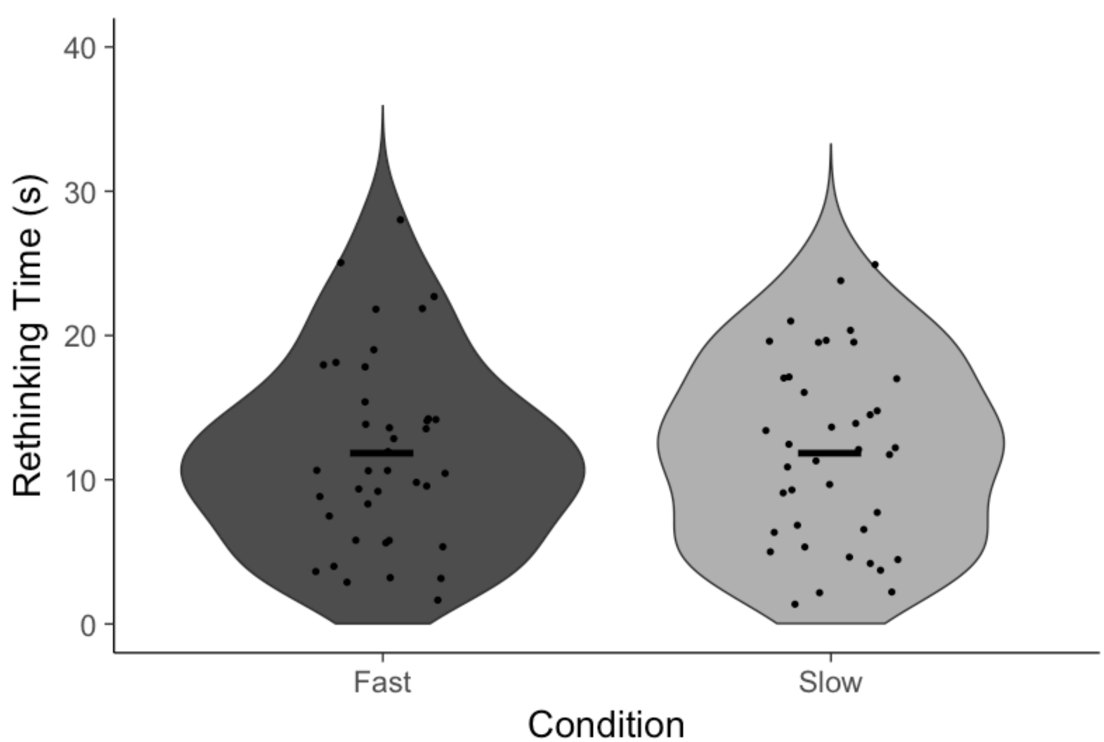


Figure 3. Rethinking Time as a function of condition. Points indicate individual participant means. Horizontal lines indicate the group mean.



Exploratory Analyses. Given our response execution manipulation was blocked and counterbalanced, we analyzed whether our measures of reflection differed across blocks. We found that participants' Rethinking Time was lower in Block 2 (M = 10.04 s, SD = 6.06 s) than in Block 1 (M = 13.74 s, SD = 6.39 s), $t(38) = 5.35$, $p < .001$, $d = 0.86$, 95% CI [2.30, 5.10], BF_{10} = 4,225. Additionally, participants were less likely to change their answer in Block 2 (M = 13.09%, SD = 11.55%) than in Block 1 (M = 20.36%, SD = 10.58%), $t(39) = 3.44$, $p = .001$, d = 0.54, 95% CI [3.00, 11.54], BF_{10} = 15.87. Correspondingly, mean FORs were higher in Block 2 (M = 5.24, SD = 0.97) than in Block 1 (M = 4.98, SD = 0.79), $t(39) = 2.52$, $p = .016$, $d = 0.40$, 95% CI [-0.46, -0.05], BF_{10} = 2.74. There were no effects of the order in which the blocks were seen on FORs (M_{FS} = 5.09, SD_{FS} = 0.81; M_{SF} = 5.12, SD_{SF} = 0.85), $t(38) = 0.09$, $p = .928$, d = 0.03, 95% CI [-0.51, 0.56], BF_{01} = 3.27, Answer Change (M_{FS} = 15.74%, SD_{FS} = 9.14%; M_{SF} = 17.82%, SD_{SF} = 8.56%), $t(38) = 0.74$, $p = .461$, $d = 0.24$, 95% CI [-3.58, 7.75], BF_{01} = 2.60, nor Rethinking Time, (M_{FS} = 10.95 s, SD_{FS} = 5.81 s, M_{SF} = 14.34 s, SD_{SF} = 8.59 s), $t(38) = 1.46$, p = .151, $d = 0.46$, 95% CI [-1.30, 8.08], BF_{01} = 1.40.

