

## Causal reasoning in understanding Newton's third law

Cheng Chen,<sup>1,2</sup> Lei Bao<sup>1,2,\*</sup>, Joseph C. Fritchman<sup>1,2</sup>, and Hemin Ma<sup>1,†</sup>

<sup>1</sup>East China Normal University, Faculty of Education,  
Department of Education & Institute of Schooling Reform and Development,  
Shanghai 200062, China

<sup>2</sup>The Ohio State University, Department of Physics, Columbus, Ohio 43210, USA



(Received 26 December 2020; accepted 29 March 2021; published 21 April 2021)

Causal reasoning is a fundamental cognitive process that supports a wide range of learning activities such as making predictions and inferences, explaining phenomena, and drawing conclusions. In physics learning, causal reasoning plays an important role in developing conceptual understanding and problem-solving skills. This study builds off the existing work to investigate aspects of students' causal reasoning in understanding Newton's third law. Quantitative assessment is conducted to explicitly probe four attributes of causal reasoning including the causal relation between a pair of interaction forces, magnitudes of interaction forces, the time order of interaction forces, and the action-reaction language. Interviews are also conducted to identify details of students' reasoning pathways. The subjects of this study include high school and college students from China. The results suggest that after instruction on Newton's third law many students still hold a belief of causal relation between the pair of interaction forces. Further analysis on students' response patterns and interviews suggests that students' belief in causality is strongly associated with asymmetries in scenario contexts due to dominant features, the scenario timeline, and the action-reaction language. In addition, the results also demonstrate a progression of four developmental levels of causal reasoning that each reveals a set of unique thinking pathways for using the asymmetries in reasoning to determine causality. The findings of this study suggest that the lack of correct causal reasoning may cause students to develop fragmented conceptual understanding, which should be addressed in future research and instruction. Detailed analysis of the findings and educational implications on assessment and instruction are discussed.

DOI: [10.1103/PhysRevPhysEducRes.17.010128](https://doi.org/10.1103/PhysRevPhysEducRes.17.010128)

### I. INTRODUCTION

Research has shown that traditional instruction resulted in many students failing to achieve a deep conceptual understanding [1–3]. Countering these problems, advances in science, technology, engineering, and mathematics (STEM) education have focused on both how students are taught and what content and skills students learn [4,5]. Through education research, many interactive-engagement methods have been developed to promote deeper conceptual understanding by building connections between knowledge elements [6,7]. Furthermore, a range of higher-end skills including reasoning, creativity, and complex

problem solving have been identified as essential education goals of 21st century learning [4–8]. Among the many descriptions of the 21st century skills, student abilities in scientific reasoning is the most commonly noted and widely studied [6,9]. Scientific reasoning emphasizes evidence-based decision making in multi-variable causal conditions which requires causal reasoning as an essential component of developing these higher-end skills. In addition, these higher-end skills are intrinsically linked to a student's ability to develop deeper conceptual understanding. A recent study on students' understanding of Newton's third law emphasizes this point by demonstrating that students' causal reasoning often underpins many intuitive misconceptions [10].

In the existing literature, students' difficulties in learning Newton's third law (N3L) have been well studied [11–14]. However, these previous studies focus on students' understanding of the N3L phenomenon but not their understanding of the fundamental causal mechanism. Helping students develop correct causal reasoning should aid developing a deeper conceptual understanding. Therefore, this study probes students' causal reasoning and more deeply explores the connection between causal

\*Corresponding author.  
bao.15@osu.edu

†Corresponding author.  
hmma@dedu.ecnu.edu.cn

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

reasoning and deep understanding of Newton's third law. The results of this study can also provide a new perspective in studying students' misconceptions by emphasizing the mechanistic understanding that underlies the students' phenomenal behaviors.

### A. Causal reasoning and knowledge forming

Causal reasoning is an important aspect of knowledge forming. Although some philosophers have argued that the aim of physics is not to explain, but rather represent phenomena absent of causality [15], many scientists and philosophers assert the important role that causality holds in science and, in particular, physics [16–20]. Among others, Max Planck wrote, “Scientific thought aspires to causality; indeed it is the same thing as causal thought, and the final aim of any science must be to take the causal viewpoint to its last consequences” [21] (Chap. 7). Thus, science and physics principles are causally related in nature [22,23,16]. These causal relationships determine the functional properties of concepts in physics and other scientific domains [16–20]. Guenther states that conceptual knowledge is “an understanding of the essential parts and cause-effect relationships that exist within a system” [24] (p. 289). Ogborn [20] stresses “the fundamental nature of causal action as an element of reasoning that is not detached from objects and events but considered as an essential part of their meaning.” It is clear then that understanding these causal relationships helps reasoning through the intrinsic functional properties of concepts.

More so than just essential to scientific thoughts, causal reasoning is an essential cognitive skill necessary to understanding the world, allowing us to predict, infer, and explain events and phenomena in the real world [25–30]. Researchers in cognitive development have shown links between causal reasoning and the development of knowledge. Piaget [18] studied the characteristics and the genesis of the causal idea in cognitive development, relating it to the attribution to objects of a set of operations. According to Piaget, causality differs from functional laws because functional laws concern observable, general regularities and imply operations applied to objects, whereas causality concerns connections going beyond the observables and implies operations attributed to objects.

A variety of more focused cognitive studies highlight the connection between causal reasoning and knowledge development. According to Anderson's [31] and Lakoff and Johnson's [32] “experiential gestalt of causation,” children observing interactions in the real world come to understand causation as agents acting on an object through an instrument. More experiences then lead to more complex understandings. Corrigan and Denton [27] further suggest that causal understanding is necessary for the development of naïve theories of physics and psychology and that causal understanding often begins as an automatic rather than a controlled process. Keil [22] suggests that

causal beliefs enable very young children to understand new concepts at deeper, theoretical levels. Causality thus is integral to knowledge development, especially in science domains.

### B. Defining a causal relation: Time, covariation, and mechanism

Discussions of causal reasoning require careful definitions of causality, with authors such as Bunge [16] and Michotte [33] writing books on the subject. Generally speaking, causality is the relation between cause and effect [34]. Three types of reasoning are used for establishing causality in physics [17,34–36]. First, the time element, also called the priority principle, where a cause must necessarily precede an effect temporally. Halbwachs [17] explains this as the delayed effects expressing real connections between things.

Second, the covariation element states that “an effect is attributed to one of its possible causes with which, over time, it covaries” [36] (p. 108). Covariational understanding is typically established based on experimental observations of events and requires comparing a potential cause and effect with others in order to determine the probability of the causal relationship. This often functions as the initial stage in a causal reasoning process [37]. Additionally, covariation is approached quantitatively and is explored through statistical analysis and inference, with the causal power or statistical probability judging the degree of viability of a relationship [38,34]. Covariations are then often represented by functional relations between quantities describing a system (e.g., energy conservation or the ideal gas law).

Third, the mechanistic element refers to mechanistic understandings or models of processes linking a cause to its effect, often at smaller scale or a deeper level of a theory. Mechanisms underlying causal relations can overcome insufficiencies in covariational causal reasoning [35]. Instead of basing the relationship on observed quantitative probabilities, mechanistic reasoning focuses on qualitative justification by examining the essential explanatory characteristics of the underlying mechanisms [28]. In the cognitive sense, observed covariation events provide evidence for describing “what” happens in a possible causal process, while the mechanism explains (and predicts) “why” the covariation events would (may) happen.

Altogether, these three principles form the basis of causal reasoning. A complete understanding of the causal relation in a specific domain is best established with all three elements; however, in the process of scientific discovery and learning, it is common for covariation (experimental evidence) or mechanism (conceptual theorization) to take a temporary lead in the progression of knowledge development. Nevertheless, the time element is a necessary *a priori* condition to establish a possible causal process, which is often assumed to be well established among learners.

However, the instruction on causal relations in science domains often makes emphasis on the covariation and mechanism elements and hardly addresses the time element in a causal process. Therefore, through such learning, it is possible for students to become less sensitive to the time condition and focus on the covariation or mechanism elements, which may then lead to incorrect conclusions on causality. In teaching and learning, a rudimentary understanding of a causal relation would require the explicit understanding of the time condition and at least one of the other two elements, while a complete causal reasoning requires a cohesive understanding with all three elements.

### C. Causal reasoning and physics learning

Causal reasoning is essential to knowledge formation in physics [27]. Causal relations used in physics are often described by qualitative representations of the underlying mechanisms and quantitative covariations using variables and formula [34]. However, causal reasoning is rarely discussed explicitly in current physics teaching and is only addressed in a few studies in physics education.

Among the small set of studies looking into causal reasoning in physics education, some researchers have proposed using causal reasoning as a basis for teaching [39,15,40–42]. Studies have also demonstrated that providing students with causal explanations improves students' conceptual understanding in physics [43,44]. In addition, Hung and Jonassen [34] compared the instruction using covariation vs mechanism-based approach in teaching physics with computer simulations. The results show that the mechanism-based instruction, which emphasizes mechanistic causal explanation, is more effective in helping students achieve a better conceptual understanding.

Other studies have looked into issues of causal reasoning in physics students. Besson [15] examined features of causal reasoning used in physics by students, highlighting differences between common and scientific usage of causal reasoning. Mustakim *et al.* [45] demonstrated the important role causal reasoning plays in student problem solving. Viennot [46] found causal explanations play a prevalent role in nearly all conceptual reasoning arguments given by students. The results of these studies suggest that the activation of causal reasoning is a crucial first step toward acquiring physicistlike systemic reasoning.

The studies discussed above demonstrate that causal reasoning plays an important role in students' learning of physics. The connection between causal reasoning and physics learning can be further understood with the cognitive studies reviewed in the previous sections. Cognitively, causal reasoning is considered as a developmental primitive in learning, which contributes to knowledge forming at all levels [27]. Therefore, students would naturally attend to causal relations in learning physics. When causality in a physics concept is not explicitly

discussed in instruction, students are likely to develop their own interpretations of the related causality. As a result, students' causal reasoning ability would influence their physics learning, which is evident from the studies reviewed above. However, most existing studies on causal reasoning and physics learning focus on using but not necessarily enhancing students' causal reasoning abilities. It is then important to understand the nature of how students may develop causal understandings of physics concepts and how instruction may be revised to address students' common difficulties in causal reasoning.

