

Reasoning with alternative explanations in physics: The cognitive accessibility ruleAndrew F. Heckler¹ and Abigail M. Bogdan²¹*Department of Physics, Ohio State University, Columbus, Ohio 43210, USA*²*Department of Physics, Seton Hill University, Greensburg, Pennsylvania 15601, USA*

(Received 18 August 2017; published 9 April 2018)

A critical component of scientific reasoning is the consideration of alternative explanations. Recognizing that decades of cognitive psychology research have demonstrated that relative cognitive accessibility, or “what comes to mind,” strongly affects how people reason in a given context, we articulate a simple “cognitive accessibility rule”, namely that alternative explanations are considered less frequently when an explanation with relatively high accessibility is offered first. In a series of four experiments, we test the cognitive accessibility rule in the context of consideration of alternative explanations for six physical scenarios commonly found in introductory physics curricula. First, we administer free recall and recognition tasks to operationally establish and distinguish between the relative accessibility and availability of common explanations for the physical scenarios. Then, we offer either high or low accessibility explanations for the physical scenarios and determine the extent to which students consider alternatives to the given explanations. We find two main results consistent across algebra- and calculus-based university level introductory physics students for multiple answer formats. First, we find evidence that, at least for some contexts, most explanatory factors are cognitively available to students but not cognitively accessible. Second, we empirically verify the cognitive accessibility rule and demonstrate that the rule is strongly predictive, accounting for up to 70% of the variance of the average student consideration of alternative explanations across scenarios. Overall, we find that cognitive accessibility can help to explain biases in the consideration of alternatives in reasoning about simple physical scenarios, and these findings lend support to the growing number of science education studies demonstrating that tasks relevant to science education curricula often involve rapid, automatic, and potentially predictable processes and outcomes.

DOI: [10.1103/PhysRevPhysEducRes.14.010120](https://doi.org/10.1103/PhysRevPhysEducRes.14.010120)**I. INTRODUCTION**

One hallmark of successful reasoning, especially in science, is the systematic consideration of multiple explanations. Besides its important contribution to valid scientific reasoning, this skill also plays a critical role in “hypothesis generation”, which is often used as part of the scientific process of discovery [1,2]. Yet, even as Francis Bacon had observed, when the mind “is left to itself,” such disciplined considerations often fail to occur [3]. Numerous studies have documented the tendencies of people to search for evidence that confirms their prior beliefs at the expense of considering other possibilities (confirmation bias) [4,5], and, in the context of deductive reasoning tasks, people often fail to consider alternative explanations even when such explanations are reasonably warranted by normative logic [6–12]. In the domain of science education, it was recognized relatively early that the frequency of considering alternatives depended

on context [10,13]. Further, the failure to consider alternative explanations (e.g., the variation of more than one cause) was found to be a contributing factor in difficulties with reasoning about phenomena with multiple causes [10,14–18].

But why do people commonly fail to consider an alternative explanation? Is there a general cognitive mechanism that can explain such failures for scientific reasoning? In this study, we formally propose and empirically test the domain general mechanism of cognitive “accessibility.” That is, we investigate the influence of cognitive accessibility on the extent to which university level students ignore alternative explanations for several simple physical concepts and phenomena that are commonly found in physical science curricula. It is important to note that while what we propose may appear intuitively obvious and may be aligned with common and implicit assumptions about student thinking, this study is useful for two general reasons. First, it refines and formalizes such thinking by and explicitly stating a simple and testable mechanism in the form of a rule (stated later in this section) that can give significant insight into student reasoning. Second, it empirically tests and verifies the proposed rule. The formal empirical verification of accessibility as a mechanism for modulating the consideration of alternative explanations in science reasoning is, as far as we can discern, also novel. As such, this study could be

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

seen as the refinement and application of a number of experimental methods, results, and theoretical mechanisms on conditional reasoning from cognitive psychology to physical concepts relevant to science education.

Specifically, in the context of conditional reasoning tasks, a number of studies have documented that the number and relative strength (accessibility) of explanations plays a significant role in modulating the extent to which reasoners consider alternatives [6,9,11,19]. For example, Quinn and Markovits [6] presented some participants with everyday context scenarios such as “A dog scratches constantly” and asked those participants to list every possible cause. They found that “fleas” was listed 100% of the time while “skin disease” was listed only 32% of the time. They used these data for an operational definition of strength of accessibility and claimed that, for this scenario and among participants in this population, the cause fleas was, on average, a more strongly accessible cause than skin disease. In the second phase of the study, one group of participants in the *weakly* associated condition (i.e., low accessibility) were asked to assume that the following initial premise is true: “If a dog has a skin condition, then it will scratch constantly.” Then the participants were shown the statement “If a dog scratches constantly, then the dog has a skin condition” and were asked to determine if this statement (which is normatively logically incorrect, formally known as “affirming the consequent”) follows from the accepted initial premise. Specifically, they were asked if they were certain that the statement was true, certain it was not true, or uncertain whether it was true or not true, given the premise. This was compared to the responses of the *strongly* associated condition, where participants were asked to assume that the following premise is true: “If a dog has fleas, then it will scratch constantly,” and then asked to determine if the (similarly logically incorrect) statement follows from the premise “If a dog scratches constantly, then the dog has fleas.” The results showed that participants in the strongly associated condition made more logically incorrect conclusions (responding that the statement follows from the premise) compared to the weakly associated condition. Similar to the previous experiment, the interpretation is that when a relatively highly accessible cause is presented as an explanation for an effect, reasoners tend to ignore possible alternatives, but when a relatively weakly accessible cause is presented, reasoners tend to offer alternative explanations, namely, ones that are more highly accessible. In our study, our experimental design is reminiscent of the experiment of Quinn and Markovits, but with some modifications adapted for specific simple physics concepts and phenomena.

Overall, there are two goals for this study. The first is to empirically demonstrate that the relative *accessibility* of explanatory factors, which, similar to the study discussed above, is operationally definable and measurable, plays a significant role in modulating reasoning bias for educationally important physical science concepts and phenomena. As such we are applying the general cognitive psychological phenomenon of the influence of accessibility on reasoning to

educationally relevant physics content. To more explicitly summarize and operationalize existing general cognitive psychological findings and to provide a standing hypothesis for this study, we articulate and propose the following “accessibility rule”:

The relative cognitive accessibility of explanatory factors of a given physical phenomenon affects the likelihood that alternative explanations will be considered for that phenomenon. More specifically, alternative explanations are considered less frequently when an explanation with relatively high accessibility is offered first.

We will show evidence of this effect and further propose two corollaries. For a given outcome,

Corollary 1: If explanation A is offered as an explanation for a physical scenario, the likelihood that an alternative explanation will be considered increases with increasing accessibility of another explanation B.

Corollary 2: The more likely that only a single explanatory factor A is accessible, the less likely that an alternative explanation will be considered when the factor A is offered as the explanation.

Note that for simplicity in this paper, while participants may offer a number of explanations for a given scenario, we are only considering the two most considered explanations. The top two explanations typically account for 80%–90% of the explanations given. Corollary 1 could be modified to include additional possible explanations.

The second goal for this study is to explicitly demonstrate that while certain explanatory factors may not be accessible in some contexts, they may be accessible in others. That is, accessibility is not a “hard limit” on reasoning; rather we must keep in mind the possibility that accessibility, thus student responses, depends on context. This is not only a common observation in everyday conversations (“Ah yes, I knew that, it just never came to mind”), but also in education and psychology research, as we will discuss in the next section. Yet, this observation plays an important role for interpreting the meaning and possible mechanisms for the effect of accessibility on reasoning, and this in turn may have implications for instructional strategies.

A. Theoretical background and context

There are several points to discuss in order to set the context for this study. The first is what is meant by cognitive accessibility. The notion of accessibility, or what comes to mind, has played a major role in psychological research over the past 50 years [20,21]. The model consists of the idea that knowledge which is “available” to a person (e.g., stored in memory) is “accessible” if it is somehow activated by context, and is consequently used to make decisions (see also Ref. [22]). Therefore, which knowledge is accessed can have an effect on the decision made. For example, the study

of Quinn and Markovitz discussed in Sec. I can be set within this general idea of accessibility: explanatory factors that are strongly associated with the reasoning context are considered highly accessible. Similar to other studies, in this study we will operationally define relative accessibility as the relative number of times that an explanation is listed in a free-recall task describing a physical scenario with some outcome (such as one pendulum having a longer period than another).

The second point to consider is that the effect of accessibility on reasoning could be modeled in terms of the influence of unconscious, automatic processes rather than of deliberate thinking or of high level mental structures [22]. This argument is part of the general model of dual cognitive systems. The dual systems approach models human cognition in terms of system 1 (fast, automatic, heuristic processes) and system 2 (slow, deliberate, analytic processes) [23]. In this model, both the mechanisms of accessibility and the tendency to make decisions based on accessibility are due to system 1 processes. Theories involving these processes include such ideas as a “singularity principle,” namely, that people tend to think of only one explanation at a time, (perhaps due to system 1 processes), and this can lead to ignoring alternative explanations [24]. In the last decade, science education researchers have also begun to explicitly note the importance of dual process theories when explaining student responses to tasks in science education settings [17,25–28].

The final point is that there is evidence that the effect of relative accessibility is modulated by the direction of the reasoning task, and this is an effect that we tested in our study (experiment 3). Specifically, some studies have examined both predictive reasoning tasks (provided a cause, predict the effect) and diagnostic reasoning tasks (provided an effect, diagnose the cause). Fernbach, Darlow, and Sloman [11] found that in the case where there are multiple possible single causes, diagnostic reasoning was affected by the relative strength (accessibility) of alternatives, but predictive reasoning was not. We tested their finding for instructionally pragmatic reasons since, when providing practice examples for students to reason with, it is important to understand the relative benefits and challenges of forward or backward reasoning.

