

## **Insights from an intervention designed to support consistent reasoning**

Thomas Fittswood, Drew J. Rosen, MacKenzie R. Stetzer,

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

An emerging body of research suggests that poor student performance on certain physics questions may stem, at least in part, from the nature of human reasoning itself. While students may demonstrate that they possess the requisite knowledge and skills to reason correctly on one question, they may abandon that same line of reasoning on an analogous question containing a salient distracting feature. As part of a larger effort to investigate and support student reasoning in physics by leveraging dual-process theories of reasoning, we developed and tested an intervention aimed at helping students draw upon the knowledge and skills they already possess to address such reasoning inconsistencies. In this study, we also explored specific factors to see if they were related to student reasoning and how students engage with the intervention. We found that the intervention was effective in helping students reason more productively and consistently, but its effectiveness appears to be related to students' cognitive reflection skills. In addition, out of the students who initially answered two analogous physics questions inconsistently, those who were able to successfully apply their correct reasoning from one question to the other question upon explicit prompting were more likely to revise their thinking and demonstrate consistent reasoning after the intervention.

## I. INTRODUCTION

Over many decades, research in physics education has improved our understanding of the learning and teaching of physics and has led to the development of research-based instructional materials and strategies. These materials and strategies substantively improve student learning outcomes in introductory physics courses and beyond [1,2]. However, even after research-based instruction, students who are able to correctly apply relevant *mindware* (a term from cognitive science that refers to conceptual knowledge and procedural skills [3]) on one physics question often do not access and apply that mindware on an analogous question probing the same physics concepts. These persistent patterns of inconsistent reasoning have been documented across numerous topics and contexts within physics [4-8].

Research suggests that domain-general reasoning phenomena (*i.e.*, phenomena related to the nature of human reasoning itself) may play an important role in accounting for these documented inconsistencies in student reasoning [4,6]. Researchers in physics education are thus increasingly turning to dual-process theories of reasoning (DPToR), from cognitive science, to better model student reasoning [4,7-10].

DPToR are a collection of theories that model human reasoning and decision-making as consisting of two distinct processes [11,12]. Process 1 (the heuristic process) is fast, automatic, and subconscious, and is responsible for generating a first-available mental model. Process 2 (the analytic process) is a slow, effortful process that may or may not be engaged. When process 2 is engaged, it is tasked with ascertaining whether or not the first-available model put forth by process 1 is satisfactory [11].

To date, DPToR have been useful in accounting for and/or predicting patterns of inconsistent reasoning on physics questions [4,7,10]. Moreover, the DPToR framework has recently been used to guide the development of instructional interventions in physics [13,14,15]. Indeed, the work presented here is part of an ongoing, multi-institutional effort to leverage DPToR to create instructional materials that better support students in reasoning effectively.

While all dual-process theories share the same core characteristics, our work has been informed by the extended heuristic-analytic theory proposed by Evans [11]. In this model, a student's thought process starts with a first-available mental model generated by process 1. At this point, process 2 may or may not be engaged to evaluate the model. If process 2 is not engaged, an answer will be given based on the initial mental model (the path of cognitive frugality). If process 2 is engaged, the initial model will be evaluated, although reasoning biases may impact that evaluation (e.g., confirmation bias). Alternative models will only be explored if the initial model is found to be unsatisfactory.

The term *cognitive reflection skills* refers to the tendency and ability of a reasoner to scrutinize (via process 2) first-available mental models. The cognitive reflection test (CRT) consists of three questions that elicit strongly appealing incorrect answers but may be answered correctly upon quick reflection [17,18]. CRT scores (0-3) are used in our analysis as an independent measure of cognitive reflection skills.

This paper describes the design, implementation, and analysis of an online *Reasoning Pathways Intervention* that flexibly supports students based on how they respond to two analogous questions. In this intervention, all students are asked about the similarity of the reasoning approaches they used on the questions before being routed to different prompt sequences based on how they answered the two questions. In this paper, however, we primarily focus on the intervention branch administered to students who answered inconsistently, which explicitly asks students to apply the reasoning they used successfully on one question to the other. The present investigation seeks to answer two research questions:

- 1) To what extent is the Reasoning Pathways Intervention effective at supporting students in reasoning more effectively?
- 2) For those students who demonstrate the requisite mindware but answer inconsistently, what factors appear to be related to the effectiveness of the intervention in helping them shift to correct reasoning on the more challenging question?