A recent study on students' understanding of Newton's third law emphasizes this point by demonstrating that students' causal reasoning often underpins many intuitive misconceptions [10]. Specifically, assessment and interview outcomes demonstrated a disconnect between students' understanding of causality and the time order of cause and effect. The results also indicated a possible strong connection between students' belief in causality and the use of the action-reaction language. In addition, some students in interviews stated that the larger or more massive object would apply the action force or cause the other force, which indicated a link between asymmetries in the problem scenario (differences in mass, etc.) and their causal views of the problem. However, the previous study was not designed to specifically address the complete set of connections between causality and its attribution conditions. Building off the existing research, this study aims to diagnose more completely student understanding of the relations among causality, time order of events, context features, and the action-reaction language in the domain of Newton's third law. Two research questions will be investigated:

1. How do students understand causality and its connections to time order, context features, and the action-reaction language in the domain of Newton's third law, and how may students' understandings vary with scenario contexts, gender, and population backgrounds?
2. What are students' thinking pathways that are commonly used in their reasoning about causal relations in Newton's third law, and how do their thinking pathways change at different developmental levels of causal reasoning?

In this paper, two terms are used frequently in analysis and discussion, which include "asymmetry" in a contextual setting and students' "thinking pathways." Asymmetry refers to nonuniformity existing in certain contextual features. For example, asymmetry in timeline represents the situation that an event occurs earlier than another event. If two events occur simultaneously, then the timeline is considered symmetric. Asymmetry also exists in contextual features such as mass, velocity, etc. If two parties involved in an interaction have different scales on a feature such as different masses, then asymmetry on the mass feature is assumed.

Thinking or reasoning pathway is a term used in studies on knowledge integration with the conceptual framework model [47,48]. It represents a unique pattern or thread of reasoning that connects contextual features with rules and processes and deeper level conceptual ideas. It is a pathway that students' knowledge can be activated by contexts to form some level of meaningful reasoning and to produce outcomes in responding to a given task.

## II. RESEARCH DESIGN

Causal reasoning has been generally defined to include three elements of understanding including time, covariation, and mechanism. The time is a necessary prerequisite for establishing the sequence of events that may be described with a causal relation. The covariation provides the observed covariation features of the events, while the mechanism offers the mechanistic explanations and predictions of how and why the covariation may occur. For a causal relation to exist, there need to be asymmetries between cause and effect in all three elements of causal understanding, which include the following: (i) asymmetry in time order such that the cause precedes the effect, (ii) asymmetry in covariation such that the observed events of cause and effect have directional covariation changes, and (iii) asymmetry in mechanism such that the mechanistic explanations provide directional conceptual links connecting the cause to the effect.

For the concept of Newton's third law, the central idea is that an interaction between two entities can lead to a pair of symmetric (i.e., equal and opposite) interaction forces being observed [10]. In this case, the two forces are the observed outcomes of an interaction and occur simultaneously, and therefore, there is no causal relation existing between a pair of interaction forces. However, as shown by research [10], students' understandings of N3L often imply a causal relation between the two interaction forces due to their considerations of dominant context features, the timeline of a scenario, and the action-reaction language, each of which provides a sense of asymmetry that can be inappropriately interpreted by students for cuing an inference on causality. Therefore, to establish a model of student understanding of causality in N3L, the three facets of asymmetries will be carefully studied and used as the basis for analyzing student reasoning.

On the surface level, each of the three asymmetries provides a covariational or mechanistic description of N3L interactions in contexts involving collision and push or pull scenarios. For scenario timeline (or storyline), which is the basis for describing covariational phenomena, an object can be placed in a position or state that leads to an interaction, or a person can initiate a push. These are often described as the action party that initiates an event early in the scenario timeline. In such cases, the force exerted by the action party is often considered by students as the cause of the other force. Similarly, one of the two objects involved in an

interaction can have certain dominant features such as traveling with a larger speed, having a larger mass, or being stronger. In these cases, the force exerted by the dominant party is often considered as the cause of the other force. Finally, the action-reaction language is often interpreted by students as an implication of causality such that the action force is considered to cause the reaction force. In addition, students often interpret this action-reaction language in resonance with dominant features and scenario timeline, with which the dominant and/or action party is considered to apply the action force, and the action force is considered to cause the reaction force. In this way, the action-reaction language provides a sense of mechanism for explaining the causal relation between the interaction forces (or action-reaction forces as traditionally defined), which is unfortunately the only meaningful description that implies a mechanistic explanation of causality in N3L among most textbooks.

In this study, the assessment questions are designed to explicitly probe four reasoning attributes including the following: (i) the causal relation between a pair of interaction forces, (ii) the magnitudes of interaction forces (dominance features), (iii) the time order of interaction forces (scenario timeline), and (iv) the action-reaction language. Among these, the attributes on magnitude and timeline are perceived observations of N3L interactions, and therefore, provide measurements on the covariation understanding of the causal relation. The timeline also provides a context to measure students' understanding of the time order of the interaction forces. Meanwhile, the action-reaction language describes an abstract level interpretation of the N3L interaction, which probes the mechanistic side of the understanding.

### A. Assessment design

Following the objectives of the measurement, a 12-question multiple-choice test on causal reasoning of N3L (CRN3L) was designed, which includes three N3L scenarios that each contains four questions corresponding to the four attributes (see Fig. 1 for the scenarios and the Appendix for the complete test). The first scenario provides a collision setting, in which two football players with different masses run into each other and a pair of interaction forces are observed. Students were asked to determine the causal relation between the interaction forces, compare the magnitudes and time order of the interaction forces, and identify the action-reaction forces. This scenario provides a symmetric timeline, as most students recognize that the two parties get into the collision simultaneously. However, since the two parties involved in the collision have different masses, the question presents a situation of asymmetry due to a dominant feature, which may influence other attributes of students' causal reasoning. Therefore, the design of this question will probe specifically the effect of dominance-based asymmetry on students' reasoning.

*Collision:* In a soccer game, two players, A and B, are running to chase a ball. Player A has twice the mass as Player B. They are both running at the same speed towards each other. Unfortunately, neither player notices the other and they collide into one another. During the collision, each player exerts a force on the other player. The force exerted by player A on player B is  $F_{AB}$ , while the force exerted by player B on player A is  $F_{BA}$ . Consider these two forces between the two players and answer the following questions.

*Book:* A textbook is at rest on top of a horizontal table. The book exerts a downward force  $F_{BT}$  on the table, while the table exerts a supporting force  $F_{TB}$  on the book. Consider these two forces between the book and the table and answer the following questions.

*Push:* Two students, A and B, are on identical low friction roller skates facing each other. They both have the same mass of 50 kg. A puts her hand on B and pushes outward causing both to move. While A is pushing B with a force  $F_{AB}$ , A also feels a force  $F_{BA}$  pushing back on her by B. Consider these two forces between A and B and answer the following questions.

For each of the above scenarios, the following four questions are asked.

1. Is there a causal relation between the two forces?
  - 1) Yes,  $F_{AB}$  causes  $F_{BA}$ .
  - 2) Yes,  $F_{BA}$  causes  $F_{AB}$ .
  - 3) No, there is no causal relation between the two forces.
2. Compare the magnitudes of the two forces.
  - 1)  $F_{AB}$  is larger than  $F_{BA}$ .
  - 2)  $F_{BA}$  is larger than  $F_{AB}$ .
  - 3) The magnitudes of the two forces are equal.
3. Does one of the forces occur before the other force?
  - 1) Yes,  $F_{AB}$  occurs before  $F_{BA}$ .
  - 2) Yes,  $F_{BA}$  occurs before  $F_{AB}$ .
  - 3) No, both forces occurred simultaneously.
4. Which force is the action force, and which one is the reaction force?
  - 1)  $F_{AB}$  is the action force and  $F_{BA}$  is the reaction force.
  - 2)  $F_{BA}$  is the action force and  $F_{AB}$  is the reaction force.
  - 3) It cannot be determined which force is the action or reaction force.

FIG. 1. Design of assessment questions on causal reasoning in Newton's third law. Collision, book, and push are three scenarios of N3L interactions. For each scenario, the same four questions are asked to probe the four attributes of causal reasoning regarding the causality, magnitude, time order, and action-reaction description of the N3L interaction forces. The complete test is shown in the Appendix.

The second scenario in Fig. 1 involves a book placed stationary on a horizontal table. Students were asked again the same four questions. This design has an implied asymmetry that students often have in their reasoning, considering that the actual placement of the book causes the interaction. This leads to the thinking that the force applied by the book on the table is the cause and the action force and that the force applied by the table on the book is the effect and the reaction force.

The third scenario is designed with the situation that one person pushes the other (the push scenario in Fig. 1), and the same four questions were asked. This design has a more pronounced asymmetry of timeline, with which students often think that the person who initiates the push (the action party) causes the interaction, and thus, the force by this

action person is the cause and the action force, and the force by the person being pushed is the effect and the reaction force.

In responding to these questions, expert level students would be able to understand that there is not a causal relation between the two interaction forces. Although the scenario may have an asymmetric timeline of events that lead to the interaction, the actual interaction is simultaneous for both interacting parties. Therefore, the two interaction forces would be equal in magnitude and occur simultaneously. Because of the nature of the symmetry in an N3L interaction, the traditionally defined action or reaction force cannot be uniquely determined. Therefore, expert level students would answer all questions with choice 3 in all scenarios.

