Investigating the impact of problem properties on introductory and advanced student responses to introductory thermodynamics conceptual problems

Mary Jane Brundage,[1] David E. Meltzer,[2] and Chandralekha Singh[1]

[1]*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pittsburgh 15260, USA*
[2]*College of Integrative Sciences and Arts, Arizona State University, Mesa, Arizona 85212, USA*

(Received 28 August 2023; accepted 10 February 2025; published 28 March 2025)

We use a validated conceptual multiple-choice survey instrument focusing on thermodynamic processes and the first and second laws of thermodynamics at the level of introductory physics to investigate the problem-property dependence of introductory and advanced student responses to introductory thermodynamics problems after traditional lecture-based instruction. The survey instrument has qualitative problems involving the same concepts across multiple problems related to internal energy, work, heat transfer, and entropy. The concepts for which we investigated problem-property dependence include, among others, (i) internal energy and entropy are state variables while work and heat transfer are path-dependent variables, (ii) internal energy is proportional to the absolute temperature for an ideal gas, (iii) work corresponds to the (signed) area under the curve on a PV (pressure-volume) diagram, and (iv) internal energy is constant but entropy increases in isolated systems undergoing spontaneous and irreversible processes. This study used survey data from over 1000 college students in introductory-level algebra-based and calculus-based physics courses as well as upper-level thermodynamics courses; the survey was administered after traditional instruction in relevant concepts for each group. Think-aloud interviews were carried out to gain additional insight into students' thinking as they responded to the survey problems. For concepts related to internal energy, heat transfer, and work, student responses for different concepts investigated often showed strong problem-property dependence, but advanced students, as a group, generally performed better than introductory students across different problems. For entropy concepts, introductory students consistently performed poorly across problem types, reflecting a persistent belief in the constancy of entropy. By contrast, upper-level students struggled consistently in cases where entropy was *not* increasing (e.g., net entropy change in a cyclic process). Our systematic investigation of problem-property dependence of student responses is novel for thermodynamics, made possible by the use of a survey that includes multiple varied problem scenarios involving the same underlying concept. In addition to yielding many previously unreported results regarding problem-property dependence in thermodynamics, our work confirms and extends, to new problem settings and student groups, findings previously reported through the study of more limited problem settings and student groups.

DOI: 10.1103/PhysRevPhysEducRes.21.010127

I. INTRODUCTION AND GOALS

A. Overview

Physics is a discipline that is devoted to explaining diverse physical phenomena using just a few basic physics principles. To learn physics effectively and to develop expertise, it is essential to unpack the underlying meaning of the abstract principles and concepts and recognize their applicability in diverse situations [1–4]. Research shows that identifying and applying relevant physics principles and concepts to a broad array of diverse problems is an important hallmark of expertise in physics. Physics courses focus on helping students learn to discern the deep similarities between problems that share the same underlying physics principle, but which have different surface features, employ different representations, or utilize different physical scenarios.

Transfer of knowledge refers to the application of knowledge and skills learned in a particular setting to other settings that might differ substantially [5–9]. This transfer could include the learning of a concept by using a particular problem type (such as learning Newton's second law using a graphical representation) and applying it correctly to a completely different problem type (such as one using a mathematical/verbal representation). However, challenges of transfer may be revealed even when problems using the same representation (for example, graphical

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.

problems) differ only in the quantities being represented (such as whether a position-time or a velocity-time graph is involved). More broadly, we may say that distinctions among diverse problems that may influence students' performance involve different problem "properties." Dual-process theories of reasoning (discussed further below) offer one approach to understanding how different modes of cognitive processing (primarily, fast and "intuitive" vs slow and "analytical") can underlie some aspects of problem-property dependence of students' performance on physics problems.

In this study, we define a "concept" as the physics principle needed to solve a problem. For example, a student may need to use concepts such as the first law of thermodynamics or the pressure or temperature dependence in the ideal gas law. The "properties" of a problem include the specific scenario or physical setting of that problem and all of its specific characteristics. For example, two similar —yet distinct—problems may involve two isothermal processes on a PV (pressure-volume) diagram that differ only in whether the volume increases or decreases. Alternatively, one problem might involve an *isothermal* expansion and the other an *adiabatic* expansion. The difference might only lie in the specific *task* presented to the student, e.g., calculate work done or calculate heat transferred. In the next paragraph, we provide a more complete definition of this terminology.

