

Using metacognitive prompts to explore student reasoning trajectories

Em Sowles, Drew J. Rosen, and MacKenzie R. Stetzer

*Department of Physics and Astronomy and Maine Center for Research in STEM Education, University of Maine,
Orono, ME 04469*

Previous research has shown that students who demonstrate sufficient skills and conceptual understanding to reason productively may perform inconsistently on analogous questions. Such inconsistencies can be explained via dual process theories of reasoning (DPToR). To gain insight into students' reasoning trajectories, we developed an exploratory sequence of DPToR-aligned metacognitive prompts and administered the sequence immediately after students answered a physics question containing salient distracting features. The metacognitive prompts asked students to: describe their first ideas, reflect on any doubts they had with respect to those ideas, compare their first ideas with their submitted responses, and characterize their reasoning approaches. In this paper, we describe how we use student responses to these prompts along with timing data to investigate students' reasoning trajectories. Students who self-reported that they revised their thinking before submitting an answer spent significantly longer answering the question than those who did not. In addition, students who retained a correct provisional response reported fewer doubts and the use of a process-first approach, whereas students who retained an incorrect provisional response reported more doubts and the use of an answer-first approach. We anticipate that a more detailed understanding of students' reasoning trajectories arising from investigations like the one reported here will be an important step in the development of effective, research-based instructional materials that better support student reasoning in physics.

I. INTRODUCTION

Previous research has established that students who demonstrate sufficient conceptual knowledge and skills on one physics question often perform inconsistently on an analogous question on the same topic [1-3]. These inconsistencies persist even after research-based instruction. Researchers are increasingly attributing such reasoning inconsistencies to the nature of human reasoning itself and explain the observed reasoning patterns via dual-process theories of reasoning (DPToR) [1-3], which posit the existence of two distinct modes of cognitive processing [4,5].

According to DPToR, when a student encounters a physics question, fast, automatic process 1 (or heuristic process) will generate a provisional mental model informed by prior knowledge, contextual cues, goals, and beliefs. The slow, analytical process 2 may or may not be engaged to evaluate the provisional model and ascertain whether or not it is satisfactory. If process 2 is not engaged, the provisional model becomes the response without further reflection and analysis [6]. If process 2 is engaged, the provisional model is evaluated (although reasoning biases may impact the quality of analysis) and if the model is accepted, the process is complete, and a final response is given. However, if the provisional model is deemed unsatisfactory, the cycle is repeated (with a new provisional model each time) until a model is accepted and yields a final response [2,5,7]. The likelihood of process 2 engagement is related to a student's cognitive reflection skills, which refer to the tendency and ability to scrutinize provisional models [8,9]. The nature of this reasoning cycle, with an emphasis on specific reasoning paths and the associated reasoning hazards, has been highlighted and represented graphically in a recent article [7].

To date, however, relatively little work has focused on investigating and documenting the reasoning pathways of students as they answer physics questions [1,10]. Written student explanations to physics questions typically do not describe the actual reasoning paths students traversed to arrive at their answers. Instructors often hope that students will review the question, select a suitable approach, and systematically apply physics principles to arrive at a solution. Research, however, suggests that students may in fact select an answer first, often cued by salient distracting features (or SDFs) of the question/context, and then construct an explanation in support of that answer [2]. As part of an ongoing effort to leverage DPToR to develop research-based materials that better support student reasoning in physics, we aim to identify, document, and characterize, to the extent possible, students' reasoning trajectories as they move from provisional model to final response.

To gain more insight into students' reasoning trajectories, we administered to introductory physics students a physics question on rigid-body dynamics immediately followed by a sequence of metacognitive prompts about students' reasoning processes and an opportunity to revisit the physics question. Using student responses to the prompts, we were able to infer their provisional responses to the question and compare them to their submitted and post-metacognitive-sequence responses. In addition, we collected question response timing data and use them to more fully document identified trajectories. In this paper, we primarily focus on characterizing the student reasoning paths that led to their submitted responses, although we briefly discuss the effectiveness of the metacognitive prompt sequence as an intervention that could potentially improve students' performance on the physics task by promoting reflection on their reasoning.