In contrast, novice students often misinterpret the asymmetries in dominant features or event timeline to infer a causal relation between the interaction forces, which is further enhanced by the action-reaction language [10]. These students are likely to answer the questions with choice 1 in all scenarios. Between novice and expert levels, there exists a range of intermediate developmental levels. Students in these stages may answer questions with a combination of expert and novice type thinking, which is also highly context dependent [6,10,49]. It is then useful to analyze the response patterns of students at different developmental levels, which can help to identify popular thinking pathways developed at different stages of learning.

### B. Data collection

The multiple-choice test of causal reasoning of N3L (CRN3L) was given to a group of college students ( $N=258$ ) and a group of high school students ( $N=395$ ) in China. All students took the whole test at one time during the fall semester of 2020. The college students were second-year engineering majors in a large comprehensive university with a Chinese national ranking around top 200. The high school students were in the second year of their three-year high school program, which is equivalent to grade 11 in the USA. The high school has a medium ranking and is in a small city near the university. The selection of the population was aimed at medium-ranking students in order to probe a wider variety of thinking and reasoning from students at different developmental stages.

The general population background of the two groups of students is comparable. The university and the high school are in the same region, and their main body of students are from a similar social-economic background. On average, half of the students graduated from the high school are able to go to the university involved in this study or to other universities with a similar ranking. The remaining high school graduates will go to colleges or universities with a lower ranking. Therefore, the college students tested in this study should have a slightly higher level of general academic performance than the high school students would have.

Both the college and high school students had completed their learning of N3L during their first year of high school in algebra-based high school physics courses. The college students had continued to study in a calculus-based college physics course during their first year in college, in which N3L was assumed to be pre-established and was applied in problem solving. But the features of N3L were not as explicitly emphasized as they were in the high school physics curriculum. For both student groups, the assessment can be considered as a delayed post-test, with a longer delay for the college students.

In addition to quantitative assessment, think-aloud interviews were conducted with students solicited from the same population pools, which included 13 high school

students and 14 college students. These students were not involved in the quantitative testing to avoid cross contamination. During the interviews, students were given the same test and were asked to respond in a think-aloud mode. Further clarification and explanation were also prompted when needed. Each interview lasted for approximately 20 min, and students were each given a small gift for their participation.

### C. Data analysis

Students' responses to the CRN3L test are analyzed to produce statistical descriptions of the students' understanding of the four attributes of causal reasoning, which provide the quantitative results for answering the first research question. The analysis uses a combination of descriptive statistics and clustering tools to identify significant associations between students' answers on the four causal attributes in different scenarios. The results also provide evidence to identify influences on students reasoning from differences in scenario contexts and population backgrounds.

In addition, students' response patterns on different questions are first sorted based on their popularity. The popular patterns are then analyzed to identify common features in student reasoning. Furthermore, the response patterns are also analyzed with latent clustering analysis (LCA) [50], which provides quantitative evidence on common patterns of students' thinking. The quantitative analysis outcomes are triangulated with interview results to identify and validate the possible thinking pathways and to establish a progression of developmental levels of student causal reasoning, which provide evidence to answer the second research question.

## III. RESULTS

### A. Student responses on the four attributes of causal reasoning in N3L

As discussed in the method section, the assessment questions are designed to explicitly probe four causal reasoning attributes including causality, dominance, time, and the action-reaction language. Students' responses to the questions of the four attributes in different scenario contexts are summarized in Table I. The average percentages of answers to different attributes are also plotted in Fig. 2 for easy comparisons.

For the answers, choice 1 represents the novice level understanding, considering the force applied by the active and dominant party causes the other interaction force. In addition, students at this level may also think that this "causing" force would have a larger magnitude, occur earlier, and be the action force. In contrast, expert level students would answer with choice 3, recognizing no causality between the two interaction forces and considering that the two forces would have the same magnitude,

TABLE I. Summary of high school and college students' answers on the four attributes of causal reasoning in different scenario contexts. The numbers are percentages of students choosing the corresponding answers.

Scenario contexts	Answers	High school ( $N = 395$ )				College ( $N = 258$ )			
		Causality	Magnitude	Time order	Action-reaction	Causality	Magnitude	Time order	Action-reaction
Collision	1	21.8	11.9	5.3	18.7	38.8	23.3	18.6	30.6
	2	6.3	3.0	3.3	4.3	7.0	8.1	7.4	10.9
	3	71.9	85.1	91.4	77.0	54.3	68.6	74.0	58.5
Book	1	38.0	1.5	8.1	32.2	58.5	16.7	23.3	52.7
	2	5.1	4.8	4.8	4.1	9.7	11.2	12.0	12.4
	3	57.0	93.7	87.1	63.8	31.8	72.1	64.7	34.9
Push	1	38.0	5.8	13.2	36.5	56.6	19.0	30.2	52.3
	2	4.1	3.8	3.8	2.8	10.1	10.9	8.9	9.7
	3	58.0	90.4	83.0	60.8	33.3	70.2	60.9	38.0
Total	1	32.6	6.4	8.9	29.1	51.3	19.7	24.0	45.2
	2	5.2	3.9	4.0	3.7	8.9	10.1	9.4	11.0
	3	62.3	89.7	87.2	67.2	39.8	70.3	66.5	43.8

occur simultaneously, and be undetermined with action or reaction if such languages is to be used. As seen in the results, most students answered with choices 1 and 3, suggesting that the expected novice and expert types of thinking have captured the reasoning of most students. Meanwhile, choice 2 is much less popular, selected by less than 10% of the students. This answer reflects the thinking of a causality due to the nondominant force, which is inconsistent with the common novice understanding. It is possible that these responses were chosen by students through random guessing; however, the origin of this kind of reasoning was not identified through interviews due to the small sample size and will not be discuss in detail.

The results in Fig. 2 show that the responses of the high school and college students on the four attributes have similar distributions, with high school students having

more expertlike answers. A two-way ANOVA indicates main effects from population groups [ $F(1) = 225.5$ ,  $p = 0.000$ ] and attributes [ $F(3) = 105.6$ ,  $p = 0.000$ ] without an interaction [ $F(3) = 0.389$ ,  $p = 0.761$ ], which confirms the similarity of response distributions between high school and college students. The results suggest that both groups of students have similar reasoning on the different attributes, with more high school students picking the expert answers (choice 3). The differences may be the result of forgetting. The high school students had just completed learning the N3L content in the previous academic year, while the college students learned the same material back in high school three years earlier. For the college students, their college physics courses emphasized more advanced topics and did not allow much time to review the basic Newton's laws in detail.

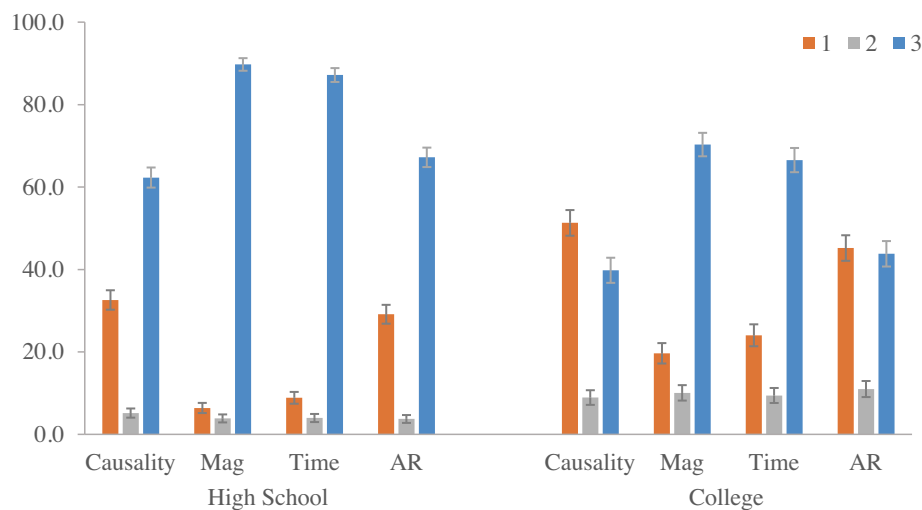


FIG. 2. Average percentages of high school and college students' answers on the four attributes of causal reasoning including causality, magnitude (Mag), time, and the action-reaction language (AR). The error bars reflect the standard errors.

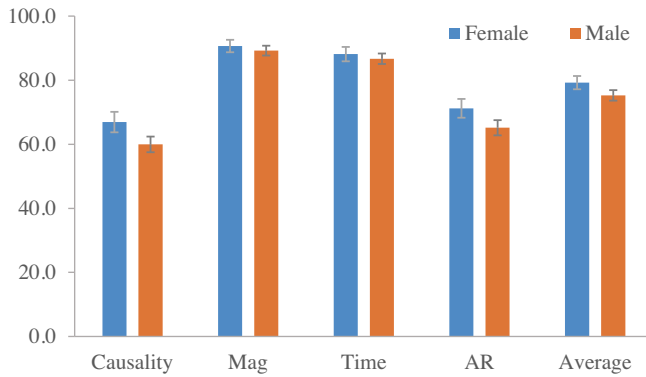


FIG. 3. Average percentages of correct answers on the four attributes of causal reasoning from male and female high school students. The error bars reflect the standard errors.

Looking into the different attributes, students were generally performing better on magnitude and time as shown in Fig. 2.  $t$  tests show no significant differences between students' answers on magnitude and time and between causality and action-reaction within each population ( $p > 0.08$ ,  $d < 0.05$ ). Meanwhile the differences are significant when comparing the combined answers on causality and action-reaction with magnitude and time within each population ( $p = 0.000$ ,  $d > 0.23$ ). The results suggest that students recognize the equal magnitude and simultaneity of the two interaction forces, which is consistent with the finding in a recent study [10]. As suggested in the study, the correct understanding of equal magnitude is likely the outcome of memorizing the definition of N3L, while the recognition of simultaneity is based on observations and intuitive interpretations of N3L events. However, over 1/3 of the high school students and nearly half of the college students still believe there is a causal relation between the interaction forces, and similar fractions of students also consider the action force being the cause. This implies a possible strong connection between students' thinking of causality and the action-reaction language, which will be further investigated in the next section.