II. EXPERIMENT 1: BASELINE ACCESSIBILITY AND AVAILABILITY OF FACTORS

The goal of experiment 1 is to operationally determine and compare measures of the baseline accessibility and availability of explanatory factors for six educationally relevant physical scenarios. This was done using both recall and recognition tasks. As an example of a physical scenario, consider determining the period of a simple pendulum. What explanatory factors are accessible to participants? We found that mass and length are the factors most frequently considered. What is the relative accessibility of these factors? On average, is length recalled more frequently (more accessible)

than mass? Note that we considered factors that participants listed regardless of whether the factors are physically valid. For example, mass does not affect the period of a pendulum, yet it is reasonable to expect that the relative accessibility of mass as an explanatory factor can play a role in participants’ ability to reason with multiple factors.

Comparing the difference between the availability and accessibility of explanatory factors for a given scenario is also important. For example, participants may not recall “lever arm length” as an explanatory factor for a tipping balance beam, but if shown this factor, they may readily recognize its relevance. Thus, in this case lever arm has relatively low accessibility but it still can be available (recognized) as a relevant factor. Again, keep in mind that physically incorrect factors are considered since they may be available and accessible to some participants.

A. Methods and materials

The participants in experiments 1 and 2 were drawn (over a period of three semesters) from a pool of over 2000 students enrolled in the first semester calculus-based physics course at Ohio State University, a large public research university. The students in this course were assigned a “flexible homework” assignment in which they completed various physics education research tasks, some of which were from this study. Over 90% of all students enrolled in the course completed the flexible homework assignment individually in a quiet room and were given full credit for participation. From this pool, and only a small portion were randomly assigned into this study. In total, 136 participants were randomly assigned to one of two conditions: either the “recall” or “recognize” conditions. Because of scheduling constraints and some design errors in early versions of the tasks, the numbers of participants are not uniform across conditions or within conditions (specific numbers are in the results section). However, since the differences in numbers were not due to systematic selection effects, both conditions are expected to have similar participants.

Six physical scenarios were used in this study (see Table I). These scenarios are commonly found in introductory physics at the university level, and many of them are also common in high school physical science curriculum. We chose simple physical scenarios that typically had two factors that the students believed (correctly or incorrectly) influenced a simple outcome, and that elicited relatively straightforward and easy to interpret explanations. Prior to the experiments reported here, pilot studies including brief interviews and trial versions of all items used in all four experiments in this study were conducted in order to ensure that participants were interpreting the questions as intended. Items were adjusted as appropriate.

Our recall condition was designed to test the *accessibility* of factors in each of the six physical scenarios. Participants in the recall condition were presented with six physical scenarios and asked to list all the possible reasons which

TABLE I. Experiment 1 recall condition questions.

Physical scenario	Question
Mass density	Block <i>X</i> and block <i>Y</i> are each made of a different material. The material of block <i>X</i> has higher mass density (kg/m^3) than the material of block <i>Y</i> . <i>On an atomic level</i> , why would material <i>X</i> have a higher mass density than material <i>Y</i> ? List all the possible reasons.
Balance scale	A rod is balanced on a pivot. A student hangs object <i>A</i> somewhere on the left side of the rod and hangs object <i>B</i> somewhere on the right side of the rod. The right-hand side immediately starts to tilt down. List the possible reason(s) why the right-hand side of the rod tilted down.
Pendulum	Pendulum <i>A</i> swings with a longer period (time) than pendulum <i>B</i> . Both are simple pendulums. List the possible reason(s) why pendulum <i>A</i> has a longer period.
Projectile	Projectile <i>A</i> and projectile <i>B</i> start at ground level and are thrown with the same speed, but <i>A</i> is in the air for a longer time than <i>B</i> . <i>Ignore any effects of air resistance or drag</i> . List the possible reason(s) why <i>A</i> has a longer flight time than <i>B</i> .
Sliding	Object <i>P</i> and object <i>Q</i> are both given a quick push and slide on a wooden floor with the same initial velocity. Object <i>P</i> comes to rest before object <i>Q</i> . List the possible reason(s) why <i>P</i> comes to rest before <i>Q</i> .
Gravitational potential energy	Object 1 has larger gravitational potential energy than object 2. List the possible reason(s) why object 1 has a larger gravitational potential energy.

might explain the scenario. For example, in the pendulum scenario, participants were asked to list all the reasons why one pendulum would have a longer period than another. The full set of questions given to students in the recall condition is shown in Table I.

Our recognize condition was designed to test the *availability* of factors in each of the six physical scenarios. Participants in the recognize condition were asked to complete the items presented in Table II. Note that for this condition major factors were provided to them (instead of asking participant to recall possible factors). The factor was marked as available if the student indicated any dependence (as opposed to “doesn’t matter”), regardless of correctness. These factors were drawn from pilot studies with the six physical scenarios.

B. Results

For the recall condition Table III presents statistics on the two most commonly listed factors for each scenario. We provided the percentage of participants who listed each factor, who listed only a given factor, who listed the factor first, and who listed both of the two most common factors. For example, for the mass density scenario, the two most commonly listed factors were atomic spacing, which was listed 70% of the time and atomic mass, which was listed 45% of the time. Further, 37% of the participants listed spacing only, while 12% of participants listed atomic mass only. Perhaps as to be expected, spacing was listed first more often than atomic mass 55% vs 27% of the time. Finally, 25% listed both factors: Note that summing up participants who only listed one factor and those listing both factors does not sum to 100% because other factors besides the top two were

sometimes also listed. In all scenarios, the other factors listed had low counts, less than 5% of responses each.

For each physical scenario, we determined whether the relative accessibility (or strength) of the two factors was significantly different. While one might operationally define relative accessibility in a number of ways (frequency of listing, listing only, listing first, etc.), we compared the relative number of times each factor was listed. Essentially, the statistical comparison is a McNemar’s test, and the relevant information is the number of participants who listed factor *A* and not factor *B* compared to those who listed factor *B* and not factor *A*. We found that there was a significant difference at the 95% confidence level between the two factors for all of the scenarios except for gravitational potential energy, where almost all (90%) of the participants list both factors, and for sliding, where it is clear upon inspection that neither of the factors is preferred over the other (despite one of them being physically incorrect). In Table III we denoted in bold face the more highly accessible (more strongly associated) factor. Note also that for each scenario the results for the relative accessibility of the factors, determined by comparing the frequency with which each factor was listed, is consistent with the frequency that the factors were only listed or were listed first. Thus, we can be confident that the indicated factors have higher accessibility in several ways.

Results for the recognize condition are present in Table IV. There are two main points to take away from the recognize condition results. The first and most important (though not surprising) point is that participants often recognize a factor as explanatory when *given* it much more frequently than they spontaneously *recall* that factor. For

TABLE II. Experiment 1 recognize condition questions. Comparison such as “larger” means larger with respect to the other object in the scenario.

Physical scenario	Question
Mass density	Block <i>X</i> and block <i>Y</i> are each made of a different material. The material of block <i>X</i> has higher mass density (kg/m^3) than the material of block <i>Y</i> . Which of the following could be the reason that material <i>X</i> has a higher mass density than material <i>Y</i> ? Circle your answer. The average atomic separation in block <i>X</i> is: larger smaller doesn't matter The average atomic mass in block <i>X</i> is: larger smaller doesn't matter
Balance scale	A rod is balanced on a pivot. A student hangs object <i>A</i> somewhere on the left side of the rod and hangs object <i>B</i> somewhere on the right side of the rod. The right-hand side immediately starts to tilt down. Which of the following could be the reason that the right-hand side tilts down? Circle your answer. The distance from object <i>A</i> to the center pivot is: longer shorter doesn't matter The mass of object <i>A</i> is: larger smaller doesn't matter
Pendulum	Pendulum <i>A</i> swings with a longer period (time) than pendulum <i>B</i> . Both are simple pendulums. Which of the following could be the reason for this? Circle your answer. Pendulum <i>A</i> is: longer shorter doesn't matter The mass of pendulum <i>A</i> is: larger smaller doesn't matter
Projectile	Projectile <i>A</i> and projectile <i>B</i> start at ground level and are thrown with the same speed, but <i>A</i> is in the air for a longer time than <i>B</i> . <i>Ignore any effects of air resistance or drag.</i> Which of the following could be the reason for this? Circle your answer. The mass of projectile <i>A</i> is: larger smaller doesn't matter The launch angle of projectile <i>A</i> is: larger smaller doesn't matter
Sliding	Object <i>P</i> and object <i>Q</i> are both given a quick push and slide on a wooden floor with the same initial velocity. Object <i>P</i> comes to rest before object <i>Q</i> . Which of the following could be the reason that object <i>P</i> comes to a rest before object <i>Q</i> ? Circle your answer. The coefficient of friction for object <i>P</i> is: larger smaller doesn't matter The mass of object <i>P</i> is: larger smaller doesn't matter
Gravitational potential energy	Object 1 has larger gravitational potential energy than object 2. Select all that apply. Which of the following could be the reason for this? Circle your answer. Circle your answer. Circle your answer. The mass of object 1 is: larger smaller doesn't matter The height of object 1 is: larger smaller doesn't matter

example, only 45% of participants recalled that atomic mass is an explanatory factor for mass density. However, when they were presented with atomic mass, 100% recognize it as explanatory. This gap between availability and accessibility can play a critical role in reasoning about multiple factors, as we shall see in experiment 2.

Second, the normative correctness of the factors plays a role in student responding, and minding the normative correctness of the factors can be helpful for interpreting the results. Consider the mass density, balance scale, and gravitational potential energy scenarios. For each of these scenarios, the two most listed factors are physically correct. In these cases, almost *all* of the students recognized that both factors were explanatory; virtually none of them respond that only one factor is explanatory. For each of the remaining three scenarios, it is important to note that one of the most common factors is *not* physically correct, and we do not see 100% of the students recognizing both factors. One could interpret these results with a simple model of two populations

of participants: one with the physically correct model of the scenario and the other without the physically correct model. For example, for the pendulum scenario, it is reasonable to judge that about 23% of the participants recognize that length plays a role in the period and mass does not. Thus, only 75% (not 100%) recognize both mass and length as explanatory. It is interesting to note that even for physically incorrect factors, such as mass in this scenario, participants recognize the factor more frequently than they list it.

