Supporting student construction of alternative lines of reasoning

Mikayla Mays and MacKenzie R. Stetzer

Department of Physics and Astronomy and Maine Center for Research in STEM Education, University of Maine, Orono, ME 04468, USA

Beth A. Lindsey

Physics, Penn State Greater Allegheny, McKeesport, Pennsylvania 15132, USA

An emerging body of research suggests that, even after research-based instruction, poor student performance on certain physics tasks may stem primarily from domain-general reasoning phenomena rather than from a lack of conceptual understanding. The reasoning patterns (and inconsistencies) reported in these studies may be explained by dual-process theories of reasoning (DPToR). In order to help students strengthen their reasoning skills and support increased cognitive reflection, there is a need to design and test instructional intervention strategies that leverage DPToR and that may ultimately guide the development of research-based curricular materials that attend to the nature of human reasoning more explicitly. This investigation focused on an intervention designed to support analytical processing in which students were asked to set aside their own reasoning and engage in alternative lines of reasoning. In the intervention, students first responded to a qualitative physics task, then constructed reasoning chains in support of answers to that task given by two fictitious students, and finally revisited the original physics task. Analysis revealed that this intervention was successful at improving student performance. Furthermore, it appears to have supported students regardless of their cognitive reflection skills, and its effectiveness may potentially be correlated with the quality of reasoning chains generated in support of the correct fictitious student's response.

I. INTRODUCTION

Even after large-scale course transformation efforts and the implementation of research-based instructional materials, researchers have documented a phenomenon in which students exhibit inconsistent reasoning patterns on two questions targeting the same physics concepts [1–4]. A significant percentage of students answer one question correctly, but do not appear to apply those same concepts on a similar question. Researchers in physics education have proposed that this type of inconsistent response pattern may stem primarily from the nature of human reasoning rather than from a lack of conceptual understanding [1, 5]. As a result, they have begun to use the framework of dual-process theories of reasoning (DPToR) when examining student reasoning in physics.

According to DPToR, human cognition can be modeled as consisting of two processes: the heuristic process (or process 1) and the analytic process (or process 2) [6, 7]. Process 1 generates a quick, intuitive response that relies upon prior knowledge, beliefs, and contextual cues. Process 2 is slower, exacting, and is frequently tasked with determining if the response generated by process 1 is satisfactory. Though the path that human reasoning takes is dependent upon a complex interplay of many factors, prior research has leveraged DPToR and related cognitive constructs to account for the inconsistent reasoning patterns described above and has recently begun to inform instructional strategies [5, 8–10]. Although this research illustrates the potential of DPToR to guide the development of research-based instructional materials, most existing research-based materials (which have been shown to substantively improve conceptual understanding) do not explicitly leverage such theories to support the development of student reasoning skills. There is thus a real need for the development and testing of targeted, DPToR-aligned interventions that may be incorporated into future instructional materials. Such interventions may also serve as probes to better understand specific factors and mechanisms that can impact student reasoning and may be leveraged during instruction.

As part of this effort, we have investigated the effectiveness of a targeted DPToR-aligned intervention aimed at helping students engage with alternative lines of reasoning. In this Constructing Alternative Reasoning (CAR) intervention, students were asked to: (1) respond to a qualitative physics task on kinematics, (2) construct reasoning chains in support of answers to that task given by two fictitious students, and (3) revisit the original physics task. The investigation was designed to answer the following three research questions:

1. To what extent can we support process 2 and assist students in overriding an incorrect intuitive response by asking them to generate reasoning chains in support of the correct answer and the incorrect intuitive answer?

2. To what extent is student performance on the target task related to cognitive reflection skills?

3. To what extent does the effectiveness of this intervention correlate with relevant factors (*e.g.*, cognitive reflection skills)?
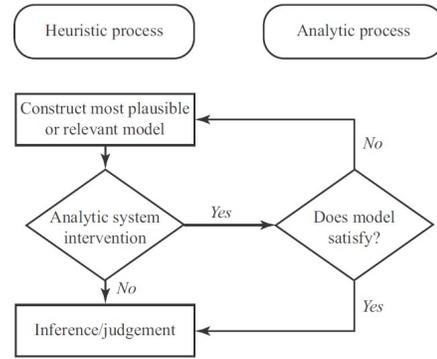

FIG. 1. Diagrammatic representation of reasoning in Evans' heuristic-analytic theory. Adapted from [7].

In Section II, we provide an overview of Evans' extended heuristic-analytic theory and related cognitive constructs. Research design and methodology are discussed in Section III. Results are presented in Section IV, followed by a brief summary of our findings in Section V.