Participants were more accurate when the mouse was not slowed (M = 62.7%, SD = 12.5%) than when it was slowed (M = 56.7%, SD = 17.0%) during the initial response, $t(39)$ = 2.34, $p = .024$, $d = 0.37$, 95% CI [0.83, 11.22], BF_{01} = 1.06, but note that the Bayesian analysis is inconclusive. There was no difference in accuracy for the second response between fast (M = 63.1%, SD = 13.0%) and slow (M = 60.76%, SD = 16.34%) trials, $t(39) = 1.10$, $p = .278$, d = 0.17, 95% CI [-1.95, 6.61], BF_{01} = 3.47. As a robustness check, we analyzed our data in a similar manner to past investigations regarding syllogistic reasoning. We were specifically interested in the effects of all the characteristics of our syllogisms (Type, Validity, Premise Believability, and Conclusion Believability) on FORs and the effects of some of them (Premise and Conclusion

Believability) on endorsement rates and whether these effects interacted with our manipulation. Due to the limited number of observations per cell when all of these characteristics are included, we decided to conduct individual ANOVAs with each characteristic and Condition as independent variables. Any three-way interactions were thus not tested, as they would likely have too few observations per cell to be reliable. Participants' FORs were greater after a correct response than an incorrect response, $F(1, 39) = 9.97$, $p = .003$, $\eta_g^2 = .02$, $BF_{10} = 4.85$, and when the syllogism was valid as opposed to when it was invalid, $F(1, 39) = 19.19$, $p < .001$, $\eta_g^2 = .03$, $BF_{10} = 3{,}159$. They were also more confident for categorical as opposed to conditional syllogisms, $F(1, 39) = 9.71$, $p = .003$, $\eta_g^2 = .02$, $BF_{10} = 14.52$. FORs were greater when the premises of the syllogism were believable, $F(1, 39) = 9.33$, $p = .004$, $\eta_g^2 = .01$, $BF_{10} = 5.12$, but not when the conclusions were believable, $F(1, 39) = 1.37$, $p = .249$, $\eta_g^2 < .01$, $BF_{01} = 5.22$. There was a statistically significant Premise Believability by Conclusion Believability interaction, $F(1, 39) = 20.95$, $p < .001$, $\eta_g^2 = .02$, $BF_{10} = 176.44$, such that the effect on FOR was largest when both the premises and conclusions were believable. None of the characteristics of the syllogisms interacted with our manipulation, all Fs ≤ 1.36. In terms of endorsement rates (i.e., participants' likelihood of responding 'valid'), participants were more likely to endorse as valid syllogisms that had believable premises, $F(1, 39) = 6.54$, $p = .015$, $\eta_g^2 = .02$, $BF_{10} = 1.29$, or believable conclusions, $F(1, 39) = 57.76$, $p < .001$, $\eta_g^2 = .19$, $BF_{10} = 1.73 \times 10^{15}$, and there was a statistically significant Premise Believability by Conclusion Believability interaction, $F(1, 39) = 18.85$, $p < .001$, $\eta_g^2 = .05$, $BF_{10} = 348.73$, such that the effect of believability was stronger when both premise and conclusion were believable.

Discussion

The response execution manipulation had no effect on participants' FOR judgments, nor did it have any effect on either of the measures of reflection. Though there was a marginally significant effect on Answer Change, examination of Figure 2 confirms that this effect is driven by a small number of participants, outside of which there is, reasonably clearly, no effect. These results suggest that participants are not using the total duration of their response from which to infer a sense of the difficulty of the task nor the likelihood they are correct. More specifically, it appears as if Answer Fluency, as opposed to some feeling of fluency derived from the totality of a response, produces participants' FORs. Additionally, our data support the claims made by Thompson and colleagues (2011) that FOR guides reflection (on the basis of the negative correlation between FOR and Rethinking Time) and that FOR itself is determined by answer fluency (on the basis of the negative correlation between Response Initiation and FOR).

It is worth noting that in Experiment 1 we observed significant block effects suggesting, over the course of the experiment, our participants were spending considerably less time on task. Participants reflected less, were quicker to respond in the second block, and had increased FORs, regardless of which order the blocks were presented. These time-on-task effects are ostensibly consistent with MRT; over time, perhaps because the task becomes more familiar, participants report increased FORs, and these increased FORs are accompanied by a decrease in reflection. That there was no effect of order counterbalance on any of our key DVs suggests that the effects of experiment fatigue likely average out across counterbalances, but it is nevertheless possible that participants' altered psychological state, as evidenced by a change in performance from the first half of the experiment to the second, affected their experience of ease in the task so as to suppress a real and meaningful effect of our manipulation.

Experiment 2

In Experiment 2, we altered our experimental procedure so that the response execution manipulation was randomly intermixed, as opposed to blocked and counterbalanced. As such, participants did not know whether the mouse would be slowed on each individual trial until they began their initial response by moving the mouse from the centre of the screen.

Method

Participants. Forty undergraduate students from the University of Waterloo participated in exchange for partial course credit. The sample size was arrived at in an identical fashion as in Experiment 1 and allowed us 80% power to detect an effect of d = .45 with a two-tailed, paired-samples t-test.

Design. The experimental design was identical to that of Experiment 1.