III. EXPERIMENT 2: REASONING WITH HIGH OR LOW ACCESSIBILITY EXPLANATIONS

Experiment 1 demonstrated that for four of the six physical scenarios studied, there is a significant difference in accessibility between the top two explanatory factors on average for this student population. The goal of experiment 2 is to use a between-student design to test our main hypothesis, the proposed accessibility rule: whether alternative explanations

TABLE III. Experiment 1 recall condition results. The two most common factors listed as explaining the outcomes for each task. Percentages of participants explicitly listing (recalling) each combination are in parentheses. The factors with significantly relatively higher accessibility (strength) are in bold. Note that the factors with asterisks are not physically correct.

Physical scenario	N	Factor listed	Only factor listed	Factor listed first	Both factors listed
Mass density	51	Spacing (70) Atomic mass (45)	Spacing (37) Atomic mass (12)	Spacing (55) Atomic mass (27)	Spacing and Atomic mass (25)
Balance scale	92	Mass (96) Position (55)	Mass (29) Position (0)	Mass (90) Position (7)	Mass and distance (55)
Pendulum	84	Length (67) Mass* (48)	Length (12) Mass* (5)	Length (49) Mass* (21)	Length and mass (32)
Projectile	80	Launch angle (61) Mass* (44)	Launch angle (35) Mass* (15)	Launch angle (58) Mass* (26)	Launch angle and Mass (21)
Sliding	84	Mass* (79) Coefficient of friction (79)	Mass* (17) Coefficient of friction (15)	Mass* (54) Coefficient of friction (40)	Mass and Coefficient of friction (60)
Gravitational Potential Energy	41	Mass (98) Height (90)	Mass (2) Height (0)	Mass (59) Height (34)	Mass and height (90)

TABLE IV. Experiment 1 recognize condition results. Percentages of participants that recognize that various *given* factors or combinations of factors can explain given outcomes are in parentheses. Note that the factors with asterisks are not physically correct.

Physical scenario	N	Only one factor recognized	Both factors recognized
Mass density	44	Spacing (0) Atomic mass (0)	Spacing and atomic mass (100)
Balance scale	44	Mass (0) Position (0)	Mass and distance (100)
Pendulum	44	Length (23) Mass* (2)	Length and mass* (75)
Projectile	44	Launch angle (30) Mass* (2)	Angle and mass* (64)
Sliding	44	Mass* (0) Coefficient of friction (11)	Mass* and μ (89)
Gravitational potential energy	44	Mass (0) Height (2)	Mass and height (98)

are considered less frequently when a relatively high accessibility factor is offered as the explanation. We test the accessibility rule using both multiple-choice and short-answer formats in order to determine whether the responses are sensitive to these differences and possibly broaden the validity of the results. Furthermore, we will also determine the extent to which the data confirm or invalidate our more specific corollaries to the accessibility rule discussed in Sec. I.

A. Method and materials

The participants were randomly drawn from the same pool as experiment 1. Participants were assigned to one of four conditions: two conditions in the multiple-choice format and two in the short-answer format (see Table V).

The two conditions in each of the answer formats (multiple choice or short answer) included questions from each of the six scenarios. In one condition, three low and three high accessibility explanations were provided. In the second condition, the compliment of this was given for each scenario. The number of participants in each format and

TABLE V. Experiment 2 conditions.

Condition	Accessibility of explanation	N
Free response format, item set 1	High or low	134
Free response format, item set 2	Low or high	98
Multiple choice format, item set 1	High or low	90
Multiple choice format, item set 2	Low or high	91

scenario is presented in the results section. Similar to experiment 1, the numbers of participants are not uniform across formats or scenarios because of scheduling constraints and some design errors in early versions of the tasks. Additionally, because of time constraints, a few students did not complete all questions in a given condition.

The questions provided the participant with a physical scenario and with either a high or low accessibility explanation and then probed whether the participant thought the explanation was valid. A list of all questions administered in all formats (both question types and with high or low accessibility explanations) are shown in the Appendix (Tables XIV and XV). For example, in the short-answer format, one question providing a highly accessible explanation states the following:

Block *X* and block *Y* are each made of a different material. The material of block *X* has higher mass density (kg/m^3) than the material of block *Y*. Is the statement below a valid conclusion based on the information given? Briefly explain.

“The atoms in block X have a smaller average separation than the atoms in block Y.”

We hypothesize, via the accessibility rule, that participants will accept this highly accessible explanation (rather than claim that some alternative explanation could explain the difference in density) more frequently than participants shown the low accessibility explanation *“The atoms in block X have a larger mass than the atoms in block Y.”* For this latter explanation we hypothesize that relatively more participants will state that there is an alternative explanation: the atoms may also be closer together, therefore the conclusion is not valid. On the other hand, consider the short-answer format for the sliding scenario:

Object *P* and object *Q* are both given a quick push and slide on a wooden floor with the same initial velocity.

Object *P* comes to rest before object *Q*. Is the statement below a valid conclusion based on the information given? Briefly explain.

(Explanation 1, condition1):

“The coefficient of friction is greater between object P and the floor than between object Q and the floor.”

(Explanation 2, condition 2):

“Object P has a greater mass than object Q does.”

Since, as demonstrated in experiment 1, both explanations had the same relative accessibility (despite the second one being physically invalid), the accessibility rules predict that there should be no difference in the frequency with which participants offer alternative explanations.

B. Results: Testing the accessibility rule

The results in Fig. 1 present the percentage of participants who offered alternative explanations when initially provided with high or low accessibility explanations in the short answer condition. These results support the accessibility rule that students consider alternative explanations less frequently when given relatively highly accessible explanations. Specifically, for three of the four scenarios with significantly different explanation accessibilities (population proportion test $p < 0.001$), there is a dramatic 40% difference (roughly 1 standard deviation) between the percentage of participants offering alternative explanations. There is still a significant difference ($p < 0.05$, Bonferroni adjusted) between the low and high availability factors for the fourth scenario (balance scale), though it is smaller, around 15% (~ 0.4 standard deviations). Further, as predicted, the two scenarios with no significant difference in accessibility in the explanations showed no significant difference in participants offering

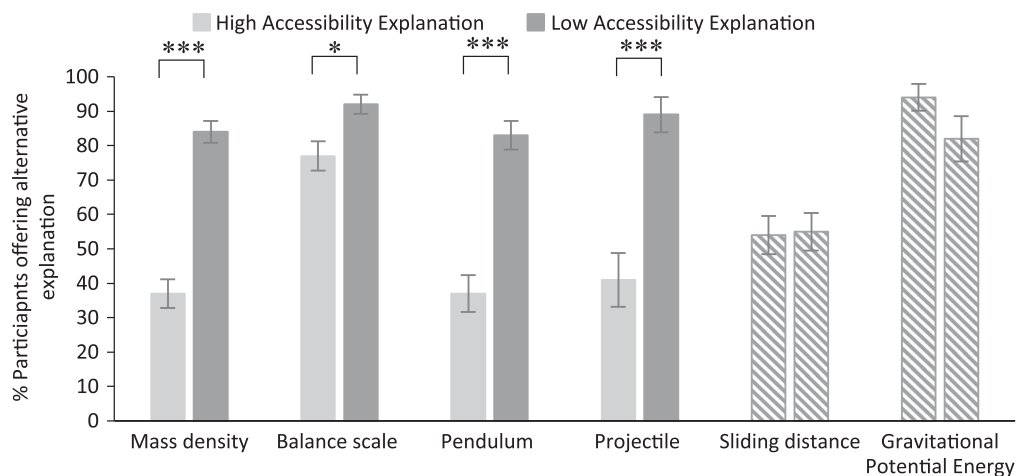


FIG. 1. Experiment 2 results for the short answer format: Percentages of participants who indicated alternative explanations may exist when presented with an explanatory factor with high or low accessibility. Hashed bars indicate scenarios in which experiment 1 determined no significant difference between the accessibilities of the explanations.

TABLE VI. Experiment 2 percentages (and ratios) of participants who indicated alternative explanations may exist when presented with an explanatory factor with the indicated accessibility. The p values are not Bonferroni adjusted. Note that the factors with asterisks are not physically correct.

Physical scenario	Relative accessibility of explanation	Short answer format	Short answer p value	Multiple-choice format	Multiple-choice p value
Density	High (spacing)	37 (50/134)	<0.001	23 (21/90)	<0.001
	Low (atomic mass)	84 (112/133)		69 (62/90)	
Balance scale	High (mass)	77 (75/98)	0.004	53 (48/91)	<0.001
	Low (distance)	92 (87/95)		81 (73/90)	
Pendulum	High (length)	37 (30/82)	<0.001	18 (7/39)	<0.001
	Low (mass*)	83 (68/82)		65 (26/40)	
Projectile motion	High (launch angle)	41 (16/39)	<0.001	23 (9/39)	0.01
	Low (mass*)	89 (33/37)		50 (20/40)	
Sliding	High (mass*)	54 (44/81)	0.9	20 (8/40)	0.3
	Low (μ)	55 (45/82)		31 (12/39)	
Gravitational potential energy	High (mass)	95 (36/38)	0.1	85 (34/40)	0.5
	Low (height)	82 (28/34)		79 (31/39)	

alternative explanations (Bonferroni adjusted population proportion test $p > 0.5$).