To explore if male or female students may respond to the questions differently, the gender-based analysis is

conducted with the high school population, which has 132 female and 263 male students. The college population includes primarily engineering majors with very limited female students, and therefore, is not used in the analysis. The gender-based results are shown in Fig. 3, which indicates no differences between responses from male and female students across all four reasoning attributes and the average outcome ( $p > 0.08$ ).

From the results in Table I, students' answers appear to be influenced by the scenario contexts. To compare more clearly the effects of different contexts on students' considerations of causality, the percentages of answers with choice 1 on the different scenario contexts are compared in Table II. It is noted that the small  $p$  values in comparing high school students' answers on the book and push questions and on some attributes are largely due to near-zero percentages, which lead to small variances. Therefore, the data from college students can provide more meaningful implications. In general, students' responses on the book and push questions are similar but are very different from their responses on the collision question. Specifically, more students responded with a belief in causality, asymmetry in time, and explicit identification of an action force on the book and push questions than on the collision question. This suggests that the asymmetry in the timelines of the book and push questions may have contributed to more students believing in causality with these two questions than with the collision question which has a symmetric timeline. For the collision scenario, the symmetry in time is naturally observed and used by many students in their reasoning, leading to a decision of non-causality. In contrast, students' answers on the magnitude questions are less influenced by the timeline feature of the scenario contexts, but with slightly more students responded with asymmetric forces in the collision problems, a likely effect of the dominant feature from the asymmetric mass.

The descriptive results discussed in this section are consistent with the expectations of the assessment design, which capture both the expert and novice level thinking. The results also provide descriptive answers to the first research question. Novice students appear to have a belief

TABLE II. Percentages of high school and college students who responded with choice 1 on the four attributes of causal reasoning in different scenario contexts. Book and push represents the average percentage of responses to the book and push question. The  $p$  value indicates the statistical significance of the difference between the means.

Scenario context	High school ( $N = 395$ )				College ( $N = 258$ )			
	Causality	Mag	Time	AR	Causality	Mag	Time	AR
Collision	21.8	11.9	5.3	18.7	38.8	23.3	18.6	30.6
Book and push	38.0	3.7	10.7	34.4	57.6	17.9	26.8	52.5
$p$ value	0.000	0.000	0.005	0.000	0.000	0.125	0.027	0.000
Book	38.0	1.5	8.1	32.2	58.5	16.7	23.3	52.7
Push	38.0	5.8	13.2	36.5	56.6	19.0	30.2	52.3
$p$ value	1.000	0.001	0.020	0.203	0.663	0.495	0.076	0.928



TABLE III. Percentage ranking of students' 4-question response patterns on the four attributes of causal reasoning in different scenario contexts. The patterns are ordered based on their popularity among college students.

Response patterns	High school ( $N = 395$ )				College ( $N = 258$ )			
	Collision	Book	Push	Average	Collision	Book	Push	Average
3333	63.3	49.4	52.4	55.0	44.6	24.4	27.9	32.3
1331	6.3	17.7	18.7	14.2	8.1	21.3	20.5	16.6
1111	2.5	0.8	2.3	1.9	11.6	12.4	15.1	13.0
1333	5.6	8.9	4.3	6.3	8.5	8.1	5.4	7.3
1311	1.0	4.1	7.3	4.1	0.4	6.2	8.1	4.9
3331	3.3	6.3	3.3	4.3	2.3	5.4	3.5	3.7
Others	18.0	12.9	11.6	14.2	24.4	22.1	19.4	22.0

in causality between the interaction forces, which is likely associated with the action-reaction language. Meanwhile an understanding of symmetry in magnitude and time order of the interaction forces appears to have been established among many students. The high school students generally perform better than the college students, which is likely due to the college students forgetting the content of N3L learned in their high school course. Based on the high school population, there appear to be no gender differences in students' thinking of the four attributes of causal reasoning. Finally, the scenario contexts have presented obvious influences on students' reasoning, suggesting that the time-line asymmetry may be a significant feature leading to students' belief in causality.

### B. Connections among the four attributes of causal reasoning in N3L

In this section, students' response patterns and interviews will be analyzed in detail to investigate the connections in student reasoning among the four attributes and to identify possible students' thinking pathways underlying their answers. First, a ranking analysis of students' response patterns on the four attributes was conducted and is shown in Table III. Students' answers to the questions on the four attributes in a scenario are combined to form a response pattern. For example, if a student answers with choice 1 on the causality, 3 on the magnitude, 3 on the time, and 1 on the action-reaction language, a pattern of "1331" is then assigned to the student. In Table III, the patterns are sorted from top to bottom based on their popularity in the college population, which has a richer set of intermediate states. The top six patterns are shown along with "others" containing the rest of the patterns combined. Among the popular patterns, there is none involving choice 2, which indicates that responses with choice 2 may likely be random answers that do not form consistent answer patterns to gain popularity. For statistical considerations, since each question has three choices, the probability for a four-question pattern to occur by random chance is  $1/81$  (1.23%). A  $\chi^2$  analysis indicates that a percentage larger

than 4.0% (for  $N = 258$ ) or 3.0% (for  $N = 395$ ) is statistically significant at the  $p < 0.05$  level.

As shown in Table III, the pattern of "3333", which represents the expertlike thinking, is the most popular one in both populations, with over half of the high school students and nearly one third of college students answering in this pattern. The novice-thinking pattern of "1111" is minimum among the high school students but is the third popular one among the college students, suggesting a consistent subgroup of novice thinkers. Interestingly, the second most popular pattern for both populations is 1331, which is a mixed pattern with expertlike answers on the magnitude and time questions but novice answers on the causality and action-reaction questions. This result suggests that for this subgroup of students the instruction on the equal magnitude is effective but does not seem to improve students' reasoning on causality. Meanwhile, the explicit identification of an action force is consistently associated with a belief in causality, indicating the thinking that attributes the action force as the cause of the reaction force. In addition, this subgroup of students also demonstrates a disconnect in their thinking of time order and causality, which is also reflected in the pattern of "1333." These students recognize that the interaction forces would occur simultaneously, however, they also consider that one of the interaction forces is the cause of the other. These types of answers indicate a serious deficit in understanding the most fundamental condition of a causal relation.

The three less popular patterns listed in Table III also reflect unique reasoning among students who are in a variety of intermediate mixed stages between novice and expert. For example, the pattern of 1333 suggests that although students seem to have correct understandings of the magnitude, time, and action-reaction attributes of N3L, they still hold a belief in causality between the two interaction forces, without realizing the needed asymmetry in time order for establishing a causal relation. Meanwhile, the 3331 pattern reveals the strong influence of the action-reaction language on students who would otherwise have achieved expertlike understanding. On the other hand, the "1311" patterns shows the beginning of students'

TABLE IV. Percentage ranking of students' 12-question response patterns on the four attributes of causal reasoning in different scenario contexts. The patterns are ordered based on their popularities among college students. Patterns not shown are all with popularities less than 2%.

Patterns	High school	College
333333333333	38.2	17.1
111111111111	0.5	8.1
333313311331	4.3	6.6
133113311331	2.3	2.3
333333331331	3.3	1.9
133313311331	0.8	1.6
Others	50.6	62.4

transitioning from the novice state, which starts with the correct understanding of the magnitude likely due to the instructional emphasis of the equal-and-opposite property of N3L. These mixed patterns demonstrate that the progression of learning often goes through a wide range of intermediate stages, in which students develop a variety of fragmented knowledge structures [47,51,48]. Seeing the students' difficulties, it is then important to make emphasis on teaching the correct causal reasoning in the context of N3L in order to help students develop an integrated knowledge structure and achieve a deep understanding of N3L [10].

The four-question patterns are formed with questions within each scenario context. To explore how the response patterns may vary with different contexts, the complete twelve-question patterns are also analyzed. The top 6 most popular patterns based on the college population are listed in Table IV. Because of the large number of possible combinations, the occupation frequencies of individual patterns are much smaller than that of the four-question patterns, however, the probability for a pattern to occur by random chance is also very small (in the order of  $10^{-6}$ ), which makes repeated and similar patterns statistically valid for inferences on possible cognitive traits.

As seen from Table IV, the all-3 and all-1 patterns are two most popular ones for the college population indicating consistent subgroups holding the expert and novice views. Meanwhile, the results of the high school students show a much larger all-3 pattern with only two students selecting the all-1 pattern, revealing that more high school students consistently responded with the expert-like answers than the college students did. Both populations show various mixed patterns that are combinations of 1331 and 3333, which are the two most popular four-question patterns. While this result clearly demonstrates the influence of the scenario context, it also indicates that the student reasoning exhibits more consistency within each scenario context than between different scenarios. For example, the fourth pattern in Table IV shows a subgroup of students consistently using the 1331 type of thinking in all three

scenarios. In comparison, the third pattern suggests a subgroup starting to use the 3333 type of thinking in the collision scenario, while keeping the 1331 reasoning in the book and push scenarios. This difference demonstrates the influence on student thinking from the time-line asymmetry, which exists in the book and push questions but is absent in the collision question.

The ranking of the response patterns provides a simple method to identify popular patterns that are directly observed in testing; however, due to the large number of possible combinations, many less popular patterns, which may only differ slightly from the popular ones, are often difficult to analyze. For a more comprehensive analysis, clustering methods, which combine similar patterns into a single cluster class, are often used. In this study, the latent class analysis (LCA) is used to identify common cluster classes of students' twelve-question response patterns.

LCA is part of the family of latent variable models, which assume that a latent variable is responsible for the relations between the observed variables [50]. These analysis methods are often used to characterize student patterns of mental model in science education [52–54]. Latent class models are defined by the unconditional probabilities of belonging to each class and the conditional probabilities of each answer option given that a student belongs to a specific class. To identify the best fit, models with different numbers of classes are calculated, and the best fitting model is determined based on achieving the lowest Bayesian information criterion (BIC) and a large relative entropy ( $>0.90$ ) [55,56]. In this study, the LCA was performed using the *poLCA* R Package [57].