Procedure. The procedure was exactly the same as in Experiment 1, with the exception that the response execution manipulation (i.e., the slowing of the mouse) was randomly intermixed within each subject at the trial level, rather than blocked and counterbalanced as in Experiment 1.

Results

Preregistered Analyses[2]. Forty-one trials (2.14%) were removed as outliers using the same criteria as in Experiment 1. There was, once again, no difference in Response Initiation Time between fast (M = 8.78 s, SD = 3.24 s) and slow (M = 8.61 s, SD = 3.10 s) trials, $t(39)$ = 0.78, p = .438, d = 0.12, 95% CI [-0.27, 0.61], BF_{01} = 4.20. Paired-samples t-tests were once again conducted to test the effect of the response execution manipulation on FOR and our measures of reflection. There was no statistically significant difference in FOR between fast (M = 4.57, SD = 0.78) and slow (M = 4.48, SD = 0.89) trials, $t(38)$ = 1.24, p = .222, d = 0.20, 95%

CI [-0.06, 0.25], BF_{01} = 3.83 (Figure 4). There was also no statistically significant difference in Answer Change between fast (*M* = 17.78%, *SD* = 10.74%) and slow (*M* = 16.85%, *SD* = 11.61%) trials, $t(38) = 0.50$, $p = .621$, $d = 0.08$, 95% CI [-2.83, 4.67], BF_{01} = 4.02 (Figure 5), nor in Rethinking Time between fast (*M* = 13.28 s, *SD* = 6.40 s) and slow (*M* = 13.01 s, *SD* = 6.40 s) trials, $t(39) = 0.83$, $p = .411$, $d = 0.13$, 95% CI [-0.38, 0.90], BF_{01} = 4.24 (Figure 6). The comparison of Rethinking Time was not preregistered.

Figure 4. FOR as a function of condition. Points indicate individual participant means. Horizontal lines indicate the group mean.

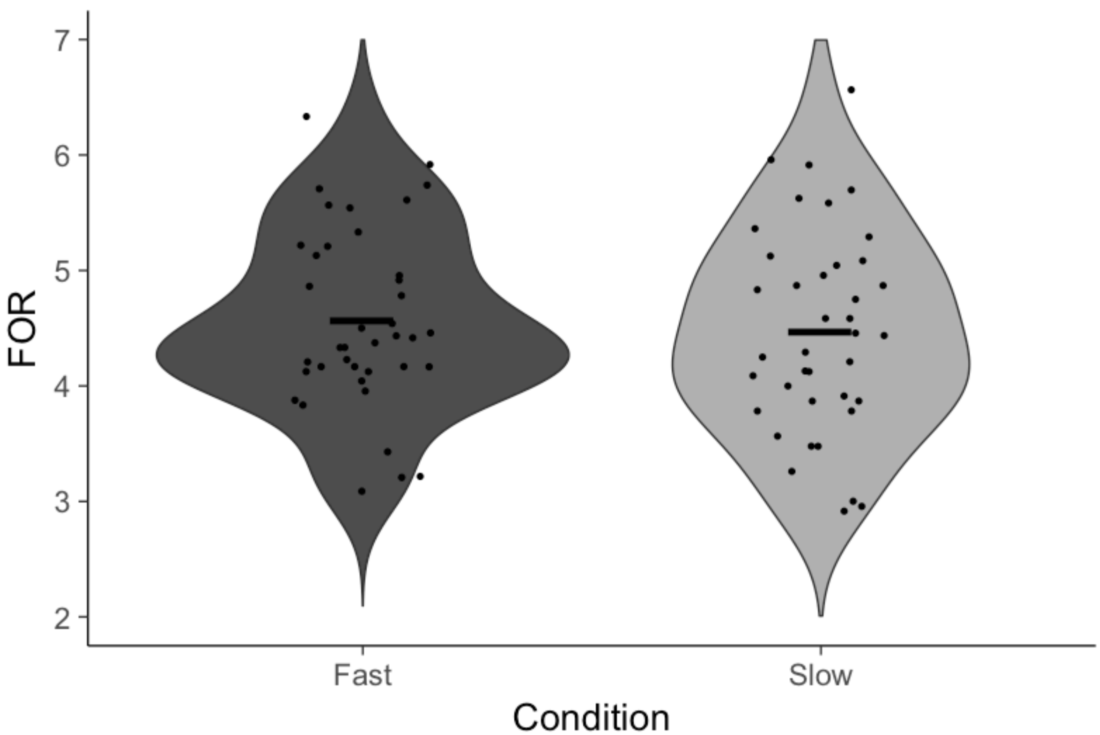


Figure 5. Answer Change as a function of condition. Points indicate individual participant means. Horizontal lines indicate the group mean.