II. METHODS

In this section, we provide an overview of our research task design and general methods for data analysis. Since multiple analyses were performed, analysis-specific methods are discussed with the associated analysis in Section III.

Core to this investigation was the selection of a challenging physics question that students had not seen before and that was likely to trigger intuitively appealing incorrect responses due to salient distracting features. Our research task was built around a question on rigid-body dynamics (FIG. 1) adapted from *Tutorials in Introductory Physics* [11] and the accompanying research [12]. To answer the question correctly, students needed to recognize that, since the rods' masses are equal and the magnitudes of the applied forces (and net forces) are equal, the magnitudes of the accelerations of the centers of mass of both rods must be equal at the instant shown via

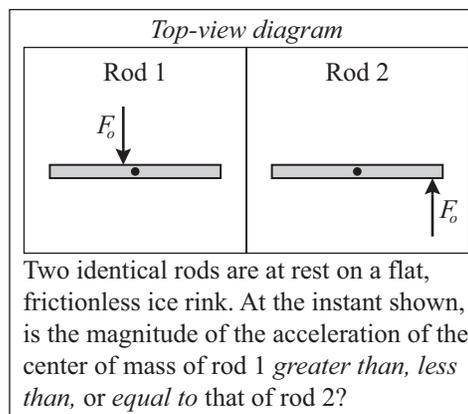


FIG. 1. Physics question on rigid-body dynamics.

Newton’s second law. Research has shown that students struggle with such questions and often focus on the proximity of the force to the center of mass (which effectively serves as an SDF) when comparing the (linear) accelerations [12].

A sequence of metacognitive prompts (Table I) was presented immediately after the target question and these DPToR-aligned questions probed students’ reasoning pathways leading to their submitted responses to the physics question. Students were asked to describe their first ideas, to reflect on any doubts they had with respect to those ideas, and to compare their first ideas with their submitted responses. These three questions provided insight into their reasoning trajectories and allowed us to infer (in most cases) their provisional responses (which may or may not differ from their submitted responses). Another prompt asked students to characterize their reasoning pathways as being either a process-first or an answer-first approach.

After the metacognitive prompts, students were given an opportunity to revisit the original physics question and change their responses if they wished.

The complete research task was administered as part of an online participation-based homework assignment given to students enrolled in calculus-based introductory mechanics after all relevant instruction on rigid-body dynamics. These weekly homework assignments were administered via Qualtrics [13].

Both multiple-choice data and free-response data were collected as part of this investigation. For the physics question, a correct response required a multiple-choice answer of ‘equal’ acceleration along with correct and sufficient supportive reasoning; all other responses (including, for example, a multiple-choice answer of ‘equal’ supported by insufficient or incorrect reasoning) were considered incorrect. Explanations were coded as correct and sufficient if they (1) discussed Newton’s second law, (2) rejected torque or point of application arguments, or (3) asserted that rotational and linear motions are independent. While most of the metacognitive prompts were multiple choice, the first prompt (probing for first ideas) was free response. The codes for these responses were developed through an iterative process of reading responses, developing categories, and validating these categories through group discussion. Consensus coding was performed during the development of the codebook to ensure reliability [14]. Once codes for the first ideas were completed, we were able to use those responses in conjunction with the change from first idea metacognitive responses in order to construct students’ provisional responses as described in Section III.A.

Additional data were also collected to allow for further characterization and triangulation of students’ reasoning trajectories. In particular, the time each student spent answering (but not explaining) the physics

TABLE I. Metacognitive sequence prompts.

1.	When you first saw this question, what was the first idea or thought that came to mind, regardless of correctness? (FR)
2.	At any point before selecting your answer, was there anything that made you doubt your first idea? (MC)
3.	Did your final answer differ from your first idea that you stated above? (MC)
4.	When you were answering the original question, which of the following best describes the approach you took? (MC)
a.	I started with an intuitive answer or gut feeling for which answer was correct, and then I used physics arguments to validate my choice
b.	I started with an idea of the physics concepts or approaches I needed to draw upon, then used them to arrive at a result, and finally checked to see which answer matched the result I obtained.

question was recorded (*i.e.*, the time between loading the physics question page and submission of a multiple-choice answer). Due to the wide variety of data collected and analyzed in this investigation, several different statistical analyses were performed and are described in Section III.