As an aside, one might comment that the difference in claiming alternative explanations for the pendulum and projectile scenarios could be interpreted at least in part by the fact that some students knew the physically correct answer (which here are the high accessibility explanations) thus the consideration of alternatives would consequently be suppressed since there are no other (reasonably expected) explanations, thus this percentage should be lower, as seen in the results. While this is true, the power of the accessibility rule in this paper is that it already includes this possibility: for participants who “know” that only one factor of the two is valid, that factor will have high (100%) accessibility and the other will have low (0%) accessibility.

Table VI presents the more detailed results for both question formats, revealing similar results for both answer formats, further supporting the accessibility rule. For all physical scenarios accepting the gravitational potential energy and sliding scenarios, there are significant differences between low and high accessibility explanations at the $p \leq 0.05$ level using the population proportion test and a *post hoc* Bonferroni correction for the comparisons made.

C. Results: Testing the corollaries

One might reasonably expect there to be a more quantitative relationship between the accessibility of given explanation (experiment 1) and the likelihood that a reasoner will consider alternatives when that explanation or an alternate explanation is given (experiment 2). Such correlations are proposed in Corollaries 1 and 2 in Sec. I.

One issue is that experiments 1 and 2 are two separate groups of students. That is, we do not have within-student

data on measures of accessibility and considering alternatives. Instead we only have between-student data: population average accessibility for each explanation from the students in experiment 1 and population average of offering alternative explanations from the students in experiment 2. Nonetheless, we can test these corollaries by comparing the mean availability data for each scenario and factor from experiment 1 to the corresponding means of offering alternative explanations from experiment 2. That is, we can *predict* the mean percentage of students (in experiment 2) who will consider an alternative explanation by examining the mean availability of explanations from the students in experiment 1—a different set of students.

Specifically, one can construct a simple quantitative model that follows from Corollary 1: for a given scenario, the average probability that a participant will consider an alternative explanation when given explanation A is equal to the average probability that a participant will list a different explanation B when asked to list explanations for that scenario. Recall here that for reasons of simplicity, we are considering only two possible explanations. Mathematically, this is (for a given scenario)

$$P_{\text{ave}}[\text{consider alternative, given } A] = P_{\text{ave}}[\text{list } B]. \quad (1)$$

Likewise, one could construct a simple model that follows Corollary 2: for a given scenario, the average probability that a participant will consider an alternative explanation when given explanation A is equal to one minus the average probability that a participant listed A *only* when asked to list explanations for the scenario. Mathematically, this is (for a given scenario)

$$P_{\text{ave}}[\text{consider alternative, given } A] = 1 - P_{\text{ave}}[\text{list } A \text{ only}]. \quad (2)$$

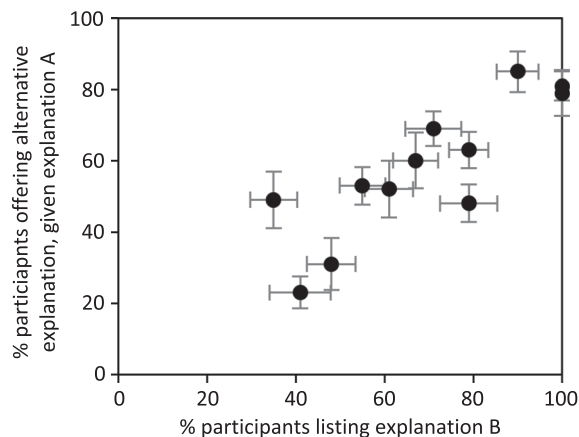


FIG. 2. Compilation of results from experiment 1, Table III (horizontal axis) and experiment 2, Table VI (vertical axis) in order to test Corollary 1. Each point represents a high or low accessibility explanation for each physical scenario. Error bars represent standard errors.

To test these two corollaries, we matched the accessibility data from experiment 1 with the consideration of alternatives from experiment 2 for each scenario. Since each scenario had high and low accessibility explanations, this resulted in 12 data points (6 scenarios, two explanations per scenario). We plotted the points on a graph to determine whether the trend was consistent with Eqs. (1) and (2).

Figures 2 and 3 present the results, which provides evidence supporting both corollaries. Essentially, Figs. 2 and 3 are plotting the points from experiment 1, Table III on the horizontal axis, and matching them with the corresponding points from experiment 2, Table VI on the vertical axis. We used on the multiple-choice format data, but the free response data gives similar results. To analyze for a fit, we performed linear regressions (with bootstrapping, and weighted by inverse standard error) on the twelve

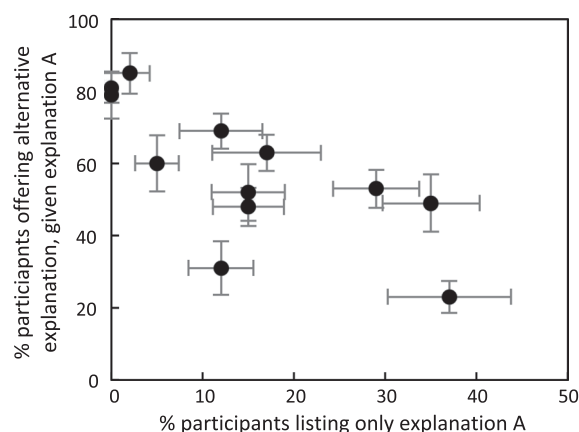


FIG. 3. Compilation of results from experiment 1, Table III (horizontal axis) and experiment 2, Table VI (vertical axis) in order to test Corollary 2. Each point represents a high or low accessibility explanation for each physical scenario. Error bars represent standard errors.

data points for each graph and found that the coefficient for $P_{ave}[\text{list A}]$ in Eq. (1) is consistent with unity (95% confidence, $0.4 < \text{coef.} < 1.1$) and the intercept is consistent with zero (95% confidence, $-17 < \text{intercept} < 29$). Likewise for Corollary 2, the coefficient for $P_{ave}[\text{list A only}]$ in Eq. (2) is consistent with -1 (95% confidence, $-1.8 < \text{coef.} < -0.5$), however, the intercept is not consistent with unity (95% confidence, $0.6 < \text{intercept} < 0.8$, where we converted from percent to proportion).

It is important to note that Eqs. (1) and (2) were only constructed for purposes of developing a more quantitative test of the two corollaries. The important point to take away from these results is the striking and somewhat regular dependence of the consideration of alternatives on the independently and empirically determined accessibility of available explanations. In fact, the linear regressions shown in Figs. 2 and 3 result in explaining 71% and 54% of the variance in considering alternatives, respectively. Put another way, the correlation between considering an alternative to explanation B and the accessibility of explanation A is quite large at 0.84.

IV. EXPERIMENT 3: REPLICATION AND EXTENSION

The goal of experiment 3 is to increase the external validity of the results by replicating and extending experiments 1 and 2 for a different student population and for a small variation in the wording of the tasks. Additionally, it aims to determine whether, for the six physical scenarios studied here, the effects of accessibility depend on “predictive” reasoning from cause to effect compared to “diagnostic” reasoning from effect to cause. This latter extension is motivated by the prior results indicating that the direction of reasoning does affect the influence of relative accessibility, as discussed in the introduction.

A. Method and materials

The method and materials were identical to the combination of experiments 1 and 2, except for four modifications (see Table VII for an overview of the design). First, the pool of participants in experiment 3 was drawn from the first semester of an algebra-based introductory physics course (rather than the calculus-based course). These students tend to be less prepared mathematically and tend to have different career interests: the majority of algebra-based students are interested in healthcare professions and life sciences, while the overwhelming majority of calculus-based students are interested in engineering and the physical sciences.

The second modification was that only the multiple-choice format was used for the consider-the-alternative tasks since there were no notable differences in the results between the short answer and multiple-choice formats in experiment 2. The third modification was the addition of conditions with slightly different wording for the consider-the-alternative

TABLE VII. Experiment 3 conditions.

Condition	Accessibility of explanation	<i>N</i>
Recall		61
Recognize		63
Backward reasoning (diagnose), standard wording	High or low	63
Backward reasoning (diagnose), standard wording	Low or high	60
Backward reasoning (diagnose), alternate wording	High or low	55
Backward reasoning (diagnose), alternate wording	Low or high	58
Forward reasoning (predict), standard wording	High or low	62
Forward reasoning (predict), standard wording	Low or high	60

tasks. For each multiple-choice question in experiment 2, the third choice always began with the phrase “It is not certain whether...” For example, in the strongly accessible version of the mass density question the third choice was “It is not certain whether the atoms in block *X* have a *larger or smaller average separation* than the atoms in block *Y*.” In addition to this condition, experiment 3 adds conditions in which “It is not certain whether...” is replaced with the phrase “It is possible that...” We added this third modification because we were concerned that the somewhat negative phrasing “it is not certain” may be suppressing participants from full consideration of this option.

The final modification was the addition of tasks in which the reasoning was in the opposite direction of those in experiment 2. For example, in the pendulum task in experiment 2, an outcome was stated (difference in period) and the participant was asked to consider a given explanatory factor (difference in mass of the pendula). In experiment 3, we added additional conditions in which the reasoning was reversed, that is and explanatory factor was given (mass of pendulum *A* is greater than pendulum *B*), and the participant was asked about the outcome with regards to period. See Tables XV and XVI Appendix for a sample of the questions administered.

B. Results

The results of the recall and recognition conditions are presented in Tables VIII and IX. These results are somewhat similar to the results discussed in experiment 1, with some small but notable exceptions. Specifically, the designation for high and low accessibility explanations is the same except for two physical scenarios. First, for this population, there is a significant difference in accessibility between the two explanatory factors for the gravitational potential energy scenario. This was not the case for experiment 1. Second, unlike experiment 1, no difference was found in accessibility for the explanations for the pendulum scenario using our standard approach of comparing the relative frequency each factor was listed. However, it should be noted that comparing which explanation was listed first or “only” reveals that length may be more accessible than mass ($p < 0.05$), at least by these two measures.