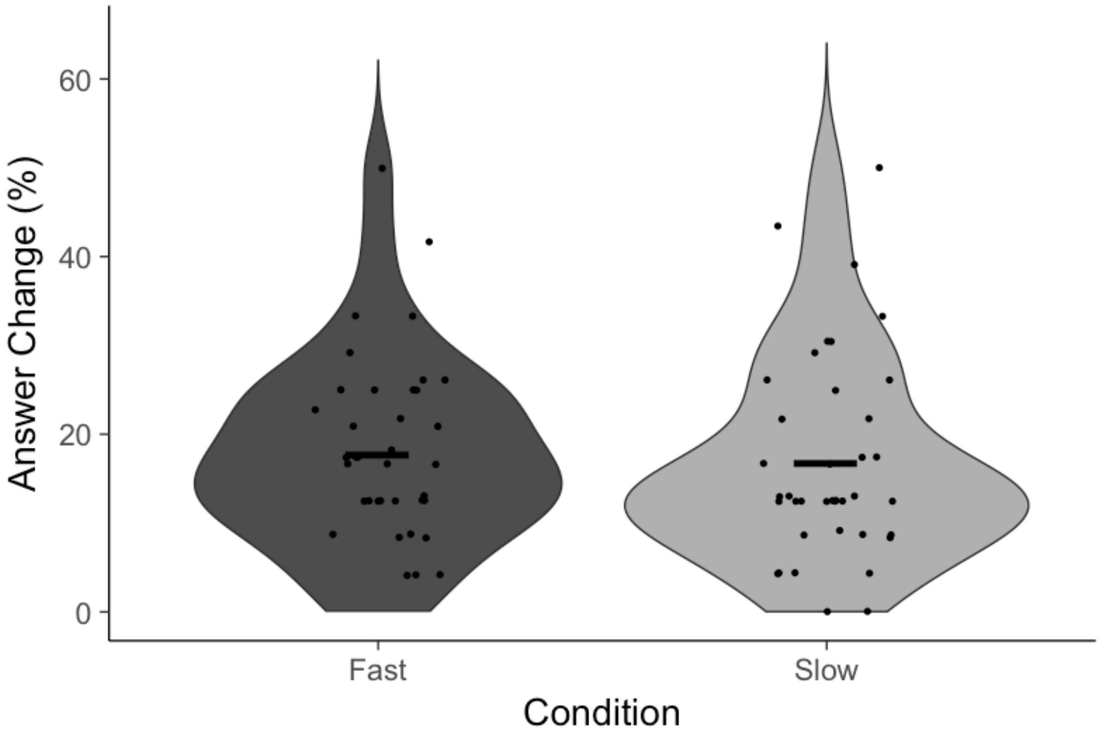
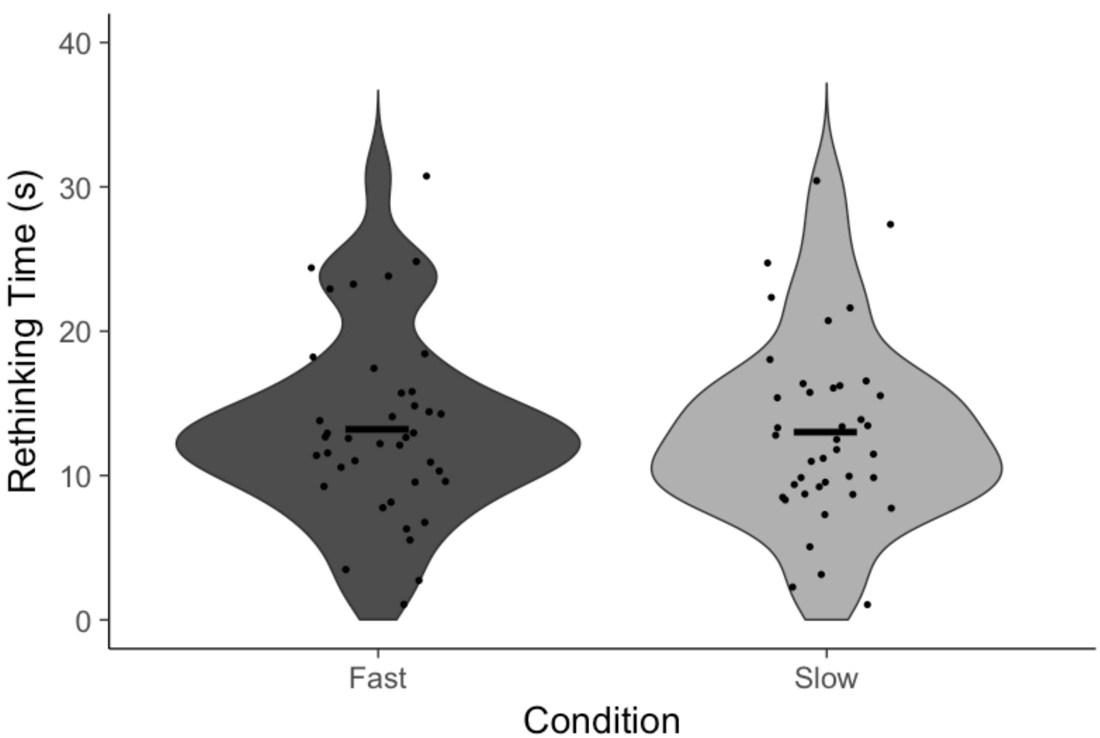


Figure 6. Rethinking Time as a function of condition. Points indicate individual participant means. Horizontal lines indicate the group mean.