II. EXTENDED HEURISTIC-ANALYTIC THEORY AND RELATED COGNITIVE CONSTRUCTS

For this work, we have used Evans' extended heuristic-analytic theory of reasoning to guide intervention design and analysis [7]. Figure 1 summarizes how heuristic-analytic theory models human reasoning. Upon encountering a (physics) problem, process 1 will generate a 'first-available' mental model based on the context, goals, and experiences of the reasoner as well as the features of the problem under consideration *(relevance principle)*. If there is no intervention by process 2 at this point, an inference or judgement will be made in accordance with the first-available mental model. If process 2 does intervene, the first-available mental model will be evaluated to determine whether or not it is satisfactory. According to heuristic-analytic theory, only one mental model is considered at a time *(singularity principle)*, and a new one is only generated if the previous model has been deemed unsatisfactory *(satisficing principle)*. Thus, even with the intervention of process 2, a failure to override an incorrect first-available mental model may result in the rationalization of an incorrect answer (often due to reasoning biases). If the model is deemed unsatisfactory, another plausible model will be generated and (possibly) tested.

It is important to note that process 1 may generate a first-available mental model informed by features of the problem that are *irrelevant* to the solution but still capture students' attention. Such features are known as salient distracting features (SDFs) since they tend to distract students from the correct line of reasoning due to the generation of incorrect, SDF-cued first-available mental models that impact the students' reasoning trajectories [11].

Two additional cognitive constructs were leveraged for intervention design and analysis. The concept of mindware refers to the "knowledge bases, rules, procedures, and strategies" required for successful performance on a task [12]. In this study, we embraced a screening-target methodology in which an independent measure of mindware (the screening question) was used to identify incorrect answers on the question of interest (the target question) rooted in a lack of conceptual understanding. By screening such responses, we could focus on incorrect target responses that stemmed from type 2 processing issues [4, 5, 9].

Cognitive reflection skills refer to a reasoner's tendency to critically reflect upon their first-available mental models. The Cognitive Reflection Test (CRT) is a widely used three-item instrument developed by Frederick to measure cognitive reflection skills by testing respondents' abilities to override incorrect intuitive answers [13]. For each item, there is an incorrect intuitive answer that is relatively easy to discard upon quick reflection, and an individual's CRT score corresponds to the total number items answered correctly. Prior research in physics education has used the CRT to investigate student reasoning and performance in physics [8, 9, 14].

III. RESEARCH DESIGN AND METHODOLOGY

This study was conducted in an introductory calculus-based mechanics course offered at a medium-sized public university in New England. Approximately 350 students, primarily engineering majors, were enrolled in the course. Synchronous, online lectures that included frequent opportunities for interactive engagement were held for 50 minutes three times a week. Students were required to attend weekly online laboratory sessions as well as online small-group discussion sessions that utilized *Tutorials in Introductory Physics* [15] as group activities. The intervention sequence presented below was included as part of an online participation-based homework assignment. Students received full credit for these weekly online assignments regardless of the correctness of their responses. Only data from students who had responded to all portions of the intervention were analyzed (N = 222).

The rationale and design of the Constructing Alternative Reasoning (CAR) intervention was guided by Evans' heuristic-analytic theory [7]. Due to the satisficing principle, a reasoner is unlikely to consider alternative models unless red flags are raised about their first-available mental model and process 2 concludes that it is not satisfactory. The CAR intervention was designed to prompt students to engage in analytical processing in support of both the correct answer and the common incorrect answer. By justifying the answers of these hypothetical students, students were asked to set aside (at least temporarily) their first-available mental models and construct alternative lines of reasoning. Thus, our hypothesis was that asking students to justify both correct and common incorrect answers would necessitate analytical engagement with both models (not just the "intuitive" common incorrect

The motions of two cars are described by the position *vs.* time graphs shown above. At which of the three labeled times are the magnitudes of the velocities (*i.e.,* the speeds) of the cars the same?
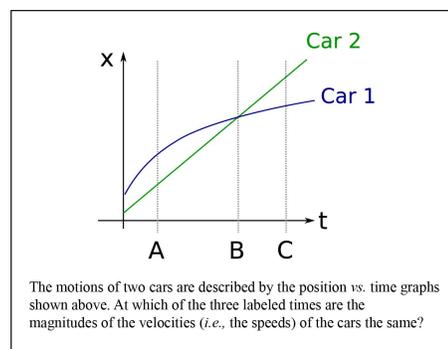
FIG. 2. Kinematics graph task used as target question. Adapted from [1].

one) and increase the likelihood of abandoning the incorrect line of reasoning in favor of the correct one.