The LCA analysis was performed separately on the datasets from the high school and college students. Each dataset was classified into five different latent models with increasing numbers of classes (1-, 2-, 3-, 4-, and 5-class models). The fitting parameters are listed in Table V, which suggests that a 4-class model is optimal for the college dataset, and a 3-class model is optimal for the high school dataset. For both datasets, their first two most popular classes are similar. The third class of the high school dataset appears to be a mixture of classes 3 and 4 from the college dataset, which are combined due to small probabilities. Therefore, for a more direct comparison, the four classes of

TABLE V. Latent class analysis goodness of fit measures for the college and high school datasets.

Number of classes	College			High school		
	df	BIC	Entropy	df	BIC	Entropy
1	234	5495.4	...	371	5763.8	...
2	209	4165.9	0.981	346	4873.4	0.901
3	184	3980.3	0.981	321	4645.3	0.926
4	159	3832.1	0.946	296	4677.1	0.919
5	134	3886.5	0.949	271	4709.4	0.913

TABLE VI. Class probabilities using the 4-class model fit for college and high school students.

Classes	College	High school
	Probability	Probability
1	0.301	0.534
2	0.392	0.365
3	0.171	0.018
4	0.135	0.084

the college students' 4-class model are used as the bases for both the college and high school datasets. The probabilities of belonging to each class are then calculated and listed in Table VI.

The response patterns of the four classes are plotted in Fig. 4, which show similar features of the popular patterns listed in Table IV. Class 1 matches well with the all-3 patterns shown in Table IV, which contained students who answered most questions consistently with the expert view. Class 2 matches the common factor of many combinations of the 1331 pattern, which indicates a connection between student understanding of a causal relation and the identification of action and reaction forces. Class 3 contained students who consistently answered all questions using the novice thinking (the all-1 pattern in Table IV). Class 4 contained students who answered most questions using a mixture of the novice view and the less meaningful choice 2 answer, which indicates the possibility of guessing or random selection.

The agreement between the LCA clusters and the popular response patterns provides additional statistical

support for using the response patterns to analyze student thinking. To further validate the implications of students' responses and to explore the details of students' actual reasoning, student interviews are analyzed in the next section to identify the thinking pathways of students with the different response patterns.

### C. Student thinking pathways and developmental levels of causal reasoning in N3L

Think-aloud interviews were conducted with 13 high school students and 14 college students solicited from the respective population pools. These students did not take the quantitative assessment and were given the same questions to solve in the think-aloud mode with prompts for further explanation when needed. The students' interviews were analyzed to identify the detailed thinking pathways at different levels of reasoning between expert and novice. Both college and high school students appeared to have similar reasoning, and their interviews were analyzed together without distinguishing the population background.

#### 1. Expertlike thinkers

From the interviews, students who answered the questions with the 3333 pattern also appeared to have expertlike reasoning and explanation. For example, in the collision scenario, a college student stated "*there is no causal relation between the two forces. The two forces should happen simultaneously. Each of the forces is the action and the reaction force; they are the interaction forces and are equal and opposite.*" Some other students explicitly referred to the N3L: "*I think there is no causal relation.*

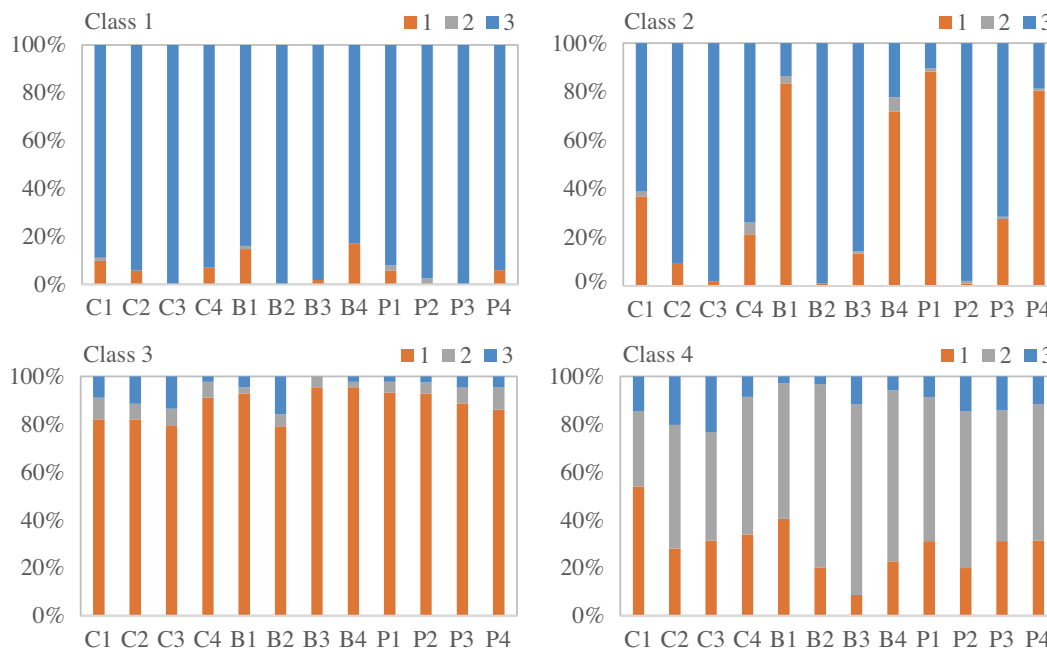


FIG. 4. The response patterns of the four classes obtained based on the four-class LCA model of the college dataset. C1-4, B1-4, and P1-4 represent the four questions in the three scenarios of collision, book, and pushing, respectively.

According to N3L, the forces are equal and opposite. Therefore, they should not have a time difference. They are interaction forces, and each can be the action force depending which one you are looking at. If you are looking at A, then A is the action and B is the reaction, and vice versa." Apparently, these students recognized the symmetry in time and used it as their basis for determining the noncausal relation. The asymmetry in the dominant feature of mass does not seem to impact these students' reasoning.

For the book and push scenarios, the design of the time-line asymmetry can confuse certain students as the time difference between the interaction forces. However, the expertlike students appeared to have clearly distinguished the time-line asymmetry from the simultaneity of the interaction forces, which is evident from the following interview excerpts: "(on the push question) ... from the logic of the story, A should have pushed first, which caused A applied by an oppositive reaction force. But looking from the physics process, the two forces would occur simultaneously, and do not have a causal relation. They are both the action force and the reaction force and are equal and opposite." Other students did not explicitly comment on the time-line asymmetry but clearly identified the simultaneity of the interaction forces: "(on the book question) ... there is no causal relation. Once your put the book down, the two forces would just occur simultaneously. Therefore, there is no time difference. They are interaction forces and have an equal magnitude."

For these expertlike students, their reasoning pathways on causality, magnitude, time, and action-reaction may slightly differ from one another but are all structured in coherent and well-integrated logical forms. They are able to distinguish between the asymmetry in scenario timeline and the simultaneity of interaction forces and use the simultaneity in their determination of a noncausal relation. Their identification of action and reaction forces is flexible and relative to the changing perspective of interpretation, reflecting a consistent understanding of the nature of interaction. Additionally, these students' consideration of equal magnitude is a certainty, which is often commented last without much emphasis in their decision on causality.

## 2. Novice thinkers

The novice thinkers, who answered the questions with the 1111 pattern, appear to have consistent reasoning based on asymmetries in the dominant feature and scenario timeline. In the collision scenario, students often considered that "... there must be a causal relation, because A is twice heavier and should be stronger, A will exert a larger force. Since A is heavier than B, A's force should be the first to occur and is the action force. The force by B is the reaction force." Apparently, students were using the asymmetry in mass as a reason for determining causality and other attributes including magnitude, time, and action-reaction. Their understanding of the connections between

the asymmetry in mass and other attributes appear to stay at the factual statement level without further explanation other than that the heavier person could be stronger. This indicates a lower quality of reasoning that lacks deeper level mechanistic richness and overall logical consistency.

In responding to the book and push scenarios, these students consistently used the scenario timeline as the reason to determine a causal relation: "(on the book question) ... the book applies a force on the table, which causes the table to reaction and provide a support force. There is a causal relation. The two forces have a time difference. The book's force occurs first. The book's force is the action force, and it should be larger than the support force by the table." "(on the push question) ... there is a causal relation. It is because A pushes B, and then B pushes back as a reaction. Therefore, A's force occurs first and is the action force. B's force is the reaction force. The two forces are in opposite directions, and the magnitudes, ... let me think. The magnitude of the A's force should be larger, and B would be pushed a bit farther in the distance."

Based on the interview results discussed above, novice students seem to consistently rely on the asymmetry due to a dominant feature or scenario timeline in their reasoning to determine a causal relation. The form of their reasoning is rather simplistic, which directly matches the asymmetry to causality without further mechanistic explanations available. However, the connection between the recognition of an asymmetry and the belief in causality is strong and persistent.

## 3. Intermediate thinkers

As shown by the response patterns in Table III, there exists a group of intermediate level students with a wide variety of mixed types of reasoning. The most popular one is the "1331" pattern: "(on the collision question) ... there is a causal relation. From the N3L point of view, the two forces are equal and opposite. When they collide, both would contact at the same time, and the two forces would occur simultaneously. The two persons will both apply a force to the other. The force applied by the person with a larger mass is the action force, and the one with less mass applies the reaction force. The action force is the cause, and the reaction force is the effect." Apparently, this student considered the two forces are simultaneous, demonstrating a disconnect between causality and time. Additionally, this student explicitly linked the action force as the cause of the reaction force, indicating a strong connection between the thinking of causality and the action-reaction language. For this student, the equal and opposite idea of the interaction forces is well established and not influenced by the difference in mass; however, the asymmetry in mass is still used as the reason to determine the action force, which leads to the belief in causality.