Once again, average within-subjects correlations were computed between Response Initiation and FOR. The average within-subjects correlation between Response Initiation and FOR was negative and significantly different from zero (-.17), $t(39) = 4.63$, $p < .001$, $d = 0.73$, 95% CI = [-.25, -.10], $BF_{10} = 555.57$. Similarly, the average within-subjects correlation between FOR and Rethinking Time was also negative and significantly different from zero (-.42), $t(39) = 13.34$, $p < .001$, $d = 2.11$, 95% CI = [-.48, -.36], $BF_{10} = 1.50 \times 10^{13}$. Trials on which participants changed their answer ($M = 3.36$, $SD = 0.97$) had a lower mean FOR than did trials on which participants did not change their answer ($M = 4.64$, $SD = 0.85$), $t(38) = 9.69$, $p < .001$, $d = 1.55$, 95% CI [-1.55, -1.01], $BF_{10} = 1.37 \times 10^{12}$. Trials on which answers changed ($M = 18.80$ s, $SD = 7.81$ s) also had a greater mean Rethinking Time than did trials where answers did not change ($M = 12.21$ s, $SD = 6.00$ s), $t(38) = 10.61$, $p < .001$, $d = 1.70$, 95% CI [5.33, 7.85], $BF_{10} = 278.01$.

Exploratory Analyses. To mirror the block analysis from Experiment 1, we analyzed differences in FOR, Rethinking Time, and Answer Change between the first half of trials and the second half of trials. There was no statistically significant difference in FOR between the first half of trials ($M = 4.52$, $SD = 0.77$) and the second half ($M = 4.53$, $SD = 0.93$), $t(38) = 0.07$, $p = .945$, $d = 0.01$, 95% CI [-0.20, 0.19], $BF_{01} = 5.78$. There was also no statistically significant difference in Answer Change between the first half of trials ($M = 18.79\%$, $SD = 11.79\%$) and the second half ($M = 15.68\%$, $SD = 12.00\%$), $t(38) = 1.36$, $p = .182$, $d = 0.22$, 95% CI [-1.52, 7.74], $BF_{01} = 2.48$, but note that the Bayesian analysis provides no substantial evidence for the null or alternative hypothesis. There was, however, a statistically significant difference in Rethinking Time between the first half of trials ($M = 14.20$ s, $SD = 7.15$ s) and the second half ($M = 12.10$ s, $SD = 6.01$ s), $t(39) = 3.46$, $p = .001$, $d = 0.55$, 95% CI [0.87, 3.32], $BF_{10} = 23.56$.

Participants' accuracy was nearly identical in the Fast (M = 60.44%, SD = 13.40%) and Slow (M = 60.98%, SD = 15.31%) conditions for the first response, $t(39) = 0.18$, $p = .857$, d = 0.03, 95% CI [-6.61, 5.52], $BF_{01} = 4.25$, and the second response (M_F = 63.07%, SD_F = 14.24%; M_S = 63.62%, SD_S = 15.17%), $t(39) = 0.17$, $p = .864$, $d = 0.03$, 95% CI [-6.98, 5.89], BF_{01} = 4.25. As with Experiment 1, we also analyzed our data in a similar manner to past investigations of syllogistic reasoning (i.e., with each characteristic of the syllogisms being analyzed in an ANOVA). Participants' FORs were greater following a correct response than following an incorrect response, $F(1, 39) = 19.94$, $p < .001$, $\eta_g^2 = .03$, $BF_{10} = 385.46$, and when the syllogism was valid as opposed to when it was invalid, $F(1, 39) = 32.01$, $p < .001$, $\eta_g^2 = .04$, BF_{10} = 158,527. They were also more confident for categorical as opposed to conditional syllogisms, $F(1, 39) = 25.23$, $p < .001$, $\eta_g^2 = .03$, BF_{10} = 18,126. This was also true when the premises of the syllogism were believable, $F(1, 39) = 25.23$, $p < .001$, $\eta_g^2 = .03$, $BF_{10} = 1.27 \times 10^8$, and when the conclusions were believable, $F(1, 39) = 17.22$, $p < .001$, $\eta_g^2 = .02$, $BF_{10} = 491.75$. This time, there was no statistically significant Premise Believability by Conclusion Believability interaction, $F(1, 39) = 2.64$, $p = .112$, $\eta_g^2 < .01$, $BF_{01} = 4.48$. None of the characteristics of the syllogisms interacted with our manipulation, all Fs ≤ 1.62. In Experiment 2, participants were no more likely to endorse as valid syllogisms that had believable premises, $F(1, 39) = 0.56$, p = .459, $\eta_g^2 < .01$, $BF_{01} = 21.02$, but were more likely to endorse as valid those that had believable conclusions, $F(1, 39) = 72.54$, $p < .001$, $\eta_g^2 = .35$, $BF_{10} = 7.02 \times 10^{25}$, and there was no statistically significant Premise Believability by Conclusion Believability interaction, $F(1, 39)$ = 1.53, $p = .223$, $\eta_g^2 < .01$, $BF_{01} = 9.47$.