Individual physics problems, even those intended to assess thinking on a single, very specific physics concept, may differ from each other in many ways. In thermodynamics, those differences may include the physical setting or scenario of the problem, the nature of the process or processes involved (e.g., isothermal or adiabatic), whether or not the problem includes a PV diagram, the specific design features of that PV diagram, whether parameters such as temperature, pressure, and volume are increasing or decreasing, the specific thermodynamic quantities that students are asked to calculate, and innumerable other characteristics. In this paper, we use the term properties to refer to the *complete set* of characteristics belonging to a specific problem, including its physical setting and scenario. (An example of a physical setting is a container of ideal gas in thermal contact with a thermal reservoir. An example of a scenario is the compression of that ideal gas by a frictionless piston.) So, for example, we use three ideal-gas problems to assess understanding of the proportionality between temperature and internal energy: one involves an isochoric process in which the temperature increases, the second an isobaric compression, and a third problem includes both an adiabatic compression and an isochoric process in which the temperature increases. We refer to these three problems as having different properties. To the expert, the characteristics distinguishing these problems from each other might all be called "surface features" since the very same physics concept may readily be applied to solve all of them. In cases where such surface features draw immediate attention and cue unproductive lines of reasoning, they may be termed "salient distracting features." (We discuss this further below.) It is important to emphasize that there are no hard and fast dividing lines distinguishing "features," "tasks," "settings," "scenarios," "characteristics," or "properties," but for convenience, we use *properties* here as the most general term that is intended to encompass all the others.

Past research has shown that experts and novices tend to classify physics problems differently [10,11]. Problems that novice students, who are developing expertise in physics, consider as similar due to surface features may involve very different physics principles and concepts, requiring quite different approaches to solve correctly. By contrast, experts may group problems together because the same physics principle or concept can be used to solve them. For example, a study on the categorization of introductory mechanics problems [10] based upon similarity of solutions indicates that, while experts are likely to place two problems in different categories because the best or most straightforward solutions involve different physics principles (e.g., energy conservation and Newton's second law), novice students may group the same two problems together because—they say—they both involve a pulley with hanging blocks. The findings suggest that novices are more likely to be distracted by surface features of problems, often leading them to choose unproductive solution pathways. A notorious example of this is Newton's third law: students may learn that two objects of different mass in contact with each other exert equal-magnitude forces on each other, but if friction is added to the problem or if the objects are changing speed, students often lose sight of the key governing principle.

Helping students recognize the applicability of the physics principles and concepts they have learned and use them correctly in different physical settings is an important goal of physics education for science and engineering students. For that reason, many research studies have investigated students' ability to apply their knowledge to diverse problems that differ from each other in some of their specific properties [12–15]. Cognitive theory suggests that transfer can be difficult, especially if the source (from which the concept was learned) and the target (a problem at hand to which the concept is to be applied) do not share similar surface features, that is, the problems have different properties. The source may incorporate, for example, a specific physical scenario such as the free expansion of an ideal gas, while the target may involve a very different scenario (such as heat transfer between two objects at different temperatures). The failure to appropriately transfer knowledge can often be attributed to the fact that the knowledge may be encoded in long-term memory (LTM) with the setting in which it was learned, and the features of the target problem may not lead to accessing relevant resources in LTM even though the two problems share deep features [1,16].

Research shows that both the robustness of the students' knowledge structure and the specific setting in which the student originally acquired the knowledge can affect a student's ability to apply knowledge flexibly across different problem types [10–13,17–19]. For this reason, prior studies have used various scaffolding mechanisms to assist students in learning to transfer their knowledge effectively [20,21]. For example, using isomorphic (similar structure) problems and analogical reasoning can help students link problems that involve the same underlying principles and concepts [20–24]. Isomorphic problems have the same underlying principles or concepts but have different surface features; students may not easily recognize the relationship between them due to their different features. To improve transfer of knowledge, it is important to help students contemplate the applicability of the same physics principles and concepts in different physical settings and learn to decontextualize knowledge and store it in their LTM at a more abstract level [1].