The kinematics graph task (see fig. 2) reported on by Heckler [1] served as the target question for this intervention. In this question, students were given position *vs.* time graphs of two cars and asked to identify the time at which the cars have the same speed. To answer correctly, students had to recognize that the cars have the same speed at time A since the slopes of both graphs are the same at that time. However, the intersection point has empirically been shown to serve as a salient distracting feature, and thus the most common incorrect answer is time B [1].

Two screening questions (shown in fig. 3) were used to ascertain whether students possessed the mindware required to answer the target question correctly [16]. In each question, students were given a position *vs.* time graph and were asked to determine the time at which the speed of each car was the greatest. In the complete CAR intervention sequence, the screening questions were administered before the target question. In order to avoid cueing correct reasoning on the target question via proximal practice, several other physics questions were placed between the screening questions and the first instance of the target question. Both the screening questions and the target were administered in multiple-choice format, with a free-response prompt to explain their reasoning. (It should be noted that the term "positions" was inadvertently used in place of "times" in the multiple-choice target question prompt, but a careful analysis of student responses revealed

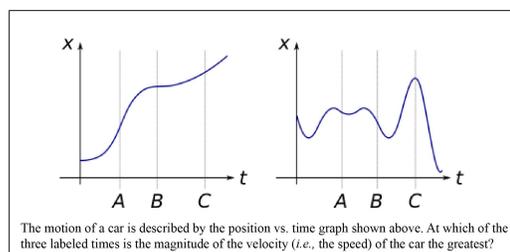
The motion of a car is described by the position vs. time graph shown above. At which of the three labeled times is the magnitude of the velocity (*i.e.,* the speed) of the car the greatest?

FIG. 3. Screening questions on kinematics graphs. [16]

Items

$\Delta x_{t_1 \to t_2} = \int_{t_1}^{t_2} v(t)\,dt$

$v = \frac{dx}{dt}$

the integral $\int h(r)\,dr$ is the area under the graph of h(r) vs. r

the derivative $\frac{dh(r)}{dr}$ at a specific point is the slope of the tangent line of the h(r) vs. r graph at that point

velocity is given by the value of the slope of a position vs. time graph

displacement is given by the area under the velocity vs. time graph

the lines intersect at time B

slopes are the same at time A

Custom:

Custom:

Custom:

Reasoning Space

Connecting Words

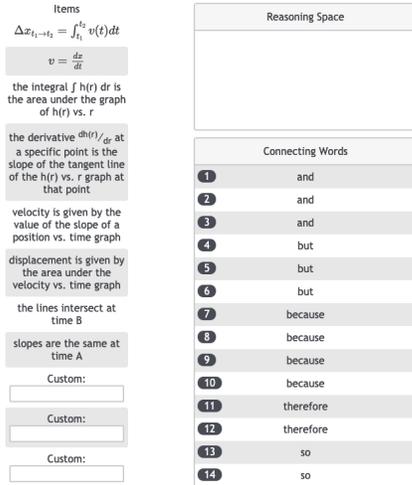| 1 | and |
2	and
3	and
4	but
5	but
6	but
7	because
8	because
9	because
10	because
11	therefore
12	therefore
13	so
14	so

FIG. 4. Reasoning chain construction format.

no evidence of this error impacting student reasoning.)

After students answered the target question, they were presented, one at a time, with the hypothetical answers of two fictitious students, one of which is correct (A) and the other of which is the most common incorrect answer (B). For each hypothetical answer, students were asked to construct a line of reasoning that the fictitious student might have used to reach their conclusion via the reasoning chain construction format [17], a modified card sort activity implemented within Qualtrics' "Rank/Group/Sort" question format (Fig. 4) [18]. In this format, students constructed a reasoning chain by drawing from the provided reasoning elements and placing them in the reasoning space. Each element was either a first principle of physics or mathematics, a derived heuristic (*i.e.*, a simple relationship taught in class that stems from the combination of two or more first principles), or an observation about the graph. Students were explicitly told that all elements are true statements. Students were also provided with a few customizable elements if needed for their arguments. After the intervention, students were asked to respond to the target question again. They subsequently completed the CRT.

IV. ANALYSIS AND DISCUSSION

Overall, student performance was quite strong on both the screening and target questions. Of the 222 student responses analyzed, 89% of students answered both screening questions correctly with explanations sufficient to demonstrate evidence of mindware. Additionally, 81% of all students answered the pre-intervention target question correctly with a correct explanation, a somewhat higher percentage than anticipated based on previous research using this task [1, 16]. After the intervention, 89% of students answered the target question correctly with a correct explanation.