For the book and push scenarios, students seem to recognize the simultaneity of the interaction forces but

are still using the asymmetry in the scenario timeline as the reason for determining causality: “(on the book question) ... because you first put the book on the table, then the table provide the book with a support force; therefore, there is a causal relation. But there is no time difference between the two forces, i.e., when the book is put on the table, the two forces would occur simultaneously. The book gives an action force to the table, and the table gives the book a reaction force. The two forces have the same magnitude and in opposite directions.” “(on the push question) ... there is a causal relation. Because A pushes B first, so that A would then feel the reaction force applied by B. A applies the action force, and B applies the reaction force. I think there is no time difference between the two forces. When they contact each other, the forces would occur simultaneously. The two forces are equal and opposite.”

For these students, they are successful in recognizing the simultaneity of the interaction forces but still use the asymmetry in scenario timeline to determine the causality. In this process, the action-reaction language is again used as the explanation that connects the scenario timeline to causality. What is encouraging is that these students, unlike the novice thinkers, seem to be able to distinguish between the asymmetry of scenario timeline and the simultaneity of the interaction forces; however, they still arrive at a belief in causality based on the timeline asymmetry combined with the action-reaction language.

Other types of mixed reasoning, such as the 3331, 1333, and 1311 types, have similar traits as the 1331 type. The 3331 type students appear to be very close to the expert level. The only thing they are missing is the action-reaction language. But they are not using this in their reasoning of causality: “(on the push question)... there is not causal relation, because the two forces occur simultaneously and are equal and opposite ... A pushes B so A’s force is the action force, and B’s force is the reaction force.” The action-reaction language seems to be used simply as a way to describe the interaction forces, which follows the traditional definition.

For the 133 type, students recognize the action-reaction language is a way to describe interaction forces but still use the asymmetry in scenario timeline to determine the causality: “(on the push question)... there is a causal relation because A pushes B, which causes B to push back. A’s force is the cause, and B’s force is the result. The two forces do not have a time difference. They occur simultaneously and are equal and opposite. ... Depending on who you are looking at, both forces are action and reaction forces. ...”

On the other hand, the 1311 type students are close to the novice thinkers. They only recognize the equal magnitude but have similar reasoning as the novice students on other attributes: “(on the book question)... there is a causal relation. The book applies a force first, and then the table applies the support force. The book’s

force occurs first and is the action force. ... The two forces are equal and opposite.”

From the interview results discussed above, there appears to be a clear progression in students’ reasoning from the novice type to the many intermediate types, and to the expert type. Meanwhile, students at different developmental levels of reasoning seem to present unique combinations of response patterns, which can be used to help identify such levels.

#### 4. Developmental levels of causal reasoning in N3L

To further explore how the different response patterns may reflect the developmental levels, the percentage distributions of the six popular four-question patterns are computed for students at different overall performance levels. This analysis combines both college and high school datasets ( $N = 653$ ). The total score of the 12-question assessment is binned into six performance levels as shown in Table VII. For students within each performance level, the percentages of the six response patterns are computed and listed in Table VII. The results are also plotted in Fig. 5 to show more clearly how each pattern may vary at different performance levels. Here, the primary goal of the analysis is to identify unique response patterns and thinking pathways that can signify the important steps in the development of causal reasoning in N3L. It is also noted that all students took the assessment at one time, which produces a cross-sectional measure of the students at different developmental stages. The development of causal reasoning was derived from the salient patterns of students from different groups of test performance, under the assumption that these groups represent different stages of causal reasoning.

TABLE VII. Distribution of the 4-question response patterns from students at different overall performance levels. The data combines both college and high school datasets ( $N = 653$ ). Students are grouped in 6 levels based on their total scores on the 12-question test. The second level is set with a wider score range of 3 points to accommodate for the smaller number of students in that range. The numbers are the percentages of individual patterns normalized based on the total possible patterns among all three scenarios within each performance level. Values that are important in identifying the reasoning levels are each marked with an asterisk (\*).

Reasoning levels	Lower		Upper		Expert	
	Novice	intermediate	intermediate			
Subgroup N	65	55	63	117	115	238
Score range	0–1	2–4	5–6	7–8	9–10	11–12
3333	0	2	4	21*	42*	94*
3331	0	0	1	7*	12*	2
1333	0	1	5	11*	17*	3
1331	0	10*	46*	41*	15*	0
1311	1	13*	15*	8	2	0
1111	52*	11	1	1	0	0
Others	48	62	29	12	11	1

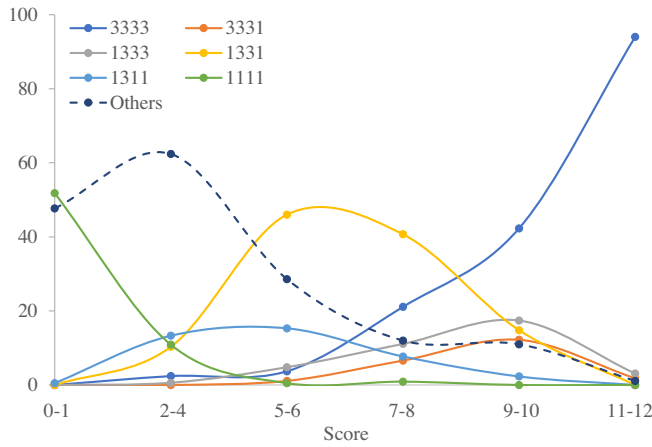


FIG. 5. The pattern response curve of the percentage distributions of the 6 popular four-question patterns for students at different overall performance levels. The single dashed line represents the combined outcome of other patterns.

Based on the results in Table VII and Fig. 5, student reasoning can be sorted into four developmental levels including novice, lower intermediate, upper intermediate, and expert, which are also indicated in Table VII. As is clearly shown, the pattern of 1111 is an obvious indicator of the novice thinkers. Additionally, many students at the novice and lower intermediate levels also answered with other less popular patterns, which indicates less consistent thinking and a larger possibility of guessing. The transition to intermediate levels starts with more students responding with the 1311 and 1331 patterns. The 1311 pattern suggests an initial transition likely due to the effect of instruction on the equal-and-opposite feature of interaction forces in N3L. On the other hand, the 1331 pattern appears to be a significant sign of students achieving a meaningful progress in reasoning. As shown in Fig. 5, the 1331 pattern has an obvious peak among the intermediate level students, indicating an important developmental step in reasoning.

For more advanced students, the 3331 and 1333 patterns start to get popular. The distributions of these two patterns in Fig. 5 are nearly parallel, suggesting that they start to appear and change similarly with students’ overall performance. From the interviews, students responding with the 3331 pattern seem to have slightly more advanced reasoning than the students with the 1333 pattern. The students

with the 3331 pattern were able to clearly distinguish between the asymmetry in scenario timeline and the simultaneity of the interaction forces, and they used the simultaneity to determine the noncausal relation. The only thing these students were missing was the action-reaction identification, which were interpreted as the traditional definition of interaction forces. In comparison, the students answering with the 1333 pattern were also able to distinguish the asymmetry in scenario timeline from the simultaneity of the interaction forces; however, these students used the asymmetry in scenario timeline to determine a causal relation. Performance wise, the average score of students with the 3331 pattern is slightly higher than that of the students with the 1333 pattern (76.8% vs 73.4%), which is on the borderline of statistical significance ( $p = 0.034$ ,  $d = 0.248$ ). Summarizing the analysis, the results provide stronger support to consider the 3331 pattern as reflecting a more advanced level of reasoning than the 1333 pattern. Nevertheless, these two patterns emerge among students with more developed reasoning that approaches expertlike thinking, and therefore, can signify a higher intermediate level of reasoning being achieved.

Finally, the 3333 pattern represents the expert level students. Students at the higher intermediate level also start to answer with the 3333 patterns on some of the scenarios. As students develop more integrated understanding and reasoning, which unifies the contextually fragmented elements their thinking, more students will consistently respond with the 3333 pattern across all scenario contexts and become a comprehensive expert thinker.

Synthesizing the analysis on the interview results and student response patterns, the popular response patterns seem to signify important steps of developmental levels in students’ causal reasoning in N3L. As shown in Table VII, the signifying distributions of the response patterns are shaded to highlight their correspondence to the different developmental levels. For example, the 1331 pattern appears to signify an important developmental step spanning across the wide-ranging intermediate stages. Additionally, these response patterns and the corresponding students’ explanations from interviews also help to identify several key reasoning pathways that are representative of students’ reasoning at different developmental levels, which are summarized in Table VIII.

TABLE VIII. Students’ reasoning pathways at different developmental levels. Factors of reasoning marked with a single asterisk (\*) are used in determining causality. Factors marked with double asterisks (\*\*) represent mixed usage meaning that they are sometimes used in determining causality. Unmarked factors are not used in determining causality. The determination of use is based on if the related factors are explicitly stated in the “because” part of an explanation.

Reasoning levels	Causality	Magnitude	Scenario timeline	Force simultaneity	Action-reaction
Novice	Causal	Asymmetry*	Asymmetry*	Unaware	Use*
Lower intermediate	Causal	Equal	Mixed*	Mixed	Use*
Upper intermediate	Mixed	Equal	Distinguished**	Identified**	Mixed**
Expert	Noncausal	Equal	Distinguished	Identified*	Distinguished

Among the reasoning pathways, recognizing equal magnitude of interaction forces appears to be a rudimentary step in the beginning of the transition from the novice level to the intermediate levels. Then, being able to distinguish the asymmetry in scenario timeline from the simultaneity of interaction forces seems to be another important step separating lower intermediate and upper intermediate students. Some of the lower intermediate students are starting to distinguish the asymmetry in scenario timeline but still use it as the reason for a causal relation. Meanwhile, most of the upper intermediate level students can distinguish the asymmetry in scenario timeline from the simultaneity of interaction forces, and many of them are able to use the simultaneity to determine a noncausal relation, although some may still use the asymmetry in scenario timeline to determine a causal relation, indicating a mixed state of understanding. Additionally, understanding the action-reaction language as a description of interaction rather than implication of causality is another important step that only upper intermediate level students are able to demonstrate. The manifestations of different coexisting thinking pathways indicate significant mixing in reasoning among the students at the intermediate levels. Moving forward, the expert level students are able to demonstrate consistent expertlike thinking on all the related reasoning attributes, indicating the achievement of a well-integrated knowledge structure and a deep understanding.