One test of the robustness of students' knowledge and their ability to access and apply it in different settings is to present them with a variety of problems that are all focused on the same physics concept and yet differ in the physical setting or scenario and/or other characteristic properties in which they are set. If students are successful in solving all of the problems regardless of their differing characteristic properties, a plausible inference might be that instruction on that topic was relatively successful. By contrast, if students are successful with some problem settings and not with others, one might infer that instruction was insufficient to enable students to overlook superficial differences when attempting to access and apply concepts that they have studied. In the latter situation, an understanding of the *specific* problem properties that led to student difficulties and the *specific* reasons for those difficulties can be extremely helpful both to instructors and to curriculum developers who seek to improve the effectiveness of instruction on these topics.

B. Impact of diverse problem properties on student performance in thermodynamics

Here we discuss an investigation of how consistently and correctly introductory and upper-level students access and apply their knowledge about thermodynamic concepts to tasks incorporating diverse problem properties. We employed a validated 78-item conceptual multiple-choice survey administered to students after traditional lecture-based instruction in relevant content. The full set of 13 concepts on which we focused can be found in Table I; some examples are (i) internal energy (E) and entropy (S) are state variables; (ii) work (W) and heat transfer (Q) are path-dependent variables; and (iii) internal energy is proportional to the absolute temperature (T) for an ideal gas.

The reason we chose to investigate performance on problems involving these particular concepts is that they are important learning goals in typical introductory physics courses and they each have a fairly narrow focus. That is, they can be clearly and succinctly stated, and problems can be posed (as in the survey instrument used) where the solution depends rather narrowly on one concept alone, without the simultaneous involvement of other concepts. By studying student responses to problems with diverse properties, we can gain a better assessment of students' expertise regarding those concepts after traditional lecture-based instruction. Multiple survey problems framed within

TABLE I. List of 13 concepts relating to thermodynamic variables on which we focus in this study. The table also lists the survey items related to each of the 13 concepts.

Item No.	Concepts
33, 34, 69	1. E is proportional to T for an ideal gas
13, 23, 50, 72	2. $\Delta E = 0$ for an isolated system undergoing a spontaneous process
6, 25, 47	3. E is a state variable (also implying $\Delta E = 0$ for a system undergoing a cyclic process)
61, 65	4. $\Delta E = 0$ for an isothermal expansion
2, 33, 34, 44, 45	5. The sign of ΔE for an ideal gas undergoing isochoric, isobaric, and adiabatic processes is determined by whether pressure and/or volume are increasing
22, 49, 71	6. $W = 0$ when there is no net system expansion against a movable boundary wall
7, 10, 42, 43, 57, 58, 59	7. W may be interpreted as the (signed) area under the curve in a PV diagram
3, 62	8. W is positive for an expansion whether adiabatic or isothermal
4, 32, 60	9. In a reversible isothermal process, $Q \neq 0$ and the sign of Q is determined by whether volume is increasing or decreasing
9, 12, 26, 30	10. Q is a path-dependent variable
17, 21, 53, 67, 75	11. Entropy of the universe increases in a spontaneous/irreversible process
8, 24	12. $\Delta S = 0$ for a system undergoing a cyclic process
5, 63	13. The sign of ΔS for a system undergoing a reversible isothermal process is determined by whether the volume is increasing or decreasing