Since our intervention was primarily designed to support

TABLE I. Student performance on target question pre- and post-intervention.

Target Response (Students with mindware)	Pre-intervention (N=198)	Post-intervention (N=198)
Correct	83%	91%
Incorrect	17%	9%

those students who possessed the requisite mindware but didn't override an incorrect first-available mental model (*i.e.*, students who answered the screening questions correctly but gave the common incorrect response on the first target), the data analysis that follows contains only students who demonstrated evidence of mindware (N = 198).

Pre- and post-intervention target response data are shown in Table I for those students with mindware. There was a statistically significant improvement in performance on the target question (McNemar, p=.0004) with a large effect size (Cohen's g=.40) after the intervention. These results indicate that the intervention was effective in shifting students with mindware toward the correct response.

Based on DPToR, we predicted that students with weaker cognitive reflection skills would be less likely to scrutinize an incorrect mental model and more likely to answer the target question incorrectly. We thus expected a correlation between pre-intervention target response and CRT score. In accordance with previous literature, CRT scores of 0 or 1 were classified as low, while scores of 2 or 3 were classified as high. Table II shows the pre-intervention target responses for students with both low and high CRT scores. There is a statistically significant correlation between pre-intervention target response and CRT score (Pearson chi squared, p=.021) with a small effect size (Cramer's V=.165), as predicted.

Given that our intervention directed students to engage in analytical processing in support of both answers, it was not clear whether the effectiveness of the intervention would correlate with cognitive reflection skills. However, one could argue that a student's tendency to scrutinize their first-available mental model is not likely to be a factor, as the intervention essentially sidesteps constraints associated with the singularity and satisficing principles by expressly asking students to analytically engage with two different models. Table III

TABLE II. Pre-intervention target performance data for students with low and high CRT scores.

	Target Performance (Students with mindware)	Low CRT (N=84)	High CRT (N=114)
Pre-Intervention	Correct	76%	88%
	Incorrect	24%	12%
Post-Intervention	Correct	87%	95%
	Incorrect	13%	5%

TABLE III. Change in target performance data for students with low and high CRT scores.

Change in Target Response (Students with mindware)	Low CRT (N=20)	High CRT (N=13)
Change: Incorrect to correct	45% (9)	70% (9)
Remained incorrect	55% (11)	30% (4)

TABLE IV. Change in target performance data for students with and without physics mapping in reasoning chains supporting the correct answer.

Change in Target Response (Students with mindware)	No Physics Mapping (N=15)	Physics Mapping (N=18)
Change: Incorrect to correct	33% (5)	72% (13)
Remained incorrect	67% (10)	28% (5)

shows pre to post shifts (or non-shifts) in target response for students with both low and high CRT scores. No statistically significant correlation between pre to post target shift and CRT score was observed (Fisher exact with Bonferroni correction for 2 tests, p=.284>0.025), but this may stem from a lack of statistical power (which we plan to address via additional data collection). It is worth noting, however, that the correlation between post-intervention target response and CRT score (from data shown in Table II) is on the edge of statistical significance (Pearson chi squared, p=.052). If this intervention supports students of all levels of cognitive reflection skills, one would expect that the correlation between student performance and CRT score post intervention, if present, would be weaker than pre intervention, which is what we observe. Thus, while our results are potentially consistent with the intervention minimizing the role of cognitive reflection in effective analytic processing, this claim cannot be made without a larger sample size.

Since our results (though limited in statistical power) do not suggest a strong correlation between the effectiveness of our intervention and cognitive reflection skills, we were interested in exploring whether its effectiveness was related to the quality of the reasoning chains generated in support of the correct *hypothetical answer*. For students who initially answered the target question incorrectly, the reasoning chains they generated in support of the correct answer were coded on the basis of whether or not relevant features of the position *vs.* time graph were mapped to appropriate physics concepts. A chain demonstrating physics mapping had to include both an observation statement ("the slopes are the same at time A") and a physics statement (indicating the relevant physics concepts) from the given elements. Appropriate physics statements could include either "velocity is given by the value of the slope of a position vs time graph" or a combination of "$v = dx/dt$" and "the derivative, dh/dr, at a specific point is the slope of the tangent line of the $h(r)$ *vs.* r graph at that

point". Table IV shows pre to post shifts (or lack thereof) for students who did and did not demonstrate physics mapping. While we did not observe a statistically significant correlation between pre to post target shift and presence of physics mapping after correcting for 2 tests (Fisher exact with Bonferroni correction, p=.038>.025), the proximity to the statistical significance threshold suggests the null result likely stems from low statistical power. In the absence of a larger sample, these results raise the possibility that, of students who initially answered the target incorrectly, those who were able to invoke relevant physics concepts in support of the correct hypothetical answer may have been more likely to switch to the correct answer after the intervention.