Section II covers the specifics of the intervention design and data collection. In Section III, results are presented and discussed, and Section IV concludes with a brief summary of our findings and a discussion of next steps.

## II. DESIGN AND IMPLEMENTATION

Here we describe the Reasoning Pathways Intervention, linking it to the theoretical underpinnings, and provide an overview of its implementation.

### A. Intervention design

In this work, we used a *screening-target question pair* methodology, which was designed to disentangle, to the extent possible, reasoning approaches from mindware [6,7,16]. Both questions in a screening-target pair require students to apply the same mindware to arrive at a correct answer. However, the target question typically elicits intuitively appealing incorrect answers that are often cued by salient distracting features (SDFs), whereas the screening question does not. In this way, the screening question serves as a proxy for mindware. Thus, when students answer the screening question correctly and the target question incorrectly, they demonstrate that they possess the requisite

mindware, but they do not appear to access it when answering the target question.

The Reasoning Pathways Intervention was constructed around a single screening-target question pair. After answering the screening and target questions, students were shown their answers to the two questions and were asked if they used similar lines of reasoning when answering the two questions (*check consistency*). After the *check consistency* prompt, students were routed to one of three pathways based on the answers they gave to the screening and target questions, as shown in Fig. 1.

Students who answered both questions correctly were simply asked to revisit the target question after responding to the *check consistency* prompt. Students who answered the screening question correctly (thereby demonstrating mindware) and the target question incorrectly were our primary population of interest for this investigation. While the relevant mindware appeared to be available, these students did not access it when answering the target question. They were served a series of two prompts (*identify features* and *apply screening reasoning*, shown in Table I). By explicitly asking students to apply the reasoning they successfully used on the screening question to the target question, the intervention was designed to stimulate a productive engagement of process 2 and aid students in considering an alternative mental model when answering the target question. Students who answered the screening question incorrectly were assumed to lack relevant mindware and were routed to a brief intervention designed to remind them of key concepts. All three groups of students were given an opportunity to change their answer to the target question at the end of the intervention. This allowed for a comparison of performance on the target question before and after the intervention.

The intervention was built around a screening-target pair (shown in Fig. 2) on the topic of single-loop, multiple-battery circuits. Both the screening and the target questions required the same mindware.

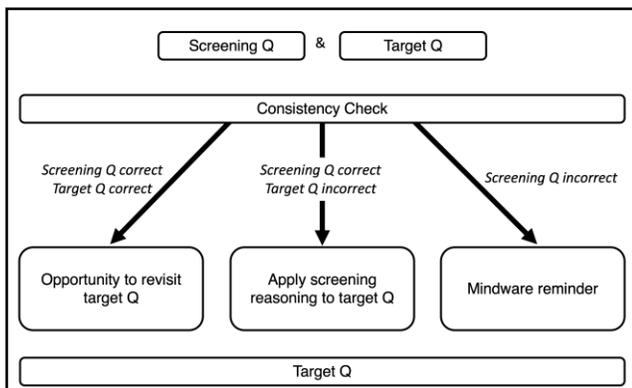


FIG. 1. Diagram of the Reasoning Pathways Intervention.

TABLE I. Prompt sequence following the screening-target pair for students who answered screening correctly and target incorrectly.

Questions served to students with screening correct and target incorrect.
<b>Check consistency.</b> Did you use similar lines of reasoning when answering these two questions?
<b>Identify features.</b> Were there certain features of the physical scenario that led you to choose your answer? If so, what were they?
<b>Apply screening reasoning.</b> If you were to use your line of reasoning from the screening question on the target question, at which answer would you arrive?
<b>Revisit target.</b> Now that you have had time to reflect more carefully on the target question, indicate your final answer.

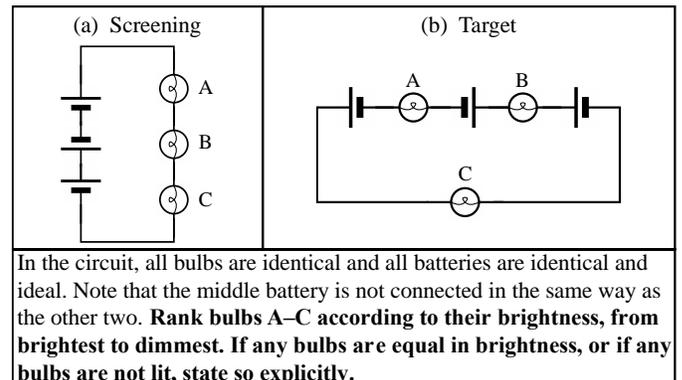


FIG. 2. Screening and target question pair used in intervention.