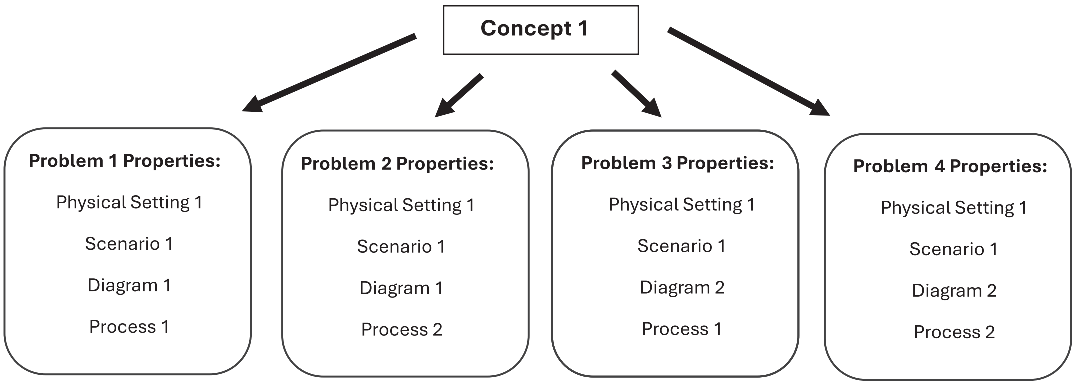

FIG. 1. For one specific concept, student performance across different problems is compared. In this illustrative model, two different diagrams and two different thermodynamic processes are used in various combinations while all other problem properties remain unchanged.

different scenarios or differing in type or surface features were thus used to investigate student thinking regarding each targeted concept, as represented schematically in Fig. 1. (We emphasize that this schematic is purely illustrative and does not exactly model any of our actual problem sets.) We note that if students in a particular group (e.g., those enrolled in calculus-based introductory courses) consistently perform well on problems with diverse properties involving a particular thermodynamic concept, e.g., that entropy is a state function, it could indicate a high level of expertise regarding that concept within that population. On the other hand, if students in a particular group *consistently* perform poorly on problems regarding a specific concept even when the problems have distinctly different properties, it could indicate that they have a weak ability to invoke and apply that concept in practice. If they perform poorly with some problem properties but well in others for the *same* concept, it may be a sign of the presence of alternative conceptions and/or distracting surface features in cases in which they perform poorly.

It is important to note that no prior educational research related to E, W, Q, and S concepts has investigated the dependence of student responses on problem properties in a systematic manner. Prior studies have mainly focused on student performance on a single or perhaps two different problems targeting the same concept involving the same thermodynamic variable. By contrast, our survey instrument has many problems involving the same concept posed within three, four, or more problems differing in type, representation, setting, task, or other properties.

As a specific example, one concept that instructors want their students to learn is that during all spontaneous and irreversible processes, the entropy of the universe *increases*. In the STPFaSL-Long, this concept is applied in five different problem settings: (i) heat transfer between two solids in contact; (ii) heat transfer between two gases in contact; (iii) free expansion of a gas; (iv) mixing of two noninteracting gases; and (v) an irreversible isochoric process. The answer options (increases, decreases, remains the same, not enough information) are similar or identical to each other from problem to problem and do not embed within them any "hidden" alternative conceptions. This allows us to compare the relative difficulty of different problem settings for the same concept, that is, which were "easier" and which were "harder" for the students to answer. Moreover, we use interview data to gain insight into some of the specific reasons that students found these problem settings to be easier or harder. Information regarding particularly difficult problem settings and the reasons students find them difficult (or easy) can be valuable for instructors, curriculum developers, and researchers.

Knowledge of the dependence of student performance on specific problem properties can be extremely valuable for planning effective instruction and for developing well-targeted curricular materials. And yet, as valuable as it may be, it is rarely the target of systematic investigation precisely because it is challenging and time consuming to study, requiring as it does the use of a large set of problems having diverse properties yet focused on the same concepts. An example of a potentially important outcome of such an investigation is the finding of a problem setting or scenario that offers particular challenges to a particular student group—even if the underlying concept was successfully applied in other settings. This type of finding is a high-confidence indicator that additional instructional focus and research-based curricula may be required to address these special challenges to students' thinking. Also, if a particular student group consistently performs poorly across different problem types targeting a particular concept, it may indicate that the concept is particularly challenging for that student group. It may be that there are alternative conceptions and/or failures to integrate and apply key ideas that are impacting student performance *independent* of the problem type. We also note that since all upper-level students were once introductory students, the consistency with which they solve introductory problems involving a particular concept across various problem types can be an indicator of concepts that pose particularly persistent difficulties for student learners at all levels. Combined with information about solution methods and reasoning pathways typically employed by introductory students after traditional instruction, these data obtained from upper-level students can provide insights into the learning process that are potentially helpful to instructors and researchers.