The intuitive first idea (or provisional model) is an elusive but vital element in the reasoning trajectory and our prompts were designed to draw out whatever students considered to be their first ideas. We recognize, however, that there is a limit to what (if anything) students can recall about their first, subconscious ideas. The construction of students’ provisional responses is also subject to limitations due to its reliance on self-reported accounts of first ideas and their differences from submitted responses. Given the importance of better understanding students’ reasoning trajectories, this exploratory study attempts to triangulate the metacognitive prompt data with other (non-self-reported) data when possible to mitigate such limitation and strengthen our claims.

III. RESULTS & DISCUSSION

In this section, we construct and analyze patterns in student response trajectories using the collected data.

A. Construction of student response trajectories and examination of intervention effectiveness

Student reasoning trajectories were assembled by characterizing the correctness or incorrectness of three different student responses: the *provisional response* (inferred by researchers), the *submitted response*, and the *post-metacognitive-sequence response*. Only students who had responses for all three were included in our analysis (N=202). Students’ *submitted responses* and *post-metacognitive-sequence responses* were evaluated directly on the basis of their responses given during the research task. *Provisional responses* were

constructed by first examining the *change from first idea* prompt responses. If a student reported no change, we assigned their *submitted response* to a preliminary *provisional response* code. Next, the *first idea* prompt responses were individually evaluated, and *provisional response* codes were assigned for all students. Finally, the latter codes were checked for consistency with any preliminary *provisional response* codes (if present). If there was a lack of consistency, then code associated with the more coherent response was taken as the final *provisional response* code for analysis. For example, if a student provided a very clear *first idea* response aligned with their *submitted response*, but also reported *changing from their first response*, we assigned a final *provisional response* code that corresponded to their *submitted response*. Only 6 inconsistencies were identified, and they were resolved via discussions with the research team.

Using these data, we can visualize the flow of students' responses (correct and incorrect) from provisional to submitted to post-metacognitive sequence using a Sankey diagram (FIG. 2) [15]. From provisional to submitted, some students demonstrated consistently correct or consistently incorrect responses and thinking. However, there is evidence of students switching both away from and to correct responses, consistent with DPToR. We note that there appears to have been more switching between provisional and submitted responses than between submitted and post-metacognitive-sequence responses.

Effectiveness of metacognitive prompt sequence as intervention. Although not the primary focus of this investigation, we wanted to see if the metacognitive prompt sequence alone could help students reflect on their own reasoning and possibly revise their thinking. Before the metacognitive prompt sequence, 41.1% of students' submitted responses were correct with correct reasoning, which is consistent with previous research [12]. After the metacognitive intervention, 44.4% of students' responses were correct with correct reasoning. We compared the submitted and post-metacognitive-

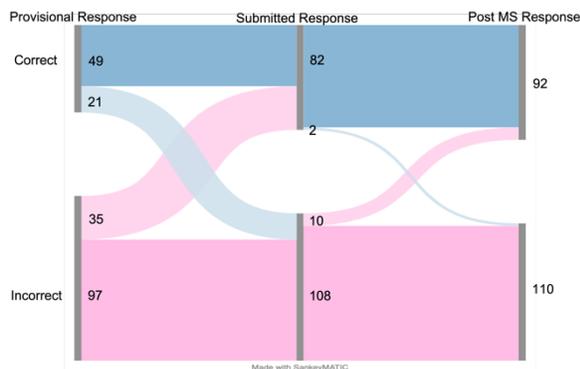


FIG. 2. Sankey diagram of response trajectories [15].

sequence responses using the McNemar test (to probe student shifting) and the binomial test (to probe for differences in overall response distribution) and found that there were no statistically significant differences as a result of the metacognitive prompt sequence ($p=.092$, $p=.181$, respectively). While it is perhaps unsurprising that such a relatively “soft” intervention (prompting reflection on one’s own reasoning trajectory) didn’t yield significant changes, our primary goal was to explore and characterize students’ reasoning trajectories from provisional to submitted responses. For this reason, in the analysis that follows, we focus exclusively on the provisional-to-submitted reasoning trajectories, leveraging timing data and metacognitive response data to provide more insight.