The key piece of mindware is that current is the same through elements in series. Additional relevant mindware includes recognition that current may flow backward through a battery and that batteries in series can be combined algebraically. Applying this mindware to both questions yields the correct answer that all bulbs are equally bright ( $A=B=C>0$ ). On both questions, students were asked to select a multiple-choice answer and explain their reasoning.

The target question (Fig. 2b) was drawn from prior work on student understanding of circuits [19]. In that investigation, approximately 30% of students gave answers inconsistent with Kirchhoff's junction rule (which requires current to be the same through elements in series), with explanations often focusing on the locations of bulbs A and/or B between like terminals (*i.e.*, between two positive or two negative terminals). These results suggested that bulb locations between like battery terminals may act as a salient distracting feature (SDF), cuing incorrect mental models that are inconsistent with Kirchhoff's junction rule.

In order to have a measure of mindware, the screening question (Fig. 2a) was designed to require the exact same mindware as the target question without the inclusion of the SDF. Its solution was thus isomorphic to that of the target question, and student explanations could be analyzed for evidence that students articulated the key piece of mindware (same current through elements in series).

TABLE II. Performance on screening and target questions.

N = 131	% Correct with correct reasoning
Screening	70%
Pre-intervention target	51%
Post-intervention target	59%

### B. Implementation

This online intervention was administered to students in a second-semester calculus-based introductory physics course. It was included as part of an online participation-based homework assignment given via Qualtrics [20] after all relevant instruction. Students received participation credit, and correct answers and illustrative explanations were made available to students at the end of the assignment.

A total of N=131 valid responses were collected. Duplicate responses from the same student and responses that showed a lack of engagement (e.g., missing explanations) were excluded from analysis. Explanations were coded as ‘correct with correct reasoning’ on the screening and the pre/post-intervention target questions if they provided an explanation that included a reference to any of the three pieces of mindware in support of the correct multiple-choice answer. The screening question responses were also separately coded for whether or not students demonstrated evidence of the key piece of mindware based on the presence or absence of statements about the bulbs being connected in series or in a single loop. Several different statistical tests were performed in our analyses, and details of the tests can be found in reference [21]. Unless stated otherwise, the p-value threshold of significance was 0.05 for all tests.

## III. RESULTS AND DISCUSSION

### A. Overall effectiveness of intervention

Results from the pre-intervention screening and target questions as well as the post-intervention target question are shown in Table II. As expected, performance was strongest on the screening question. Of those students who answered the pre-intervention target question correctly, very few answered the screening question incorrectly, suggesting that the screening-target pair performed as intended.

In order to answer research question 1, two statistical tests were performed to compare performance on the pre- and post-intervention target questions. McNemar’s test provided a measure of whether students were primarily shifting in a particular direction (e.g., from incorrect to correct). A binomial test was used to determine if the difference in the overall distribution of responses was significant. The intervention led to a statistically significant shift in responses in the desired direction with a large effect size (McNemar,  $p=0.0075$ ,  $g=0.36$ ), and significantly changed the response distribution with a small effect size (binomial,  $p=0.045$ ,  $h=0.16$ ). These results suggest that the

intervention helped students reason more productively on the target question but its impact is limited in scope.

Given that the intervention was designed to promote type 2 processing of an alternative mental model, we had hoped that the intervention would improve student reasoning independent of cognitive reflection skills. We thus expected that post-intervention target question performance would correlate less strongly with CRT score than pre-intervention target question performance.

The CRT was administered in a separate participation-based homework, and matched CRT scores were available for a total of N=122 intervention responses. Performance on the pre-intervention target question had a statistically significant correlation with CRT score with a small effect size (Mann-Whitney U,  $p=0.007$ ,  $r=0.24$ ). Post-intervention target performance, however, had a stronger correlation with CRT as indicated by the larger effect size (Mann-Whitney U,  $p<0.001$ ,  $r=0.31$ ). Despite the principles guiding its design, the intervention appeared to widen the gap between students with stronger and weaker cognitive reflection skills.