Finally, we compared all of our dependent variables of interest in Experiment 1 to Experiment 2. Though the lack of an effect of our manipulation was clearly constant across experiments, it is possible that participants' FORs, reflection, or just performance in general changes when the manipulation is intermixed as opposed to blocked, and it is possible that these changes are informative in and of themselves. There was no difference in accuracy for the first response between Experiment 1 (*M* = 59.74%, *SD* = 12.45%) and Experiment 2 (*M* = 60.73%, *SD* = 10.82%), *t*(78) = 0.38, *p* = .705, *d* = 0.08, 95% CI [-6.18, 4.20], BF_{01} = 4.04, nor was there a difference in accuracy for the second response between Experiment 1 (*M* = 61.94%, *SD* = 13.15%) and Experiment 2 (*M* = 63.38%, *SD* = 10.72%), *t*(78) = 0.53, *p* = .595, *d* = 0.12, 95% CI [-6.78, 3.91], BF_{01} = 3.80. There was, however, a statistically significant difference in Response Initiation Time between Experiment 1 (*M* = 7.16 s, *SD* = 1.89 s) and Experiment 2 (*M* = 8.70 s, *SD* = 3.10 s), *t*(76) = 2.64, *p* = .010, *d* = 0.60, 95% CI [-2.71, -0.38], BF_{10} = 4.49. There was also a statistically significant difference in FOR between Experiment 1 (*M* = 5.11, *SD* = 0.82) and Experiment 2 (*M* = 4.52, *SD* = 0.80), *t*(77) = 3.19, *p* = .002, *d* = 0.72, 95% CI [0.22, 0.95], BF_{10} = 16.55. There was, however, no statistically significant difference in Answer Change between Experiment 1 (*M* = 16.78%, *SD* = 8.81%) and Experiment 2 (*M* = 17.29%, *SD* = 9.55%), *t*(77) = 0.25, *p* = .805, *d* = 0.06, 95% CI [-4.63, 3.60], BF_{01} = 4.17, nor in Rethinking Time (M_1 = 11.91 s, SD_1 = 5.85 s; M_2 = 13.15 s, SD_2 = 6.32 s), *t*(77) = 0.91, *p* = .367, *d* = 0.20, 95% CI [-3.97, 1.49], BF_{01} = 3.00.

Discussion

In Experiment 2, we replicated the critical patterns of data from Experiment 1. Once again, the response execution manipulation had no effect on participants' FOR judgments, nor did it have any effect on either of the measures of reflection. This time, there was a clearer non-effect of our manipulation on Answer Change. Our results again replicate the key predictions made by MRT, specifically that speed of first response is negatively correlated with FORs, that

FORs are negatively correlated with Rethinking Time, and that changed answers are associated with decreased FORs. In addition to this, our data from both Experiments 1 and 2 generally fall in line with extant observations in the syllogistic reasoning literature. The results of Experiments 1 and 2 suggest that participants, in a two-response task, do not infer FORs from the duration with which they are able to complete a response. Importantly, the results of Experiment 2 also provide a replication of the results from Experiment 1, one benefit of which is to rule out the possibility that the results in Experiment 1 were due to the blocked nature of the response execution manipulation.

General Discussion

Across two experiments, we manipulated the speed of responding in a two-response reasoning task. The purpose of this manipulation was to identify whether processes occurring after the *mental* generation of an answer (but before the *physical* completion of the response) factor into metacognitive feelings about a response, and specifically FORs. It has been previously demonstrated (Thompson et al., 2011) that answers that come to mind more easily are associated with greater FORs and are subsequently associated with a decreased tendency to reflect. Critically, the operationalization of ease of responding here was the speed with which a response could be completed. From the Dual-Process perspective, easier Type I outputs are less likely to be reflected upon; more difficult Type I outputs prompt reflection. The response execution manipulation employed here differentiated two possible accounts: 1) that the relation between speed and FORs is due to the ease with which an answer comes to mind, and 2) that this relation is based on the totality of each response, including aspects of said response occurring after the point when an answer comes to mind. In the present investigation, we found no evidence that FORs nor reflection are influenced by a manipulation of response execution. In addition, we reproduced the key predictions of MRT; FORs were negatively correlated with both Response Time (as approximated here by Response Initiation Time) and Rethinking Time, and participants were less likely to change answers when higher FORs were given.

The results of the present investigation suggest that metacognitions in reasoning are not driven by inferences made about the length of time it takes to complete a response to a problem. More broadly, the present investigation provides evidence that the relation between speed of response and FORs is likely due to the speed with which an answer is generated. That is, events occurring after the generation of a response, at least as operationalized here, do not influence

FORs. This is consistent with Thompson and colleagues' (2011) suggestion that answer fluency is the critical variable in determining FORs.