Investigations of student difficulties do not generally focus on the dependence of student responses on specific problem properties nor on whether or to what degree students are consistently invoking and applying their knowledge across many diverse problem types. This limited focus—understandable from a practical researcher's standpoint—can potentially lead to significant gaps in our understanding of students' thinking. For example, a common difficulty previously reported by researchers is the student idea that temperature and internal energy of the system do not change in an adiabatic process. When investigating with a specific focus on problem properties, we compare student performance on a group of problems that all focus on change in internal energy. However, we use one problem that deals solely with an adiabatic process and another that includes both adiabatic and isochoric processes. This broader focus allowed us to show that students' thinking regarding changes in internal energy in adiabatic processes was not consistent and was in fact distinctly dependent on problem properties, offering a fresh perspective on this aspect of students' reasoning along with clues as to possible productive directions for further research on this topic (see discussion in Sec. III A).

Although prior research suggests that both introductory and upper-level students face many difficulties with introductory thermodynamics concepts, e.g., see Refs. [25–69], the focus of this paper is not on student difficulties. (Detailed analyses of student difficulties on problems in the STPFaSL-Long are provided in two other papers; see Refs. [65,68].) Our focus here is specifically on the dependence of students' problem responses on problem *properties*. We explore both the presence and the nature of problem properties that impact students' problem solving, either positively or negatively. That is, the effects of this impact may either interfere with or aid students' ability to solve thermodynamics problems. For example, other research has shown that apparent inconsistencies in students' ability to apply ideas across problems with varied problem properties may sometimes be explained by the presence of salient distracting features. Salient distracting features (SDFs) are features (or properties) of a task that draw immediate attention away from other task features, are processed easily and rapidly, and cue unproductive lines of reasoning [67]. In addition to SDFs, we explore problem features that are *not* distracting and, in fact, seem to cue potentially productive reasoning pathways. (These might be called "salient productive features" if they draw immediate attention, although that terminology is not apparently in use.) Alternatively, some problem properties may draw immediate attention and yet cue lengthy reasoning chains that *sometimes* turn out to be productive; we examine these cases as well. The framework of dual-process theories of reasoning may offer additional insight into this process by illuminating differences between rapid and intuitive reasoning pathways on the one hand, and slow, deliberate, and analytical reasoning pathways on the other, thus shedding light on how the use of particular pathways cued by various problem features may influence students' ability to choose correct answers with consistency [66,67]. Although we do not explore those specific issues here, they do offer a potentially important direction for future research in this area.

C. Research questions

As noted above, there are many factors that can affect students' ability to access and apply relevant knowledge consistently across different problem types for the same concept. These factors include (but are not limited to) the setting in which the concept was learned, how similar the surface features of the different problems are to those of the problems they have solved in the past, and whether certain properties of the problems involving different settings act as distractors. Our research questions are as follows:

RQ1: For each thermodynamics concept that we investigated (see Table I), to what extent are the responses of introductory and upper-level student dependent on the specific properties of the problems?

RQ2: How do the problem-property dependences of introductory and upper-level student performance on conceptual problems involving thermodynamic variables differ from each other?

RQ3: In cases where problem-property dependence exists, what specific problem properties can be identified as either "distracting" or "potentially productive" for one or more student groups, and what is the nature of their influence?

To put these research questions into perspective, we point out that by including many target concepts previously explored in research, we are in effect performing a validation and consistency check on many previously published findings. By investigating whether, to what degree, and how exactly different student groups' responses depend on the specific setting, diagram features, task, and other properties of ordinary test problems, we are able to confirm, disconfirm, or add nuance to previous findings and generalize them—or restrict them—to specific student groups (in particular, upper level, calculus-based, or algebra-based physics students). That is, we are not merely searching for the existence of problem-property dependences. Rather, we are investigating the specific nature of those dependences in detail and applying our results both to broaden perspective on previously reported findings and to reveal new, previously unreported aspects of students' thinking in thermodynamics.