The remaining analysis focuses exclusively on those students who were routed to the middle intervention branch in Fig. 1 and whose screening question responses demonstrated evidence of the key piece of mindware. There were N=31 such responses. Students in this group necessarily gave an incorrect answer on the pre-intervention target question. The following analysis focuses on whether or not students shifted to the correct target reasoning.

### B. Relating shifts to correct to possible factors

In order to answer research question 2, two factors were considered that may relate to the effectiveness of the intervention for students who reasoned inconsistently. *Check consistency* responses and *apply screening reasoning* responses were investigated for correlation with a shift to the correct target response. Fisher’s exact tests were performed to test for correlation, with a Bonferroni correction to account for 2 tests reducing the significance threshold from 0.05 to 0.025.

#### 1. Identification of inconsistency

Since all students considered in this analysis answered the screening-target pair inconsistently, self-reporting a possible inconsistency may be a sign that the initial model was being reconsidered, a hallmark of the productive engagement of process 2. However, the correlation between recognition of inconsistency and a shift to the correct response (Table III) was not statistically significant (Fisher exact with Bonferroni,  $p=0.077>0.025$ ). Given the small number of responses examined, statistical power may be an issue, so we plan to collect additional data.

TABLE III. Pre-post target response vs. *check consistency*.

Pre-post target	Recognize inconsistency?	
	No (22)	Yes/not sure (9)
Shifted to correct	18% (4)	56% (5)
Stayed incorrect	82% (18)	44% (4)

TABLE IV. Pre-post target response vs. successful application of screening reasoning.

Pre-post target	Successfully apply screening reasoning?	
	No (18)	Yes (13)
Shifted to correct	11% (2)	54% (7)
Stayed incorrect	89% (16)	46% (6)

## 2. Successful application of screening reasoning

We also expected that students may have been more likely to shift to a correct target response if they were successful in applying the line of reasoning they used on the screening question to the target question. If students were successful in doing so, it can be inferred that, at least in that moment, they saw how their productive screening reasoning could lead to a different response to the target question. Based on the data in Table IV, students who gave the correct ranking when asked to apply their screening question reasoning were significantly more likely to shift to the correct answer with a medium-large effect size (Fisher exact with Bonferroni,  $p=0.017 < 0.025$ ,  $V = 0.465$ ).

## C. Trends among those who shifted to correct

A trend emerged in the free-response explanations given by students who shifted to the correct response. (See Table V for example explanations.) Some of the explanations given to the *check consistency/identify features* prompts indicated that students had not only recognized the inconsistency between their screening and target reasoning but had also already shifted to the correct target reasoning prior to the *apply screening reasoning* prompt. If explanations given for *check consistency* and *identify features* (see Table I for prompts) indicated that the student had already shifted to the correct line of reasoning, they were categorized as an early shifter. It can be surmised that these early shifters were

engaging in productive cognitive reflection early on in the intervention. Since that behavior is associated with stronger cognitive reflection skills, a correlation between early shifting and high CRT scores was expected.

A total of 9 students in this branch of the intervention shifted to the correct answer, 8 of whom had matched CRT scores. Students who shifted early had statistically significantly higher CRT scores than those who shifted later in the intervention, and this difference was characterized by a large effect size (Mann Whitney U,  $p=0.036$ ,  $r = 0.91$ ).

## IV. CONCLUSIONS AND NEXT STEPS

The Reasoning Pathways Intervention appeared to help students shift toward the correct line of reasoning, although it seemed to be more effective for students with stronger cognitive reflection skills. We have also begun to relate shifting behavior to other relevant factors, including responses to intervention prompts and cognitive reflection skills. The strength of our claims is limited by the use of a single intervention context (circuits) and the modest number of students in our analysis populations, particularly when examining early vs. late shifting ( $N < 10$ ). As we move forward, we plan to collect more data from this intervention as well as interventions in other contexts to allow us to refine our claims and to conduct longitudinal studies. In addition, we plan to use more sophisticated statistical methods, such as logistic regression, to better account for possible interplay between factors that influence how students interact with these interventions.

## 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

TABLE V. Examples of free-response explanations to intervention questions for early and late shifters.