B. Response time analysis

As shown in the Sankey diagram (Fig. 2), although the majority of the students remained consistently correct or incorrect between provisional and submitted responses, many students ($N=56$) were identified to have shifted their thinking between provisional and submitted responses, where the provisional responses were inferred by researchers based on responses to metacognitive prompts 1 and 3 as described in Section III.A. As a result, we wanted to examine the timing data (to give a submitted answer) for these different paths between provisional and submitted responses, with a particular focus on students who changed their responses versus those who didn’t. In preparation for this analysis, we controlled for outlier data by performing a very small amount of symmetric time trimming, adapted from Field [16], effectively removing 6 students and leading to the final sample of 202 students shown in FIG. 2.

In Table II, we report the mean response times for students change their responses from provisional to submitted (left column) and for those who didn’t (right column). A paired sample t-test indicated that those students who self-reported that they changed responses took a statistically significantly ($t(201)=8.305$, $p<.001$) longer time to give a submitted answer than those who didn’t, with a medium-large effect size ($d= .584$). This suggests that, on average, students who changed their responses undertook more extended periods of effortful analysis via process 2, whereas students who did not change their responses (either correct or incorrect) may have had more automatized (process 1) responses with less analysis and reflection via process 2; such results

TABLE II. Provisional to submitted mean response times with standard errors.

	Changed Response N=56	Unchanged Response N=146
Mean submitted response time	105.7 ± 23.3 s	64.2 ± 8.7 s

are consistent with other response time studies [10]. Finally, our response time analysis is consistent with and serves to validate students' self-reported switching.

C. Metacognitive response analysis

In this section, we focus on our analysis of student responses to the two metacognitive prompts that were not used to infer students' provisional responses: doubts about first ideas (prompt 2) and generalized approach (prompt 4). (See Table I for prompts).

1. Doubts about first ideas (prompt 2).

Responses to the metacognitive prompt exploring student doubts related to their first ideas are shown for all four trajectories in Table III. A Pearson's chi-squared test indicated that there is a statistically significant difference with a medium effect size ($\chi^2(3)=22.92$, $p<.001$, $V=.337$) in student doubt response distributions across the four trajectories. By analyzing the standardized residuals, we found that students who remained correct from provisional to submitted reported significantly fewer doubts than any other population. These results suggest that students who were always correct may have automatized that correct response (or approach) and have confidence in their response, while those who remain incorrect experience considerable doubts about their initial ideas. While the latter students may have had doubts about their initial ideas, they either did not engage in reflection (consistent with the relative quickness of their responses – see Section III.B) or else they failed to sustain a productive engagement of the analytic process needed to arrive at a different response, despite their doubts.

2. Reasoning approach (prompt 4)

Responses to the metacognitive prompt exploring general reasoning approaches are shown for all four reasoning trajectories in Table IV. Students selected either an *answer-first* approach (option a) or a *process-first* approach (option b). A Pearson's chi-squared test indicated that there is a statistically significant difference with a medium effect size ($\chi^2(3)=18.32$, $p<.001$, $V=.301$) in student approach distributions across the four trajectories. Based on an analysis of the standardized residuals, students who remained correct

TABLE III. Metacognitive doubt (prompt 2) responses: provisional to submitted.

Doubts Response	Correct to Correct N=49	Correct to Incorrect N=21	Incorrect to Correct N=35	Incorrect to Incorrect N=97
Yes	38.8%	76.2%	85.7%	68.0%
No	61.2%	23.8%	14.3%	32.0%

TABLE IV. Metacognitive approach (prompt 4) responses: provisional to submitted.