II. ANALYTIC FRAMEWORK AND METHODOLOGY

A. Analytic framework

Our approach in this paper is basically empirical, while at the same time acknowledging consistency with and relevance to both transfer theory and dual-process theories of reasoning. Figure 2 shows the analytic framework we used to investigate the problem-property dependence of students' performance on problems related to each individual thermodynamic concept. First, we analyzed the correct-response rates on all multiple-choice problems related to a given concept. For survey items with four answer choices, purely random guessing would correspond to 25% of students answering correctly; very few of our survey problems had correct-response rates close to that random-guessing range. However, there are distinctions between purely random guessing based on complete ignorance of a topic—which might lead to a 25% correct-response rate—and so-called "educated guessing" based on partial or imperfect knowledge, which would be expected to yield significantly higher rates of correct responses that are nonetheless well below 100%. (Calling this practice educated "guessing" seems somewhat pejorative, and a better term might be "educated reasoning." In any case we are merely referring here to the typical mental process students undertake to solve challenging problems.) However, educated guessing—in contrast to firm understanding—is relatively unreliable and inconsistent, often leading in practice to good performance on some problems and weaker performance on related problems that may have different settings, scenarios, tasks, diagrams, or other properties. For that reason, significant differences in correct-response rates on closely related problems are a good indicator of "fragile" student conceptual understanding that is associated with attempts to solve problems with imperfect knowledge.

In cases where over 75% of students gave correct responses on *all* problems within a concept group, we concluded that most students had conceptual understanding that was adequate to be applied accurately to a diversity of problems. If correct-response rates on problems within a concept group did not differ significantly from each other, even if they were below 75%, we concluded that students'

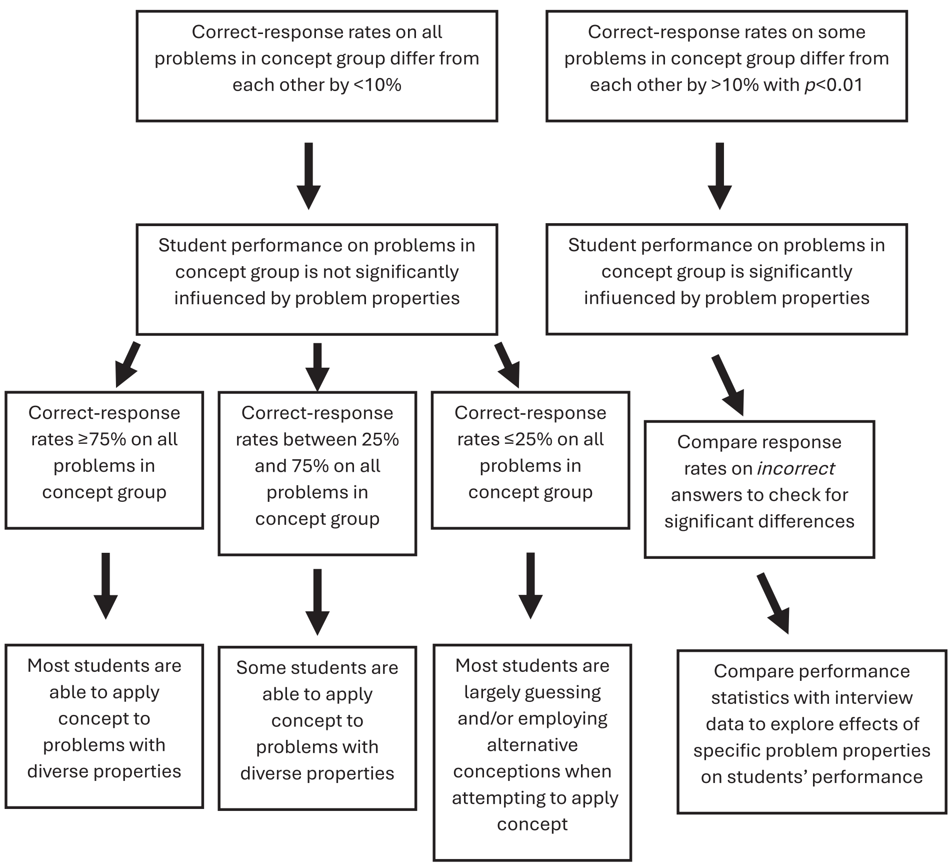

FIG. 2. Analytic framework used to determine whether and why a concept shows problem-property dependence in student responses across different problems.

performance was not substantially influenced by problem properties. If, for a given concept, there was a significant performance difference across different problems (on problems with the same number of answer options), we concluded that students' performance was probably influenced by problem properties.