	Early shifter	Late shifter
Check consistency	<i>No. I looked at the circuit differently when the bulbs and batteries were intermixed.</i>	<i>Yes. I considered what the electric potential differences and the currents that would be going through the bulbs at the different bulbs.</i>
Identify features	<i>I decided that A would not light because the batteries on either side of it are facing outward. In reality, the order of the elements did not matter because they are all in series and will have the same current running through them.</i>	<i>The fact that the batteries on each side of the bulbs are facing each other. In Q1 the bulbs were not surrounded by two opposing batteries.</i>
Apply screening reasoning	<i>[Correct ranking.] The order of the elements in the circuit is arbitrary because the whole circuit will have the same current and so the brightness will be the same.</i>	<i>[Correct ranking.] This is because two of the batteries would cancel, leaving the one battery providing the current to the rest of the circuit.</i>
Post-intervention target	<i>[Correct ranking.] The order of the elements does not matter. They all have the same current.</i>	<i>[Correct ranking.] This is the same question as before.</i>

- 
- [1] J. L. Docktor, and J. P. Mestre, Synthesis of discipline-based education research in physics, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020119 (2014).
- [2] L. C. McDermott and E. F. Redish, Resource letter: PER-1: Physics education research, *Am. J. Phys.* **67**, 755 (1999).
- [3] K.E. Stanovich, Miserliness in human cognition: The interaction of detection, override, and mindware, *Think. Reas.* **24**, 423 (2018).
- [4] 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).
- [5] A. Elby, Helping physics students learn how to learn, *Am. J. Phys.* **69**, S54 (2001).
- [6] M. Kryjevskaja and M. R. Stetzer, Examining inconsistencies in student reasoning approaches, *AIP Conf. Proc.* **1513**, 226 (2013).
- [7] 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).
- [8] M. Kryjevskaja, M. R. Stetzer, and T. K. Lê, Failure to engage: Examining the impact of metacognitive interventions on persistent intuitive reasoning approaches, in *2014 Physics Education Research Conference Proceedings (Minneapolis, MN, July 30-31, 2014)*, edited by P. V. Engelhardt, A. D. Churukian, and D. L. Jones, 143-146 (2015).
- [9] C. R. Gette, M. Kryjevskaja, M. R. Stetzer, and P. R. L. Heron, Probing student reasoning approaches through the lens of dual-process theories: A case study in buoyancy, *Phys. Rev. Phys. Educ. Res.* **14**, 010113 (2018).
- [10] 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**, 023101 (2016).
- [11] J. S. B. T. Evans, The heuristic-analytic theory of reasoning: Extension and evaluation, *Psychon. Bull. Rev.* **13**, 378 (2006).
- [12] D. Kahneman, *Thinking, Fast and Slow* (Mac Millan USA, New York, 2013).
- [13] M. Mays, M. R. Stetzer, and B. A. Lindsey, Supporting Student Construction of Alternative Lines of Reasoning, in *2021 Physics Education Research Conference Proceedings (Virtual Conference, August 4-5, 2021)*, edited by M. B. Bennett, B. W. Frank, and R. E. Vieyra, 277-282 (2021).
- [14] M. Kryjevskaja, M. R. Stetzer, B. A. Lindsey, A. McNerny, P. R. L. Heron, and A. Boudreaux, 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).
- [15] J. C. Speirs, M. R. Stetzer, B. A. Lindsey, and M. Kryjevskaja, Exploring and Supporting Student Reasoning in Physics by Leveraging Dual-Process Theories of Reasoning and Decision Making, *Phys. Rev. Phys. Educ. Res.* **17**, 020137 (2021).
- [16] M. Kryjevskaja, M. R. Stetzer, B. A. Lindsey, A. McNerny, P. R. L. Heron, and A. Boudreaux, 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).
- [17] S. Frederick, Cognitive reflection and decision making, *J. Econ. Perspect.* **19**, 25 (2005).
- [18] C.R. Gette and M. Kryjevskaja, 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).
- [19] M. R. Stetzer, P. Van Kampen, P. S. Shaffer, and L. C. McDermott, New Insights into Student Understanding of Complete Circuits and the Conservation of Current, *Am. J. Phys.* **81**, 134 (2013).
- [20] Qualtrics, Provo, Utah, USA, 2019.
- [21] A. Field, *Discovering statistics using IBM SPSS statistics 4th ed.* (Sage, Los Angeles, 2013).