In comparing the standard wording format to the alternate wording format, we only found significant differences in absolute scores for the density and the balance scale scenarios, (chi-squared test, Bonferroni corrected $p < 0.05$); however, the relative scores for the high and low accessibility explanations

TABLE VIII. Experiment 3 recall condition results. The two most common factors listed as explaining the outcomes for each task. Percentages of participants explicitly listing (recalling) each combination are in parentheses. The factors with significantly relatively higher accessibility (strength) are in bold. Note that the factors with asterisks are not physically correct.

Physical scenario	<i>N</i>	Factor listed	Only factor listed	Factor listed first	Both factors listed
Mass density	61	Spacing (70) Atomic mass (49)	Spacing (25) Atomic mass (5)	Spacing (51) Atomic mass (25)	Spacing and atomic mass (41)
Balance scale	61	Mass (93) Position (62)	Mass (30) Position (5)	Mass (87) Position (13)	Mass and distance (54)
Pendulum	61	Length (52) Mass* (53)	Length (20) Mass* (3)	Length (34) Mass* (21)	Length and mass (21)
Projectile	61	Launch angle (66) Mass* (20)	Launch angle (57) Mass* (11)	Launch angle (57) Mass* (21)	Launch angle and mass (5)
Sliding	61	Mass* (79) Coefficient of friction (77)	Mass* (8) Coefficient of friction (8)	Mass* (54) Coefficient of friction (38)	Mass and coefficient of friction (59)
Gravitational potential energy	61	Mass (92) Height (74)	Mass (20) Height (5)	Mass (67) Height (30)	Mass and height (69)

TABLE IX. Experiment 3 recognize condition results. Percentages of participants that recognize that various *given* factors or combinations of factors can explain given outcomes are in parentheses. Note that the factors with asterisks are not physically correct.

Physical scenario	N	Only one factor recognized	Both factors recognized
Mass density	63	Spacing (21) Atomic mass (13)	Spacing and atomic mass (67)
Balance scale	63	Mass (6) Position (5)	Mass and distance (89)
Pendulum	63	Length (41) Mass* (10)	Length and mass* (49)
Projectile	63	Launch angle (49) Mass* (10)	Angle and mass* (33)
Sliding	63	Mass* (6) Coefficient of friction (16)	Mass* and μ (78)
Gravitational potential energy	63	Mass (11) Height (13)	Mass and height (76)

remained unchanged for the two formats for all scenarios. Thus, in the rest of our analysis, we combined the results of the conditions from both formats. In a sense, this could be thought of as counterbalancing across wording formats.

The results for the diagnose conditions (with both wording formats combined) is presented in Fig. 4, with more detailed numeric results in Table IX. There are several conclusions to make from these results. First, the results are very similar to experiment 2 and are aligned with the accessibility rule, except for two scenarios. The pendulum scenario shows a significant difference in alternative explanation scores between the two factors. However, as mentioned earlier, the recall condition data has mixed results regarding the difference in accessibility of the two explanations, and two of the measures for accessibility indicate a difference in accessibility that predicts (via the accessibility rule) the observed differences in scores for the pendulum scenario in Fig. 4. The gravitational potential energy scenario results, however, cannot be explained

by the accessibility rule, since we would predict that the low accessibility explanation would have a higher alternative explanation score. Yet, no such difference is observed, and, in fact, the data indicate a slight trend in the opposite direction. Thus, this scenario indicates that, for at least some situations, there may be other factors that are more important than accessibility, as measured by our free recall tasks.

The results for the predict conditions are shown in Table X. Inspection of these results reveals no differences in the patterns in the scores compared to the diagnose conditions and these results appear to contradict the results of Fernbach *et al.* [11]. A population proportion test reveals that the overall scores are only significantly different for the sliding and gravitational potential energy scenarios, with participants scoring lower in the predict condition. However, since the differences between scores between the two conditions varies somewhat, it is not clear if there is a systematic pattern.

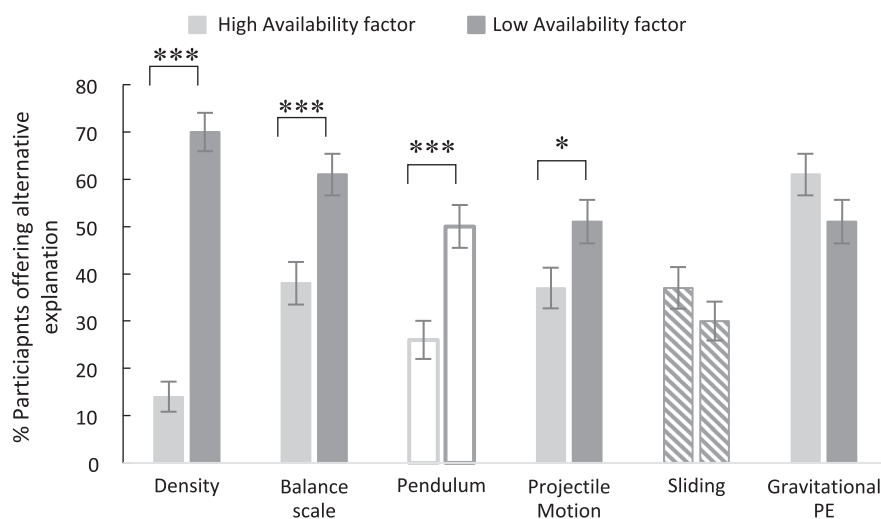


FIG. 4. Experiment 3 results for the multiple-choice format. Hashed bars indicate the scenario in which no significant difference between the accessibilities of the explanations was determined. Outlined bars are used for the pendulum scenario, in which the standard measure of differences between accessibilities showed no significance, but other measures did.

TABLE X. Experiment 3 percentages (and ratios) of participants who indicated alternative explanations or predictions may exist when presented with an explanatory factor with the indicated accessibility. The p values are not Bonferroni adjusted. Note that the factors with asterisks are not physically correct.

Physical scenario	Relative accessibility of explanation	Diagnose format	Diagnose p value	Predict format	Predict p value
Density	High (spacing)	14 (17/120)	>0.001	21 (12/58)	>0.001
	Low (atomic mass)	70 (88/125)		64 (35/55)	
Balance scale	High (mass)	38 (46/120)	>0.001	45 (26/58)	0.05
	Low (distance)	61 (76/125)		64 (35/55)	
Pendulum	High (length)	26 (31/120)	>0.001	26 (15/58)	0.05
	Low (mass*)	50 (63/125)		44 (24/55)	
Projectile motion	High (launch angle)	37 (46/125)	0.03	24 (13/55)	0.01
	Low (mass*)	51 (61/120)		47 (27/58)	
Sliding	High (mass*)	37 (44/120)	0.2	10 (6/58)	0.5
	Low (μ)	30 (37/125)		14 (8/55)	
Gravitational potential energy	High (mass)	61 (76/125)	0.1	36 (20/55)	0.7
	Low (height)	51 (61/120)		40 (23/58)	

In sum, the accessibility rule is further supported by experiment 3, which replicated the results of experiments 1 and 2 for a different population, with some small exceptions. Further, we found no notable differences between the diagnose and predict formats in regards to how accessibility affects the consideration of alternative explanations or predictions.

V. EXPERIMENT 4: WITHIN-STUDENT DESIGN

Up to this point all of the experiments have used between-student designs by determining the average accessibility with one group of participants and the consideration of alternatives with a different group of participants. To further test the accessibility rule, experiment 4 employs a within-student design. This way we can, for example, directly compare participants who consider an alternative to explanation A in one task *and* only list explanation A in another task to participants who consider an alternative to explanation A in one task *and* do not only list explanation A in another task. The accessibility rule predicts that the average frequency of the latter will be significantly greater than the former.

A. Methods and materials

The methods and materials used are identical to those in the recall, and the diagnose, standard wording conditions in

experiment 3, except for one modification: Each participant first completes the recall task and two weeks later the same participant completes one of the diagnose standard wording tasks. We included a delay between the two tasks in order to minimize possible effects of priming. See Table XI for an overview of the design.

B. Results

We compiled responses for all six physical scenarios and tabulated the number of times participants listed factor A (or A only) in the recall task and correspondingly whether they considered an alternative explanation when factor B (or A) was presented in the diagnose task. This is reported in Tables XII and XIII.

The results from these tables provide further support for the accessibility rule and its corollaries. Specifically, the results of Table XI and an odds ratio test reveals that the odds that a participant considered an alternative to explanation A is 2.6 times greater (95% conf. interval: [1.8, 3.7]) if that participant listed factor B compared to if they did not list factor B , consistent with Corollary 1.

Similarly, Table XII reveals that the odds that a participant considered an alternative to explanation A is reduced by a factor of 0.35 (95% conf. interval: [0.21, 0.57]) if that participant listed factor A only compared to if other factors were listed, consistent with Corollary 2.

TABLE XI. Experiment 4 design.

Condition	Task 1		Task 2	N
Condition 1	Recall	Two-week	Diagnose, high or low accessibility	50
Condition 2	Recall	Interim	Diagnose, low or high accessibility	52

TABLE XII. Experiment 4 average proportion of times participants did (or did not) consider an alternative explanation when factor *A* was provided as explanation, for participants who listed (or did not list) explanation *B* in the recall task. This includes 102 participants answering 6 scenarios each.

		Considered alternative to <i>A</i>	
		No	Yes
Listed factor <i>B</i>	No	0.25	0.10
	Yes	0.33	0.32

TABLE XIII. Experiment 4 average proportion of times participants did (or did not) consider an alternative explanation when factor *A* was provided as explanation, for participants who listed (or did not list) explanation *A* only in the recall task. This includes 102 participants answering 6 scenarios each.