Implications for Fluency Research

The evidence presented here is based on only one of many possible post-response production manipulations. The manipulation used herein impacted only the length of time with which a response could be completed. The observed lack of effect of the chosen manipulation precludes only a mechanism wherein elapsed time during a response is the critical factor. Perhaps a manipulation that makes responses more difficult (and is longer as a by-product of this difficulty) reduces FORs. In this vein, the results of the present investigation should not be taken as contradictory to those presented by Susser and Mulligan (2015; and Susser et al., 2017), who observed decreased JOLs for words written with one's nondominant hand. Their handedness manipulation slowed responding to the presented stimuli but was also inherently more difficult. In addition, it is of course possible that the aspects of processing that influence FORs and JOLs are different.

One possible difference between the work presented here and past work is the obvious irrelevance of the response execution manipulation. As noted by Alter and Oppenheimer (2009; and Oppenheimer, 2004), participants will in some contexts spontaneously discount manipulations they perceive as having no relevance to the task at hand. It is therefore possible that participants in the present experiments simply discounted the response execution manipulation as a source of fluency, ignoring any potential effects of said manipulation when responding with their FORs. The manipulation used in the present investigation was designed to affect only the time it took to complete a response, and a necessary tradeoff in using a manipulation so one-dimensional is a corresponding increase in its obviousness. It is difficult to imagine a manipulation that clearly does not impact the difficulty of the task but hides this fact from participants. Additionally, the characteristics of fluency manipulations that make them susceptible to spontaneous discounting are ill-defined. Though it is possibly the case that the manipulation used herein was discounted by participants, it is difficult to image an alternative that accomplishes the same goals but would not be. As such, it is possible that the response execution manipulation was simply discounted by participants, but this limitation is seemingly necessary for an investigation like the one here.

Implications for Dual-Process Research

It has been argued that the driving force behind the relation between speed and FORs is that quick responses are experienced as easier. This could be true in one of two ways. One possibility is that individuals infer ease from the speed with which they can bring an answer to mind. That is, individuals might, upon the production of a response, evaluate the amount of time it took to produce said response, and infer ease if the response was produced quickly. Alternatively, individuals might simply have access to the ease or difficulty of producing a response. In this case, the relation between time to response and FORs would simply be a by-product of the relation between speed and ease. That is, it would necessarily be true that easier responses are produced more quickly. It is worth noting that the present investigation cannot differentiate between these competing explanations, as our manipulation occurred necessarily after the point where an answer had come to mind. Rather, the above accounts can both equally explain the results herein and those of Thompson and colleagues (2011).

Irrespective of the exact monitoring mechanism, this facet of dual-process decision-making has important implications. In MRT, the experience of difficulty is an important determinant of whether individuals will choose to engage their more reflective, Type II processes, and as such, manipulations that induce feelings of difficulty have the potential to increase reflectiveness, and downstream, improve the quality of decisions. The present investigation provides evidence suggesting that successful manipulations of the fluency experience in reasoning will likely have to operate in the time before the generation of a Type I output, as this is likely the point when a feeling of ease or difficulty arises.

Through exploratory analyses, we found two novel patterns of data that are interesting vis-à-vis MRT. First, when the response execution manipulation was blocked, the second block was associated with increased FORs and decreased reflection, consistent with the predictions of MRT. However, when the manipulation was intermixed, there was no difference in FORs between the first half of trials and the second, but there was decreased reflection for at least one of our measures of reflection (Rethinking Time; Bayesian analyses were unable to conclusively rule out an effect on Answer Change). Second, (arguably) by virtue of intermixing the response execution manipulation in Experiment 2, FORs decreased without any effect on either measure of reflection. That is, participants felt less right on average, but without a commensurate increase in reflection (though the means were higher in the intermixed condition). These results provide evidence that the nature of metacognitive experience can potentially change both over time and with the nature of the manipulation (i.e., whether it is mixed or blocked). Additionally, the predicted relations between FORs and reflection are, in the data presented here, inconsistent when looking at aggregate changes; an aggregate change in FOR is not necessarily predictive of an opposite aggregate change in reflection. Put another way, while maintaining the individual correlations between FORs and reflection, MRT seems not to capture aggregate changes. Although Thompson and colleagues primarily conceptualize MRT within an individual's metacognitive experiences (i.e., trials on which that individual experiences ease will be met with relatively little reflection for that individual), it is perhaps more accurate to conceptualize MRT at only the individual level. This is consistent with the view that the crucial feature of the experience of ease is that it changes relative to past experience (Hansen & Wänke, 2013; Wänke & Hansen, 2015). That is, fluency is a relative more so than an absolute phenomenon; ease relative to expectations or recent past states influences metacognitive judgments. In this case, FORs are likely a product of experienced ease relative to previously experienced ease, and thus should probably be held to be predictive of reflection at the trial level, and not be expected, when subject to aggregate changes, to produce aggregate changes in reflection. This insight might have important implications for attempts to use introduced disfluencies as a means to increase reflection.