To sum up, the results of pattern analysis and interviews show strong connections between students' understanding of causality and asymmetries in magnitude, scenario timeline, and the action-reaction language. These connections reveal unique thinking pathways, which evolve as students' reasoning develops, indicating a progression of developmental levels in causal reasoning. The analysis and implications provide quantitative and qualitative evidence to answer the second research question.

#### IV. SUMMARY AND DISCUSSIONS

In this study, assessments of high school and college students suggest that after instruction on Newton's third law many students still hold a belief in causal relation between a pair of N3L interaction forces. This understanding in causality is influenced by scenario contexts. It also shows a possible effect of forgetting and reveals no gender differences among high school students. Further analysis on students' response patterns and interviews suggests that students' belief in causality is strongly associated with asymmetries in the scenario contexts including the asymmetries due to a dominant feature, scenario timeline, and the action-reaction language. In addition, results from interviews and response patterns also demonstrate a progression of developmental levels of causal reasoning signified by a set of unique thinking pathways that use the related asymmetries in reasoning to determine causality. Typically, novice thinkers often use

any type of asymmetry to support the claim of a causal relation. Meanwhile, lower and upper intermediate thinkers demonstrate a wide-ranging mixed thinking that relies more on the timeline asymmetry and the action-reaction language in their reasoning for causality. Upper intermediate thinkers are able to recognize the simultaneity of interaction forces and distinguish it from timeline asymmetry; however, only some are able to apply the simultaneity to determine a noncausal relation. It is not until students achieve expert level thinking that they are finally able to determine the noncausal relation with all the related reasoning attributes consistently integrated within a coherent thinking pathway.

From this study, it is evident that students' reasoning about causal relations is an important element for achieving deep understandings of basic physics concepts. In the example of N3L, most students are able to answer the traditional magnitude questions, but many of them demonstrate incorrect reasoning on the related causal relation, indicating a fragmented knowledge structure and incomplete understanding. In the long term, the lack of correct reasoning may interfere with students' learned knowledge and cause students to revert to their original novice thinking, which can be speculated as a contributing factor leading to the weaker performance of the college students as observed in this study. Therefore, making emphasis on teaching not only content but also reasoning in physics needs much attention in current STEM education research and practices [6,9], especially when thinking and reasoning is becoming a primary goal of learning outcomes in the current education initiatives [7].

More specifically in this study, it is found that asymmetry is a key attribute in causal reasoning, and students can be cued by a range of asymmetries into determining a causal relation. A difficulty here is the disconnect between time and causality, with which students would recognize the simultaneity of interaction forces but still determine a causal relation. From the surface level, it may imply that students do not rely on asymmetry to determine causality. However, through more in-depth analysis, the nature of the disconnect between time and causality is such that even if students can recognize the simultaneity of interaction forces, they use the asymmetry in scenario timeline to claim a causal relation without seeing the contradiction. Their causal reasoning still relies on asymmetry, but some students would use any asymmetry in replacement of the critical one (i.e., the asymmetry or simultaneity in time order of forces) and be satisfied to determine a causal relation. These students are not developed at the level to distinguish the difference between the timeline asymmetry and the simultaneity of the actual interaction forces and know which one is the foundation for determining causality.

The design of the assessment questions, which helps to form simple but meaningful response patterns, appears to work well in measuring students thinking pathways. When combined with interviews, analysis of response patterns

provides effective utility to identify subtle features of student reasoning and establish progression levels of reasoning development. The identification of the different pathways and their states in different developmental levels can provide useful information for future instruction and assessment.

In this study, although a number of interesting outcomes are identified, there are a few limitations worth noting when interpreting the results. First, the result on gender difference is limited to the high school population studied. Because of the predominant male concentration in college students, a reasonable analysis cannot be performed with the college population. In future research, more diverse populations should be used to explore if gender difference may exist in different education settings and population backgrounds.

Second, although the high school and college students are from the same region, they still represent two different populations. According to the background information, the high school students represent a population with weaker academic performance compared to the college population. On this ground, the results showing college students performing weaker than high school students on the N3L assessment are adequate to suggest that the weaker performance of college students may be due to forgetting. However, alternative explanations are possible and need to be explored through an actual longitudinal design using a single population in future research.

Additionally, based on the results shown in Table VIII, the assessment questions can be further improved to explicitly address the difference between timeline asymmetry and force simultaneity, which can provide direct quantitative measurements of students' understandings of these two attributes. Additional research is needed to extend the current study with a more balanced body of subjects and to address the weaknesses. Future research on instructional interventions is also warranted, which is lacking in current curriculum and needed to improve this important fundamental reasoning ability.

### ACKNOWLEDGMENTS

The research is supported in part by the National Science Foundation Grant No. DUE-1712238, and by the Peak Discipline Construction Project of Education at East China Normal University and The National Social Science Fund of China (Grant No. 17BSH003). Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agencies.

### APPENDIX: ASSESSMENT OF CAUSAL REASONING IN NEWTON'S THIRD LAW

Included below is a multiple-choice test on causal reasoning in the context of Newton's third law. The test

is designed with four context scenarios that each contains four questions on attributes of causal relations.

1. In a soccer game, two players, A and B, are running to chase a ball. Player A has twice the mass as Player B. They are both running at the same speed towards each other. Unfortunately, neither player notices the other and they collide into one another. During the collision, each player exerts a force on the other player. The force exerted by player A on player B is  $F_{AB}$ , while the force exerted by player B on player A is  $F_{BA}$ . Consider these two forces between the two players and answer the following questions.
  - 1.1 Is there a causal relation between the two forces?
    - 1) Yes,  $F_{AB}$  causes  $F_{BA}$ .
    - 2) Yes,  $F_{BA}$  causes  $F_{AB}$ .
    - 3) No, there is no causal relation between the two forces.
  - 1.2 Compare the magnitudes of the two forces.
    - 1)  $F_{AB}$  is larger than  $F_{BA}$ .
    - 2)  $F_{BA}$  is larger than  $F_{AB}$ .
    - 3) The magnitudes of the two forces are equal.
  - 1.3 Does one of the forces occur before the other force?
    - 1) Yes,  $F_{AB}$  occurs before  $F_{BA}$ .
    - 2) Yes,  $F_{BA}$  occurs before  $F_{AB}$ .
    - 3) No, both forces occurred simultaneously.
  - 1.4 Which force is the action force, and which one is the reaction force?
    - 1)  $F_{AB}$  is the action force and  $F_{BA}$  is the reaction force.
    - 2)  $F_{BA}$  is the action force and  $F_{AB}$  is the reaction force.
    - 3) It cannot be determined which force is the action or reaction force.
2. A textbook is at rest on top of a horizontal table. The book exerts a downward force  $F_{BT}$  on the table, while the table exerts a supporting force  $F_{TB}$  on the book. Consider these two forces between the book and the table and answer the following questions.
  - 2.1 Is there a causal relation between the two forces?
    - 1) Yes,  $F_{BT}$  causes  $F_{TB}$ .
    - 2) Yes,  $F_{TB}$  causes  $F_{BT}$ .
    - 3) No, there is no causal relation between the two forces.
  - 2.2 Compare the magnitudes of the two forces.
    - 1)  $F_{AB}$  is larger than  $F_{BA}$ .
    - 2)  $F_{BA}$  is larger than  $F_{AB}$ .
    - 3) The magnitudes of the two forces are equal.
  - 2.3 Does one of the forces occur before the other force?
    - 1) Yes,  $F_{BT}$  occurs before  $F_{TB}$ .
    - 2) Yes,  $F_{TB}$  occurs before  $F_{BT}$ .
    - 3) No, both forces occurred simultaneously.
  - 2.4 Which force is the action force, and which one is the reaction force?
    - 1)  $F_{BT}$  is the action force and  $F_{TB}$  is the reaction force.



- 2)  $F_{TB}$  is the action force and  $F_{BT}$  is the reaction force.
  - 3) It cannot be determined which force is the action or reaction force.
3. Two students, A and B, are on identical low friction roller skates facing each other. They both have the same mass of 50 kg. A puts her hand on B and pushes outward causing both to move. While A is pushing B with a force  $F_{AB}$ , A also feels a force  $F_{BA}$  pushing back on her by B. Consider these two forces between A and B and answer the following questions.
- 3.1 Is there a causal relation between the two forces?
- 1) Yes,  $F_{AB}$  causes  $F_{BA}$ .
  - 2) Yes,  $F_{BA}$  causes  $F_{AB}$ .
  - 3) No, there is no causal relation between the two forces.
- 3.2 Compare the magnitudes of the two forces.
- 1)  $F_{AB}$  is larger than  $F_{BA}$ .
  - 2)  $F_{BA}$  is larger than  $F_{AB}$ .
  - 3) The magnitudes of the two forces are equal.
- 3.3 Does one of the forces occur before the other force?
- 1) Yes,  $F_{AB}$  occurs before  $F_{BA}$ .
  - 2) Yes,  $F_{BA}$  occurs before  $F_{AB}$ .
  - 3) No, both forces occurred simultaneously.
- 3.4 Which force is the action force, and which one is the reaction force?
- 1)  $F_{AB}$  is the action force and  $F_{BA}$  is the reaction force.
  - 2)  $F_{BA}$  is the action force and  $F_{AB}$  is the reaction force.
  - 3) It cannot be determined which force is the action or reaction force.