		Considered alternative to <i>A</i>	
		No	Yes
Listed factor <i>A</i> only	No	0.45	0.38
	Yes	0.13	0.04

VI. CONCLUSIONS AND DISCUSSION

There are two main results from this study. The first is that in the context of physical scenarios commonly found in physics education curriculum, we provided systematic, quantitative evidence for the accessibility rule, namely, that alternative explanations are considered less frequently when an explanatory factor with relatively high accessibility is offered first. In fact, we demonstrated that the accessibility rule is a quantitatively predictive model: for the population of students studied, the average accessibility of the factors in each scenario can explain about 70% of the variance of the average proportion of students considering alternative explanations. Further, for a given student, the odds that the student will consider an alternative to explanation *A* increases by about a factor of 3, if the explanatory factor *B* is accessible to this student. We replicated this effect of accessibility with multiple formats in two populations of students (algebra- and calculus-based intro courses) and for both predictive and diagnostic reasoning tasks.

These results are not unexpected. As discussed in the introduction, a number of science education researchers have recognized the importance of context and accessibility on reasoning, and there are a number of cognitive psychology studies that have demonstrated the effect of accessibility on the consideration of alternatives. What is novel about this study is the explicit articulation of a testable cognitive accessibility rule and direct application and the empirical, quantitative demonstration of this phenomenon for physical scenarios relevant to physics education. In a sense, this study provides explicit and concrete examples of a basic cognitive mechanism at work for relevant physics education content.

Naturally, it is important to determine whether the results and mechanisms investigated in cognitive psychology studies apply to science education, because the two fields present different contexts and issues. Consider a typical everyday scenario used in the cognitive psychology study of Klaczynski and Daniel [29]: “If a person eats all the time, then he will gain weight.” In this case there is a significant amount of ambiguity: there could be many other possible causes for the outcome of weight gain, none of which are typically explicitly considered (disease, medications, etc.), and rarely are any of the imagined causes “incorrect.” Other cognitive psychology studies such as Fernbach, Darlow, and Sloman [11], speak of the participants being “sensitive” to the presence or relative strength of alternatives (such as in examples of medical diagnosis). In contrast, in the science education context, most scientific concepts are framed in a manner in which one or more factors (which in many cases may be viewed causally) are explicitly related to an outcome, and accounting for all such factors is necessary for determining the outcome. For example, considering the pressure of an ideal gas requires an accounting of three factors: the number of atoms, the volume, and the temperature. The importance of systematically considering *all* relevant alternative factors is a central feature of scientific reasoning.

Furthermore, for science scenarios, there are often explanations that are highly accessible to students that are physically incorrect. For example, many students list “mass” as a factor for the period of a pendulum. In this case, the relative accessibility of mass is playing a role in reasoning when, in fact, it should not be considered at all, at least not after instruction. As such, the issue of *incorrect factors* is an important one in science education that is rarely if ever considered in cognitive psychology. Interestingly, our study here was blind to this issue of incorrect factors. We treated all factors the same and found that the effect of accessibility still holds. Nonetheless, in the context of science education, it is sometimes the instructional goal that students *do not* consider alternatives, because there are no other physically correct ones within the context given. Thus, the consideration of alternatives is not a universally desired outcome for all scenarios relevant to science reasoning.

A minor result of this study, as noted in experiment 3, was that we found no significant interaction between accessibility and predictive vs diagnostic reasoning, contrary to Fernbach *et al.* [11]. The reason for the difference in our finding from their results is not clear but may have to do with the issues of the kinds of contexts used in that study, as discussed above. Overall this is a potential area for further investigation, since understanding the differences between posing predictive vs. diagnostic questions has practical instructional implications.

The second main result we found is that students recognize explanatory factors for physical scenarios (from a list) significantly more frequently than they can free-recall those factors (by listing them), even when the factors are physically incorrect. We have interpreted these results

in terms of the cognitive psychological framework that explanations maybe be available to students but not accessible. In short, recognition (availability) of an explanation is not sufficient; easy recall of the explanation (accessibility) plays a critical role for productive reasoning.

It is important to note, however, that the accessibility effect is a phenomenon that is distinct from the idea of explicit formal or informal theories that reasoners may “have” in their minds. This is a comment on influential and important work that discusses such formal and informal theories that students may use in the process of reasoning [15,30]. The fact that there is an operationalizable difference between the availability and accessibility of an explanatory factor calls into question what is meant by a person “having a theory.” Do they have a theory if it is available but not accessible for a given context? Consider the mass density scenario. If a student appropriately recognizes from a list that both atomic mass and atomic spacing affect mass density, but this same student does not consider both spacing and mass when reasoning about a mass density scenario, does this student “have” the correct conception of mass density? The difference between availability and accessibility may suggest more implicit processes at work that are different from the idea of explicit thinking and reasoning.

We would suggest that a productive analogy may be that accessibility is a “soft contour” that affects the path of the reasoner through the process of reasoning. While a set of knowledge may be available, accessibility is a contour that implicitly guides the use of that knowledge. Further, this effect of accessibility does not appear to be explicit but rather to be due to rapid and automatic processes.

Let us consider one final reason why the effect of accessibility may be relevant for science education. The effect of accessibility may cause an error in just one of the links in a chain of reasoning, rendering the whole chain invalid. Problem solving and conceptual understanding in

science often requires multistep processes or chains of reasoning in which a number of decisions must be made by the student either implicitly or explicitly, that is, with a combination of heuristic and analytic processes. For example, Speirs *et al.* [31] describe a multistep chain of reasoning associated with a kinematics graph task, and Kryjevskaja, Stetzer, and Grosz [28] describe multistep reasoning paths that appear to include both heuristic and analytic processes for capacitor and wave pulse tasks. In an example more closely related to a physical scenario studied here, Rosenblatt, Heckler, and Flores [32] found that, even post-instruction, many students in an introductory materials science engineering course believe that high mass density implies high melting temperature, and this conclusion is produced via the physically incorrect line of “reasoning” that high mass density implies small atomic separation, and this implies high atomic bond strength, which in turn implies high melting temperature. Notice that the first implication in the chain does not necessarily follow and could be seen as the failure to consider the alternative explanation that high mass density may be due to composition of elements with high atomic number, which was commonly seen in our current study. Therefore, this is a good candidate case for explaining a documented student difficulty via the mechanism of relative accessibility inhibiting students from considering alternatives and producing physically correct arguments. As such, this study may provide insight and may help lead to more productive interpretations of student reasoning and effective instructional interventions.

ACKNOWLEDGMENTS

Funding for this research was provided by the Center for Emergent Materials: an NSF MRSEC under Grant No. DMR-1420451.

APPENDIX: SAMPLE REASONING TASKS

Full versions of items used in the reasoning tasks in this study are shown below.

TABLE XIV. Experiment 2 short answer format questions. Participants answered either a high or low accessibility explanation for each scenario.

Physical scenario	Relative accessibility of explanation	Question
Mass density	High	Block X and block Y are each made of a different material. The material of Block X has higher mass density (kg/m^3) than the material of block Y. Is the statement below a valid conclusion based on the information given? Briefly explain. “The atoms in block X have a smaller average separation than the atoms in block Y.”
	Low	Block X and block Y are each made of a different material. The material of Block X has higher mass density (kg/m^3) than the material of block Y. Is the statement below a valid conclusion based on the information given? Briefly Explain. “The atoms in block X have a larger mass than the atoms in block Y.”

(Table continued)

TABLE XIV. (Continued)

Physical scenario	Relative accessibility of explanation	Question
Balance scale	High	A rod is balanced on a pivot. A student hangs object <i>A</i> somewhere on the left side of the rod and hangs object <i>B</i> somewhere on the right side of the rod. The right-hand side immediately starts to tilt down. Is the statement below a valid conclusion based on the information given? Briefly Explain. “Object <i>A</i> has a greater mass than object <i>B</i> does.”
	Low	A rod is balanced on a pivot. A student hangs object <i>A</i> somewhere on the left side of the rod and hangs object <i>B</i> somewhere on the right side of the rod. The right-hand side immediately starts to tilt down. Is the statement below a valid conclusion based on the information given? Briefly explain. “Object <i>A</i> is hung farther from the center pivot than object <i>B</i> is.”
Pendulum	High	Pendulum <i>A</i> swings with a longer period (time) than pendulum <i>B</i> . Both are simple pendulums. Is the statement below a valid conclusion based on the information given? Briefly explain. “Pendulum <i>A</i> has a longer string than pendulum <i>B</i> .”
	Low	Pendulum <i>A</i> swings with a longer period (time) than pendulum <i>B</i> . Both are simple pendulums. Is the statement below a valid conclusion based on the information given? Briefly explain. “Pendulum <i>A</i> has a greater mass than pendulum <i>B</i> does.”
Projectile	High	Projectile <i>A</i> and projectile <i>B</i> start at ground level and are thrown with the same speed, but <i>A</i> is in the air for a longer time than <i>B</i> . Ignore any effects of air resistance or drag. Is the statement below a valid conclusion based on the information given? Briefly Explain. “The launch angle (from horizontal) of projectile <i>A</i> is greater than that of projectile <i>B</i> .”
	Low	Projectile <i>A</i> and projectile <i>B</i> start at ground level and are thrown with the same speed, but <i>A</i> is in the air for a longer time than <i>B</i> . Ignore any effects of air resistance or drag. Is the statement below a valid conclusion based on the information given? Briefly explain. “Projectile <i>A</i> has a larger mass than projectile <i>B</i> .”
Sliding distance	Equal	Object <i>P</i> and object <i>Q</i> are both given a quick push and slide on a wooden floor with the same initial velocity. Object <i>P</i> comes to rest before object <i>Q</i> . Is the statement below a valid conclusion based on the information given? Briefly explain. “The coefficient of friction is greater between object <i>P</i> and the floor than between object <i>Q</i> and the floor.”
	Equal	Object <i>P</i> and object <i>Q</i> are both given a quick push and slide on a wooden floor with the same initial velocity. Object <i>P</i> comes to rest before object <i>Q</i> . Is the statement below a valid conclusion based on the information given? Briefly explain. “Object <i>P</i> has a greater mass than object <i>Q</i> does.”
Gravitational potential energy	Equal	Object 1 has larger gravitational potential energy than object 2. Is the statement below a valid conclusion based on the information given? Briefly explain. “Object 1 has a larger mass than object 2.”
	Equal	Object 1 has larger gravitational potential energy than object 2. Is the statement below a valid conclusion based on the information given? Briefly explain. “Object 1 is higher up than object 2.”