V. CONCLUSIONS AND NEXT STEPS

In this investigation, we examined the effectiveness of a constructing alternative reasoning (CAR) intervention in which students were guided to construct alternative lines of reasoning in support of two different fictitious students' answers to a kinematics graph task. Overall, our analysis demonstrates that the CAR intervention was successful at improving student performance on the target task. We found that pre-intervention target responses were correlated with cognitive reflection skills, as predicted by DPToR. While the effectiveness of the intervention did not appear to be correlated with cognitive reflection skills, low statistical power makes it impossible to rule out the possibility of a correlation at this time. (DPToR, however, may suggest that no correlation should exist if the intervention successfully sidesteps both the singularity and satisficing principles.) Finally, our low-N results raise the possibility that initially incorrect students who leveraged relevant physics concepts in support of the correct answer may have been more likely to shift to a correct response, which may indicate that this intervention scaffolds productive analytic processing.

This investigation is part of a larger effort to identify the extent to which DPToR can be used to inform intervention strategies that impact student reasoning *in the moment* and to ascertain the factors that may impact the effectiveness of the intervention. Given the short duration of the CAR intervention, we would not expect any long-term impact on student reasoning, but we hope to incorporate this type of intervention into a coherent instructional sequence (for implementation and testing) that includes multiple interventions, spaced over time, along with explicit efforts to foreground the importance of considering alternative models.

[1] A.F. Heckler, The ubiquitous patterns of incorrect answers to science questions: The role of automatic, bottom-up processes, Phychol. Learn. Motiv. **55**, 227 (2011).

[2] A. Elby, Helping physics students learn how to learn, Am. J. Phys. **69**, S54 (2001).

[3] E.F. Redish, A theoretical framework for physics education research: Modeling student thinking, Proceedings of the International School of Physics Enrico Fermi (2004).

[4] M. Kryjevskaia and M.R. Stetzer, Examining inconsistencies in student reasoning approaches, AIP Conf. Proc. **1513**, 226 (2013).

[5] M. Kryjevskaia, 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).

[6] D. Kahneman, *Thinking, Fast and Slow* (Farrar, Strauss, & Giroux, New York, 2011).

[7] J.St.B.T. Evans, The heuristic-analytic theory of reasoning: extension and evaluation, Psychon. Bull. Rev. **13**, 378 (2006).

[8] C.R. Gette and M. Kryjevskaia, Establishing a relationship between student cognitive reflection skills and performance on physics questions that elicit strong intuitive responses, Phys. Rev. Phys. Educ. Res. **15**, 010118 (2019).

[9] M. Kryjevskaia *et al.*, Designing research-based instructional materials that leverage dual-process theories of reasoning: Insights from testing one specific, theory-driven intervention, Phys. Rev. Phys. Educ. Res. **16**, 020140 (2020).

[10] B.D. Mikula and A.F. Heckler, Framework and implementation for improving physics essential skills via computer-based practice: Vector math, Phys. Rev. Phys. Educ. Res. **13**, 010122 (2017).

[11] S. Mamede, T.A.W. Splinter, T. Van Gog, R.M.J.P. Rikers, and H.G. Schmidt, Exploring the role of salient distracting clinical features in the emergence of diagnostic errors and the mechanisms through which reflection counteracts mistakes, BMJ Qual. Saf. **21** 295 (2012).

[12] K.E. Stanovich, Miserliness in human cognition: The interaction of detection, override, and mindware, Think. Reas. **24** 423 (2018).

[13] S. Frederick, Cognitive reflection and decision making, J. Evon. Perspect. **19**, 25 (2005).

[14] A.K. Wood, R.K. Galloway, and J. Hardy, Can dual processing theory explain physics students' performance on the Force Concept Inventory?, Phys. Rev. Phys. Educ. Res. **12**, 23101 (2016).

[15] 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).

[16] J. C. Speirs, M. R. Stetzer, B. A. Lindsey, and M. Kryjevskaia, Exploring and supporting student reasoning in physics by leveraging dual-process theories of reasoning and decision-making, submitted to Phys. Rev. Phys. Educ. Res. (under review).

[17] J.C. Speirs, W.N. Ferm Jr., M.R. Stetzer, and B.A. Lindsey, Probing student ability to construct reasoning chains: A new methodology, in *presented at the Physics Education Research Conference 2016*, Sacramento, CA, 2016.

[18] Qualtrics, Provo, Utah, USA, 2019.