- 
- [1] M. Alonso, Problem solving vs. conceptual understanding, *Am. J. Phys.* **60**, 777 (1992).
- [2] W. J. McKeachie, *Teaching and Learning in the College Classroom: A Review* (National Center for Research to Improve, Ann Arbor, 1986).
- [3] R. R. Hake, Interactive-engagement versus traditional methods: A six thousand-thousand-student survey of mechanics test data for introductory courses, *Am. J. Phys.* **66**, 64 (1998).
- [4] National Research Council, *Assessing 21st Century Skills: Summary of a Workshop* (National Academies Press, Washington, DC, 2011).
- [5] National Research Council, *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century* (The National Academies Press, Washington, DC, 2012).
- [6] L. Bao and K. Koenig, Physics education research for 21st century learning, *Discip. Interdiscip. Sci. Educ. Res.* **1**, 2 (2019).
- [7] National Research Council, *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (National Academies Press, Washington, DC, 2012).
- [8] United States Chamber of Commerce, Bridging the soft skills gap: How the business and education sectors are partnering to prepare students for the 21st century workforce, in *Center for Education and Workforce* (U.S. Chamber of Commerce Foundation, Washington, DC, 2017).
- [9] L. Bao, T. Cai, K. Koenig, K. Fang, J. Han, J. Wang, Q. Liu, L. Ding, L. Cui, Y. Luo, Y. Wang, L. Li, and N. Wu, Learning and scientific reasoning, *Science* **323**, 586 (2009).
- [10] L. Bao and J. C. Fritchman, Assessment of knowledge integration in student learning of Newton's third law, [arXiv:2104.05944](https://arxiv.org/abs/2104.05944).
- [11] D. E. Brown, Students' concept of force: the importance of understanding Newton's third law, *Phys. Educ.* **24**, 353 (1989).
- [12] J. Clement, Students' preconceptions in introductory mechanics, *Am. J. Phys.* **50**, 66 (1982).
- [13] I. A. Halloun and D. Hestenes, The initial knowledge state of college physics students, *Am. J. Phys.* **53**, 1043 (1985).
- [14] L. Bao, K. Hogg, and D. Zollman, Model analysis of fine structures of student models: An example with Newton's third law, *Am. J. Phys.* **70**, 766 (2002).
- [15] U. Besson, Some features of causal reasoning: Common sense and physics teaching, *Res. Sci. Technol. Educ.* **22**, 113 (2004).
- [16] M. Bunge, *Causality. The Place of the Causal Principle in Modern Science* (Harvard University Press, Cambridge, MA, 1959).
- [17] F. Halbwachs, Réflexions sur la causalité physique. Causalité linéaire et causalité circulaire, in *Les Théories de la Causalité*, edited by M. Bunge, F. Halbwachs, T. S. Kuhn, and J. Piaget (PUF, Paris, 1971).
- [18] J. Piaget, Causalité et opérations, in *Les Explications Causales*, edited by J. Piaget and R. Garcia (PUF, Paris, 1971).
- [19] R. Harré, *The Philosophies of Science* (Oxford University Press, Oxford, 1972).
- [20] J. Ogborn, Approche théorique et empirique de la causalité, *Didaskalia* **1**, 29 (1993).
- [21] M. Planck, *Vorträge und Erinnerungen* (S. Hirzel, Stuttgart, 1949).
- [22] F. C. Keil, *Concepts, Kinds, and Cognitive Development* (MIT Press, Cambridge, MA, 1989).
- [23] S. Carey, The origin of concepts: Continuing the conversation, in *Representation, Memory, and Development: Essays in Honor of Jean Mandler*, edited by N. L. Stein, P. J. Bauer, and M. Rabinowitz (Erlbaum, Mahwah, NJ, 2002), pp. 43–52.
- [24] R. K. Guenther, *Human Cognition* (Prentice Hall, Upper Saddle River, NJ, 1998).

- [25] W. F. Brewer, C. A. Chinn, and A. Samarapungavan, Explanation in scientists and children, in *Explanation and Cognition*, edited by F. C. Keil and R. A. Wilson (MIT Press, Cambridge, MA, 2000), pp. 279–298.
- [26] S. Carey, On the origin of causal understanding, in *Causal Cognition: A Multidisciplinary Debate*, edited by D. Sperber, D. Premack, and A. J. Premack (Clarendon Press, Oxford, UK, 1995), pp. 268–302.
- [27] R. Corrigan and P. Denton, Causal understanding as a developmental primitive, *Dev. Rev.* **16**, 162 (1996).
- [28] P. Thagard, Probabilistic networks and explanatory coherence, *Cogn. Sci. Quarterly.* **1**, 91 (2000).
- [29] A. Schlottmann, Perception versus knowledge of cause and effect in children: When seeing, *Curr. Dir. Psychol. Sci.* **10**, 111 (2001).
- [30] H. M. Wellman and S. A. Gelman, Knowledge acquisition in foundational domains, in *Handbook of Child Psychology: Cognition, Perception and Language* (Wiley, New York, 1998), pp. 523–573.
- [31] B. Anderson, The experiential Gestalt of causation: A common core to pupils' preconceptions, *Eur. J. Sci. Educ.* **8**, 155 (1986).
- [32] G. Lakoff and M. Johnson, *Metaphors We Live By* (University of Chicago Press, Chicago, 1980).
- [33] A. Michotte, *La perception de la causalité* (Vrin, Paris, 1946).
- [34] W. Hung and D. H. Jonassen, Conceptual understanding of causal reasoning in physics, *Int. J. Sci. Educ.* **28**, 1601 (2006).
- [35] M. Bullock, R. Gelman, and R. Baillargeon, The development of causal reasoning, in *The Developmental Psychology of Time*, edited by W. Friedman (Academic Press, New York, 1982), pp. 209–254.
- [36] H. H. Kelley, The process of causal attribution, *Am. Psychol.* **28**, 107 (1973).
- [37] P. W. Cheng, From covariation to causation, *Psychol. Rev.* **104**, 367 (1997).
- [38] P. Thagard, Explaining disease: Correlations, causes, and mechanisms, in *Explanation and Cognition*, edited by F. C. Keil and R. A. Wilson (MIT Press, Cambridge, MA, 2000), pp. 254–276.
- [39] D. Psillos and P. Koumaras, Multiple causal modelling of electrical circuits for enhancing, in *Learning Electricity and Electronics with Advanced*, NATO ASI series F, Vol. 115, edited by M. Caillot (Springer, Berlin, 1993), pp. 57–75.
- [40] B. A. Sherwood and R. W. Chabay, Conceptual model for understanding the behaviour of electrical circuits, in *Learning Electricity and Electronics with Advanced Educational Technology*, NATO ASI Series F, Vol. 115, edited by M. Caillot (Springer, Berlin, 1993), pp. 23–35.
- [41] B. A. Sherwood and R. W. Chabay, *Matter and Interactions* (John Wiley & Sons, New York, 2002).
- [42] J. Gutwill, J. Frederiksen, and M. Ranney, Seeking the causal connection in electricity: shifting among mechanistic perspectives, *Int. J. Sci. Educ.* **18**, 143 (1996).
- [43] D. E. Brown, Using examples and analogies to remediate misconceptions in physics: factors influencing conceptual change, *J. Res. Sci. Teach.* **29**, 17 (1992).
- [44] C. Chinn and B. Malhotra, Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks, *Sci. Educ.* **86**, 175 (2002).
- [45] M. Mustakim, J. Mansyur, M. Rizal, and S. N. Kaharu, Analysis of students' causal reasoning in physics problem solving, *J. Phys. Conf. Ser.* **1521**, 022058 (2020).
- [46] L. Viennot, Fundamental patterns in common reasoning: Examples in physics, *Tijdschrift voor Didactiek der ft-wetenschappen* **12**, 89 (1994), <http://www.fi.uu.nl/publicaties/literatuur/199402-viennot.pdf>.
- [47] R. Dai, J. C. Fritchman, Q. Liu, Y. Xiao, and L. Bao, Assessment of student understanding on light interference, *Phys. Rev. Phys. Educ. Res.* **15**, 020134 (2019).
- [48] W. Xu, Q. Liu, K. Koenig, J. Fritchman, J. Han, S. Pan, and L. Bao, Assessment of knowledge integration in student learning of momentum, *Phys. Rev. Phys. Educ. Res.* **16**, 010130 (2020).
- [49] L. Bao and E. F. Redish, Model analysis: Assessing the dynamics of student learning, *Physical Phys. Rev. ST Phys. Educ. Res.* **2**, 010103 (2006).
- [50] L. M. Collins and S. T. Lanza, *Latent Class and Latent Transition Analysis* (John Wiley & Sons, Hoboken, NJ, 2010).
- [51] Y. Nie, Y. Xiao, J. C. Fritchman, Q. Liu, J. Han, J. Xiong, and L. Bao, Teaching towards knowledge integration in learning force and motion, *Int. J. Sci. Educ.* **41**, 2271 (2019).
- [52] M. Schneider and I. Hardy, Profiles of inconsistent knowledge in children's pathways of conceptual change, *Dev. Psychol.* **49**, 1639 (2013).
- [53] M. Straatemeier, H. L. J. van der Maas, and B. R. J. Jansen, Children's knowledge of earth: A new methodological and statistical approach, *J. Exp. Child Psychol.* **100**, 276 (2008).
- [54] T. J. P. van Schijndel, S. E. van Es, R. K. Franse, B. M. C. W. van Bers, and M. E. J. Raijmakers, Children's mental models of prenatal development, *Frontiers Psychol.* **9**, 1835 (2018).
- [55] G. Schwarz, Estimating the dimension of a model, *Ann. Stat.* **6**, 461 (1978).
- [56] J. K. Vermunt and J. Magidson, Latent class analysis, in *The Sage Encyclopedia of Social Sciences Research Methods*, edited by M. Lewis-Beck, A. Bryman, and T. F. Liao (Sage, Thousand Oakes, 2004), pp. 549–553.
- [57] D. A. Linzer and J. Lewis, polCA: An R package for polytomous variable latent class analysis, *J. Stat. Softw.* **42**, 1 (2011).