TABLE XV. Experiment 2 and Experiment 3 multiple-choice (diagnostic) answer format questions. Participants answered either a high or low accessibility explanation for each scenario.

Physical scenario	Relative accessibility of explanation	Question
Mass density	High	Block <i>X</i> and block <i>Y</i> are each made of a different material. The material of Block <i>X</i> has a higher mass density (kg/m^3) than the material of block <i>Y</i> . Which of the following statements comparing the blocks is most accurate? (a) The atoms in block <i>X</i> have a <i>larger average separation</i> than the atoms in block <i>Y</i> . (b) The atoms in block <i>X</i> have a <i>smaller average separation</i> than the atoms in block <i>Y</i> . (c) It is not certain whether the atoms in block <i>X</i> have a <i>larger or smaller average separation</i> than the atoms in block <i>Y</i> .
	Low	Block <i>X</i> and block <i>Y</i> are each made of a different material. The material of block <i>X</i> has a higher mass density (kg/m^3) than the material of block <i>Y</i> . Which of the following statements comparing the blocks is most accurate? (a) The atoms in block <i>X</i> have a <i>larger mass</i> than the atoms in block <i>Y</i> . (b) The atoms in block <i>X</i> have a <i>smaller mass</i> than the atoms in block <i>Y</i> . (c) It is not certain whether the atoms in block <i>X</i> have a <i>larger or smaller mass</i> than the atoms in block <i>Y</i> .
Balance scale	High	A rod is balanced on a pivot. A student hangs object <i>A</i> somewhere on the left side of the rod and hangs object <i>B</i> somewhere on the right side of the rod. The right-hand side immediately starts to tilt down. Which of the following statements comparing the objects is most accurate? (a) Object <i>B</i> on the right side has a <i>larger mass</i> than object <i>A</i> on the left side. (b) Object <i>B</i> on the right side has a <i>smaller mass</i> than object <i>A</i> on the left side. (c) It is not certain whether object <i>B</i> on the right side has a <i>larger or smaller mass</i> than object <i>A</i> on the left side.
	Low	A rod is balanced on a pivot. A student hangs object <i>A</i> somewhere on the left side of the rod and hangs object <i>B</i> somewhere on the right side of the rod. The right-hand side immediately starts to tilt down. Which of the following statements comparing the objects is most accurate? (a) Object <i>B</i> on the right side is <i>farther from the center pivot</i> than object <i>A</i> on the left side. (b) Object <i>B</i> on the right side is <i>closer to the center pivot</i> than object <i>A</i> on the left side. (c) It is not certain whether object <i>B</i> on the right side is <i>farther from or closer to the center pivot</i> than object <i>A</i> on the left side.
Pendulum	High	Pendulum <i>A</i> swings with a longer period (time) than pendulum <i>B</i> . Both are simple pendulums. Which of the following statements comparing the pendulums is most accurate? (a) Pendulum <i>A</i> is <i>longer than</i> pendulum <i>B</i> . (b) Pendulum <i>A</i> is <i>shorter than</i> pendulum <i>B</i> . (c) It is not certain whether pendulum <i>A</i> is <i>longer or shorter than</i> pendulum <i>B</i> .
	Low	Pendulum <i>A</i> swings with a longer period (time) than pendulum <i>B</i> . Both are simple pendulums. Which of the following statements comparing the pendulums is most accurate? (a) The mass of pendulum <i>A</i> is <i>larger than</i> the mass of pendulum <i>B</i> . (b) The mass of pendulum <i>A</i> is <i>smaller than</i> the mass of pendulum <i>B</i> . (c) It is not certain whether the mass of pendulum <i>A</i> is <i>larger or smaller than</i> the mass of pendulum <i>B</i> .
Projectile	High	Projectile <i>A</i> and projectile <i>B</i> start at ground level and are thrown with the same speed, but <i>A</i> is in the air for a longer time than <i>B</i> . <i>Ignore any effects of air resistance or drag</i> . (a) The launch angle (from horizontal) of projectile <i>A</i> is <i>greater than</i> that of projectile <i>B</i> . (b) The launch angle of projectile <i>A</i> is <i>less than</i> that of projectile <i>B</i> . (c) It is not certain whether the launch angle of projectile <i>A</i> is <i>greater than or less than</i> that of projectile <i>B</i> .
	Low	Projectile <i>A</i> and projectile <i>B</i> start at ground level and are thrown with the same speed, but <i>A</i> is in the air for a longer time than <i>B</i> . <i>Ignore any effects of air resistance or drag</i> . (a) The mass of projectile <i>A</i> is <i>greater than</i> that of projectile <i>B</i> . (b) The mass of projectile <i>A</i> is <i>less than</i> that of projectile <i>B</i> . (c) It is not certain whether the mass of projectile <i>A</i> is <i>greater than or less than</i> that of projectile <i>B</i> .

(Table continued)

TABLE XV. (Continued)

Physical scenario	Relative accessibility of explanation	Question
Sliding distance	Equal	Object P and object Q are both given a quick push and slide on a wooden floor with the same initial velocity. Object P comes to rest before object Q . Which of the following statements comparing the objects is most accurate? (a) The coefficient of friction between object P and the floor is <i>greater than</i> that of object Q and the floor. (b) The coefficient of friction between object P and the floor is <i>less than</i> that of object Q and the floor. (c) It is not certain whether the coefficient friction between object P and the floor is <i>greater than or less than</i> that of object Q .
	Equal	Object P and object Q are both given a quick push and slide on a wooden floor with the same initial velocity. Object P comes to rest before object Q . Which of the following statements comparing the objects is most accurate? (a) The mass of object P is <i>greater than</i> that of object Q . (b) The mass of object P is <i>less than</i> that of object Q . (c) It is not certain whether the mass of object P is <i>greater than or less than</i> that of object Q .
Gravitational potential energy	Equal	Object 1 has larger gravitational potential energy than object 2. Which of the following statements comparing the objects is most accurate? (a) Object 1 has a <i>larger mass</i> than object 2. (b) Object 1 has a <i>smaller mass</i> than object 2. (c) It is not certain whether object 1 has a <i>larger or smaller mass</i> than object 2.
	Equal	Object 1 has larger gravitational potential energy than object 2. Which of the following statements comparing the objects is most accurate? (a) Object 1 is <i>higher</i> than object 2. (b) Object 1 is <i>lower</i> than object 2. (c) It is not certain whether object 1 is <i>higher or lower</i> than object 2.

TABLE XVI. Experiment 3 multiple choice (predictive) answer format questions. Participants answered either a high or low accessibility explanation for each scenario.

Physical scenario	Relative accessibility of explanation	Question
Mass density	High	Block X and block Y are each made of a different material. The atoms in block X have a <i>larger average separation</i> than the atoms in block Y . Which of the following statements comparing the blocks is most accurate? (a) The material of block X has a <i>higher</i> mass density (kg/m^3) than the material of block Y . (b) The material of block X has a <i>lower</i> mass density (kg/m^3) than the material of block Y . (c) It is not certain whether the mass density (kg/m^3) of block X is <i>higher or lower</i> than the atoms in block Y .
	Low	Block X and block Y are each made of a different material. The atoms in block X have a <i>larger mass</i> than the atoms in block Y . Which of the following statements comparing the blocks is most accurate? (a) The material of block X has a <i>higher</i> mass density (kg/m^3) than the material of block Y . (b) The material of block X has a <i>lower</i> mass density (kg/m^3) than the material of block Y . (c) It is not certain whether the mass density (kg/m^3) of block X is <i>higher or lower</i> than the mass density of block Y .

(Table continued)

TABLE XVI. (Continued)