Conclusion

The present investigation represents a first, basic step toward understanding the specific fluency mechanisms behind the operation of metacognitive processes within a MRT framework. Future research can further investigate the metacognitive mechanisms within Dual-Process reasoning and answer some of the questions left unanswered by the work presented here.

Footnotes

1. Link to preregistration: https://osf.io/wxhtn
2. Link to preregistration: https://osf.io/7t2fa

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review*, *13*(3), 219-235.

Alter, A. L., Oppenheimer, D. M., Epley, N., & Eyre, R. N. (2007). Overcoming intuition: metacognitive difficulty activates analytic reasoning. *Journal of Experimental Psychology: General*, *136*(4), 569.

Castel, A. D., McCabe, D. P., & Roediger, H. L. (2007). Illusions of competence and overestimation of associative memory for identical items: Evidence from judgments of learning. *Psychonomic Bulletin & Review*, *14*(1), 107-111.

De Neys, W. (2006). Dual processing in reasoning: Two systems but one reasoner. *Psychological science, 17*(5), 428-433.

De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, *106*(3), 1248-1299.

Evans, J. S. B. (2003). In two minds: dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454-459.

Evans, J. S. B. (2007). *Hypothetical thinking: Dual processes in reasoning and judgement*. Psychology Press.

Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, *59*, 255-278.

Evans, J. S. B. (2010). Intuition and reasoning: A dual-process perspective. *Psychological Inquiry*, *21*(4), 313-326.

Evans, J. S. B., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives on Psychological Science*, *8*(3), 223-241.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25-42.

Gromer, D. (2017). apa: Format Outputs of Statistical Tests According to APA Guidelines. R package version 0.2.0. https://CRAN.R-project.org/package=apa

Hansen, J., & Wänke, M. (2013). Fluency in context: Discrepancy makes processing experiences informative. In *The Experience of Thinking* (pp. 80-94). Psychology Press

Hertzog, C., Dunlosky, J., Robinson, A. E., & Kidder, D. P. (2003). Encoding fluency is a cue used for judgments about learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(1), 22.

Komsta, L. (2011). outliers: Tests for outliers. R package version 0.14. https://CRAN.R-project.org/package=outliers

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, *100*(4), 609.

Lawrence, M. A. (2016). ez: Easy Analysis and Visualization of Factorial Experiments. R package version 4.4-0. https://CRAN.R-project.org/package=ez

Littrell, S., Fugelsang, J., & Risko, E. F. (2019). Overconfidently underthinking: narcissism negatively predicts cognitive reflection. *Thinking & Reasoning*, 1-29.

Morey, R. D., & Rouder, J. N., (2018). BayesFactor: Computation of Bayes Factors for Common Designs. R package version 0.9.12-4.2. https://CRAN.R-project.org/package=BayesFactor

Navarro, D. J. (2015) Learning statistics with R: A tutorial for psychology students and other beginners. (Version 0.5) University of Adelaide. Adelaide, Australia

Oppenheimer, D. M. (2004). Spontaneous discounting of availability in frequency judgment tasks. *Psychological Science*, *15*(2), 100-105.

Pennycook, G., Fugelsang, J. A., & Koehler, D. J. (2012). Are we good at detecting conflict during reasoning?. *Cognition*, *124*(1), 101-106.

Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39-50.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reber, R., Wurtz, P., & Zimmermann, T. D. (2004). Exploring "fringe" consciousness: The subjective experience of perceptual fluency and its objective bases. *Consciousness and Cognition*, *13*(1), 47-60.

Simmons, J. P., & Nelson, L. D. (2006). Intuitive confidence: Choosing between intuitive and nonintuitive alternatives. *Journal of Experimental Psychology: General*, *135*(3), 409.

Stanovich, K. (2011). *Rationality and the reflective mind*. Oxford University Press.

Stanovich, K. E. (2012). On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. *The Oxford handbook of thinking and reasoning*, 343-365.

Susser, J. A., & Mulligan, N. W. (2015). The effect of motoric fluency on metamemory. *Psychonomic Bulletin & Review*, *22*(4), 1014-1019.

Susser, J. A., Panitz, J., Buchin, Z., & Mulligan, N. W. (2017). The motoric fluency effect on metamemory. *Journal of Memory and Language*, *95*, 116-123.

Thompson, V. A., Turner, J. A. P., & Pennycook, G. (2011). Intuition, reason, and metacognition. *Cognitive Psychology*, *63*(3), 107-140.

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

Wänke, M., & Hansen, J. (2015). Relative processing fluency. *Current Directions in Psychological Science*, *24*(3), 195-199.

Wickham, H. (2011). The Split-Apply-Combine Strategy for Data Analysis. *Journal of Statistical Software, 40*(1), 1-29. URL http://www.jstatsoft.org/v40/i01/.

Wickham, H., François, R., Henry, L., & Müller, K. (2018). dplyr: A Grammar of Data Manipulation. R package version 0.7.5. https://CRAN.R-project.org/package=dplyr