Approach	Correct to Correct N=49	Correct to Incorrect N=21	Incorrect to Correct N=35	Incorrect to Incorrect N=97
Answer first (a)	34.7%	47.6%	42.9%	69.1%
Process first (b)	65.3%	52.4%	57.1%	30.9%

were significantly more likely to self-report a process-first approach, whereas students who remained incorrect were significantly more likely to self-report an answer-first approach. The approaches reported by students who changed responses from provisional to submitted did not differ significantly from each other, regardless of the direction of the change.

IV. CONCLUSION AND NEXT STEPS

In this exploratory investigation, we have used metacognitive prompts along with timing data to gain insight into students' reasoning trajectories. We were able to show that students who self-reported that they revised their thinking before submitting an answer spent significantly longer answering the question than those who did not, which helps validate the self-reported data we are using to explore reasoning trajectories. In addition, while all students who did not revise their thinking spent less time answering, there were significant differences in terms of doubts and approaches depending on whether they retained correct or incorrect provisional models.

In future work, we plan to add a prompt explicitly asking students to indicate their provisional answers so that we will address some of the limitations associated with constructing it as we had to in this exploratory study. In addition, future tasks will include an additional physics question to provide an independent assessment of students' mindware so that we may gain more insight into the various reasons why some students retain an incorrect provisional model.

The findings of our exploratory investigation suggest that the use of metacognitive prompts along with timing data can help researchers characterize student reasoning trajectories. We anticipate that a more detailed understanding of these trajectories will be an important step in the development of effective, research-based instructional materials that better support student reasoning in physics.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under Grant Nos. DUE-1821390, DUE-1821123, DUE-1821400, DUE-1821511, and DUE-1821561.

-
- [1] A.F. Heckler, The ubiquitous patterns of incorrect answers to science questions: The role of automatic, bottom-up processes, *Psychol. Learn. Motiv.* **55**, 227 (2011).
- [2] M. Kryjevskaja, M.R. Stetzer, and N. Grosz, Answer first: Applying the heuristic-analytic theory of reasoning to examine student intuitive thinking in the context of physics, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020109 (2014).
- [3] M. Kryjevskaja and M.R. Stetzer, Examining inconsistencies in student reasoning approaches, *AIP Conf. Proc.* **1513**, 226 (2013).
- [4] J. S. B. T. Evans, The heuristic-analytic theory of reasoning: Extension and evaluation., *Psychon. Bull. Rev.* **13**, 378 (2006).
- [5] D. Kahneman, *Thinking, Fast and Slow* (Farrar, Strauss, & Giroux, New York, 2011).
- [6] K. Stanovich, Miserliness in human cognition: The interaction of detection, override and mindware, *Think. Reas.* **24**, 423 (2018).
- [7] M. Kryjevskaja, P.R. Heron, A.F. Heckler, Intuitive or rational? Students and experts need to be both, *Phys. Today* **74**, 8, 28 (2021).
- [8] S. Frederick, Cognitive Reflection and Decision Making, *J. Econ. Perspect.* **19**, 25 (2005).
- [9] M. E. Toplak, R. F. West, and K. E. Stanovich, The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks, *Mem. Cognit.* **39**, 1275 (2011).
- [10] A. F. Heckler, T.M. Scaife, Patterns of response times and response choices to science questions: the influence of relative processing time, *Cogn. Sci.* **39**(3) (2015).
- [11] L. C. McDermott and P. S. Shaffer, and the Physics Education Group at the University of Washington, *Tutorials in Introductory Physics* (Prentice-Hall, Upper Saddle River, NJ, 2002).
- [12] H. G. Close, L.S. Gomez, and P.R. Heron, Student understanding of the application of Newton's second law to rotating rigid bodies, *Amer. J. Phys.* **81**, 458 (2013).
- [13] Qualtrics, Provo, Utah, USA (2019).
- [14] K. A. Richards and M.A. Hemphill, A practical guide to collaborative qualitative data analysis, *J. Teach. Phys. Educ.*, **37**(2), 225–231 (2018).
- [15] <https://sankeymatic.com/build/>. Retrieved 4/28/2022.
- [16] A. Field, *Discovering statistics using IBM SPSS statistics* 4th ed. (Sage, Los Angeles, 2013) p. 196.