Physical scenario	Relative accessibility of explanation	Question
Balance scale	High	<p>A rod is balanced on a pivot. A student hangs object <i>A</i> somewhere on the left side of the rod and hangs object <i>B</i> somewhere on the right side of the rod. Object <i>B</i> on the right side has a <i>larger mass</i> than object <i>A</i> on the left side. Which of the following statements comparing the objects is most accurate?</p> <p>(a) The <i>right</i> side immediately starts to tilt down. (b) The <i>left</i> side immediately starts to tilt down. (c) It is not certain whether the <i>right side or the left side</i> starts to tilt down.</p>
	Low	<p>A rod is balanced on a pivot. A student hangs object <i>A</i> somewhere on the left side of the rod and hangs object <i>B</i> somewhere on the right side of the rod. Object <i>B</i> on the right side is <i>farther from the center pivot</i> than object <i>A</i> on the left side. Which of the following statements comparing the objects is most accurate?</p> <p>(a) The <i>right side</i> immediately starts to tilt down. (b) The <i>left side</i> immediately starts to tilt down. (c) It is not certain whether the <i>right side or the left side</i> immediately starts to tilt down.</p>
Pendulum	High	<p>Pendulum <i>A</i> is <i>longer than</i> pendulum <i>B</i>. Both are simple pendulums. Which of the following statements comparing the pendulums is most accurate?</p> <p>(a) Pendulum <i>A</i> swings with a <i>longer</i> period (time) than pendulum <i>B</i>. (b) Pendulum <i>A</i> swings with a <i>shorter</i> period (time) than pendulum <i>B</i>. (c) It is not certain whether pendulum <i>A</i> swings with a <i>longer or shorter</i> period (time) than pendulum <i>B</i>.</p>
	Low	<p>The mass of pendulum <i>A</i> is <i>larger than</i> the mass of pendulum <i>B</i>. Both are simple pendulums. Which of the following statements comparing the pendulums is most accurate?</p> <p>(a) Pendulum <i>A</i> swings with a <i>longer</i> period (time) than pendulum <i>B</i>. (b) Pendulum <i>A</i> swings with a <i>shorter</i> period (time) than pendulum <i>B</i>. (c) It is not certain whether pendulum <i>A</i> swings with a <i>longer or shorter</i> period (time) than pendulum <i>B</i>.</p>
Projectile	High	<p>Projectile <i>A</i> and projectile <i>B</i> start on level ground, are thrown with the same speed, and later land on the ground. The launch angle (from horizontal) of projectile <i>A</i> is <i>greater than</i> that of Projectile <i>B</i>. <i>Ignore any effects of air resistance or drag.</i></p> <p>(a) <i>A</i> is in the air for a <i>longer</i> time than <i>B</i>. (b) <i>A</i> is in the air for a <i>shorter</i> time than <i>B</i>. (c) It is not certain whether <i>A</i> is in the air for a <i>longer or shorter</i> time than <i>B</i>.</p>
	Low	<p>Projectile <i>A</i> and projectile <i>B</i> start on level ground, are thrown with the same speed, and later land on the ground. The mass of projectile <i>A</i> is <i>greater than</i> that of Projectile <i>B</i>. <i>Ignore any effects of air resistance or drag.</i></p> <p>(a) <i>A</i> is in the air for a <i>longer</i> time than <i>B</i>. (b) <i>A</i> is in the air for a <i>shorter</i> time than <i>B</i>. (c) It is not certain whether <i>A</i> is in the air for a <i>longer or shorter</i> time than <i>B</i>.</p>
Sliding distance	Equal	<p>Object <i>P</i> and object <i>Q</i> are both given a quick push and slide on a wooden floor with the same initial velocity. The coefficient of friction between object <i>P</i> and the floor is <i>greater than</i> that of object <i>Q</i> and the floor. Which of the following statements comparing the objects is most accurate?</p> <p>(a) Object <i>P</i> comes to rest <i>before</i> object <i>Q</i>. (b) Object <i>P</i> comes to rest <i>after</i> object <i>Q</i>. (c) It is not certain whether object <i>P</i> comes to rest <i>before or after</i> object <i>Q</i>.</p>
	Equal	<p>Object <i>P</i> and object <i>Q</i> are both given a quick push and slide on a wooden floor with the same initial velocity. The mass of object <i>P</i> is <i>greater than</i> that of object <i>Q</i>. Which of the following statements comparing the objects is most accurate?</p> <p>(a) Object <i>P</i> comes to rest <i>before</i> object <i>Q</i>. (b) Object <i>P</i> comes to rest <i>after</i> object <i>Q</i>. (c) It is not certain whether object <i>P</i> comes to rest <i>before or after</i> object <i>Q</i>.</p>

(Table continued)

TABLE XVI. (Continued)

Physical scenario	Relative accessibility of explanation	Question
Gravitational potential energy	Equal	Object 1 has a <i>larger mass</i> than object 2. Which of the following statements comparing the objects is most accurate? (a) Object 1 has <i>larger</i> gravitational potential energy than object 2. (b) Object 1 has <i>smaller</i> gravitational potential energy than object 2. (c) It is not certain whether object 1 has a <i>larger or smaller</i> gravitational potential energy than object 2.
	Equal	Object 1 is <i>higher</i> than object 2. Which of the following statements comparing the objects is most accurate? (a) Object 1 has <i>larger</i> gravitational potential energy than object 2. (b) Object 1 has <i>smaller</i> gravitational potential energy than object 2. (c) It is not certain whether object 1 has a larger or smaller gravitational potential energy than object 2.

- [1] J. R. Platt, Strong inference, *Science* **146**, 347 (1964).
- [2] A. E. Lawson, What is the role of induction and deduction in reasoning and scientific inquiry?, *J. Res. Sci. Teach.* **42**, 716 (2005).
- [3] F. Bacon, *Novum Organum* (Clarendon Press, Oxford, 1878).
- [4] J. Klayman, Varieties of confirmation bias, *Psychol. Learning Motivation* **32**, 385 (1995).
- [5] R. S. Nickerson, Confirmation bias: A ubiquitous phenomenon in many guises, *Rev. Gen. Psychol.* **2**, 175 (1998).
- [6] S. Quinn and H. Markovits, Conditional reasoning, causality, and the structure of semantic memory: Strength of association as a predictive factor for content effects, *Cognition* **68**, B93 (1998).
- [7] R. M. Byrne, O. Espino, and C. Santamaria, Counterexamples and the suppression of inferences, *J. Mem. Lang.* **40**, 347 (1999).
- [8] B. K. Hayes, G. E. Hawkins, and B. R. Newell, Consider the alternative: The effects of causal knowledge on representing and using alternative hypotheses in judgments under uncertainty, *J. Exper. Psych.: Learning, Memory, Cognit.* **42**, 723 (2016).
- [9] P. A. Klaczynski, M. J. Schuneman, and D. B. Daniel, Theories of conditional reasoning: A developmental examination of competing hypotheses, *Dev. Psychol.* **40**, 559 (2004).
- [10] A. E. Lawson, What do tests of “formal” reasoning actually measure?, *J. Res. Sci. Teach.* **29**, 965 (1992).
- [11] P. M. Fernbach, A. Darlow, and S. A. Sloman, Neglect of alternative causes in predictive but not diagnostic reasoning, *Psychol. Sci.* **21**, 329 (2010).
- [12] P. M. Fernbach, A. Darlow, and S. A. Sloman, Asymmetries in predictive and diagnostic reasoning, *J. Exp. Psychol. Gen.* **140**, 168 (2011).
- [13] M. C. Linn, C. Clement, and S. Pulos, Is it formal if it’s not physics? (The influence of content on formal reasoning), *J. Res. Sci. Teach.* **20**, 755 (1983).
- [14] A. Keselman, Supporting inquiry learning by promoting normative understanding of multivariable causality, *J. Res. Sci. Teach.* **40**, 898 (2003).
- [15] D. Kuhn, What is scientific thinking and how does it develop?, *The Wiley-Blackwell Handbook of Childhood Cognitive Development*, 2nd ed. (Wiley-Blackwell, Chichester, England, 2010), pp. 497–523.
- [16] D. Kuhn, J. Black, A. Keselman, and D. Kaplan, The development of cognitive skills to support inquiry learning, *Cognit. Instr.* **18**, 495 (2000).
- [17] V. Talanquer, Chemistry education: Ten heuristics to tame, *J. Chem. Educ.* **91**, 1091 (2014).
- [18] A. Boudreaux, P. S. Shaffer, P. R. Heron, and L. C. McDermott, Student understanding of control of variables: Deciding whether or not a variable influences the behavior of a system, *Am. J. Phys.* **76**, 163 (2008).
- [19] D. D. Cummins, Naive theories and causal deduction, *Memory Cognit.* **23**, 646 (1995).
- [20] E. T. Higgins, Knowledge activation: Accessibility, applicability, and salience, *Social Psychology: Handbook of Basic Principles*, edited by E. T. Higgins and A. W. Kruglanski (Guilford Press, New York, US, 1996).
- [21] N. Schwarz, H. Bless, M. Wänke, and P. Winkielman, Accessibility revisited, *Foundations of social cognition: A Festschrift in honor of Robert S. Wyer, Jr*, edited by G. V. Bodenhausen and A. J. Lambert (Lawrence Erlbaum, Mahwah, New Jersey, US, 2003), pp. 51–77.
- [22] D. Kahneman, A perspective on judgment and choice: mapping bounded rationality, *Am. Psychol.* **58**, 697 (2003).
- [23] J. S. B. Evans, Dual-processing accounts of reasoning, judgment, and social cognition, *Annu. Rev. Psychol.* **59**, 255 (2008).
- [24] J. St. B. T. Evans, D. E. Over, and S. J. Handley, A theory of hypothetical thinking, in *Thinking: Psychological*

- Perspectives on Reasoning, Judgement and Decision Making*, edited by D. Hardman and L. Maachi (Wiley, Chichester, England, 2003), pp. 3–22.
- [25] V. Talanquer, Commonsense chemistry: a model for understanding students' alternative conceptions, *J. Chem. Educ.* **83**, 811 (2006).
- [26] A. F. Heckler, The Ubiquitous Patterns of Incorrect Answers to Science Questions: The Role of Automatic, Bottom-up Processes, *Psychol. Learn. Motivation, Adv. Res. Theory* **55**, 227 (2011).
- [27] A. F. Heckler and T. M. Scaife, Patterns of response times and response choices to science questions: The influence of relative processing time, *Cogn. Sci.* **39**, 496 (2015).
- [28] M. Kryjevskaja, M. R. Stetzer, and N. Grosz, Answer first: Applying the heuristic-analytic theory of reasoning to examine student intuitive thinking in the context of physics, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020109 (2014).
- [29] P. A. Kłaczynski and D. B. Daniel, Individual differences in conditional reasoning: A dual-process account, *Thinking Reasoning* **11**, 305 (2005).
- [30] D. Kuhn and S. Franklin, *The Second Decade: What Develops (and How)* (John Wiley & Sons Inc., New York, 2006).
- [31] J. C. Speirs, W. N. Ferm, M. R. Stetzer, and B. A. Lindsey, Probing Student Ability to Construct Reasoning Chains: A New Methodology, *Proceedings of the Physics Education Research Conference 2016, Sacramento, CA*, edited by D. L. Jones, L. Ding, and A. Traxler (AIP, New York, 2016).
- [32] R. Rosenblatt, A. F. Heckler, and K. Flores, A Tutorial Design Process Applied to an Introductory Materials Engineering Course, *Adv. Engin. Educ.* **3**, n3 (2013).