In a sense, the goal of this paper is to uncover and elucidate some of the mental dynamics associated with the type of educated reasoning process described above. We uncover its presence by analyzing differences in correct-response rates ("performances") on problems that are focused on the same concept and which—to the expert—appear essentially identical. In those cases, we endeavor to identify the specific problem properties that either aid or detract from students' performance. One of the tools we use for this is to check response rates on *incorrect* answer options to see if there are any significant differences. We also carefully compare performance statistics with interview data to see if students' explanations help account for performance differences. (We note that even in cases where the correct-response rates for different problems are the same or only marginally different, there may still be problem-property dependence present that is reflected in varied response rates on incorrect-answer options. We explicitly address this possibility in the Results section in cases where it was relevant.)

A more detailed analysis of the dynamics of these reasoning processes is provided within the framework of dual-process theories of reasoning, as referenced above. Certain problem properties may cue rapid, intuitive thinking processes that can lead to erroneous, yet persuasive conclusions. Other properties may aid students in carrying out the slower, more deliberate, and generally more accurate reasoning processes that can test or replace conclusions arrived at through the rapid processes. However, we also found in certain cases that specific problem properties actually seemed to cue a slower and more deliberative reasoning process that was *consistently* inaccurate and unproductive; we take special note of these cases in the Results and Summary. We must leave it to future investigations to analyze in more detail the dynamics of the influential problem properties identified in the Results section, since our own data are not fully adequate for that task.

For a specific concept, e.g., E is proportional to T for an ideal gas, students in our survey had to invoke and apply their knowledge to problems with differing properties involving isothermal, isochoric, adiabatic, and other processes. Comparison of how well and how consistently students performed on problems involving the same concept that nonetheless differed in their specific properties provided insight into the level of student expertise and the possible presence or absence of salient distracting features. The problem types we employ to probe student thinking on specific concepts are diverse, yet not exceedingly broad. They do not involve just *any* problem related to, e.g., internal energy E, but just a small group of relatively simple processes (commonly encountered in introductory thermodynamics) for which the answers—if one understands and can apply the concept—should be relatively easy to determine. The focus on performance across diverse problems provides valuable perspectives that complement the focus on student difficulties.

We applied the analytic framework, shown in Fig. 2, to the 13 concepts selected for this investigation. The approach for selecting the concepts involved the following:

- Must be important concepts typically discussed in introductory physics courses and related to specific thermodynamic variables.
- Must be directly relevant to more than one problem on the survey instrument.
- Must be very narrowly defined, such that a simple one-sentence statement (such as those in Table I) can encompass the relevant concept that is—in principle—necessary and sufficient to solve a particular set of problems.

The specific choice of concepts, aside from the criteria described above, relies on the typical learning goals of many experienced instructors instead of, e.g., classical test theory or factor analysis. (A future factor analysis might indeed be of interest but is not part of the present investigation.) As outlined in our previous papers [31,64,70], we have engaged in prolonged and extended discussions with thermodynamics instructors to determine the optimum choice of appropriate survey topics, and it is the results of those discussions that led to the final version of the survey instrument employed in this investigation.

Our discussion of results involving thermodynamic concepts related to various thermodynamic variables are grouped together by variable; that is, there is an "internal energy" group, an "entropy" group, and so on. This organization facilitates the comparison of problem-property dependence for different concepts related to that variable. The variables themselves are discussed in sequence. We first examine students' responses to problems involving various concepts related to internal energy, each concept treated one by one under the "Energy" heading. We then move on to the variable of work, then heat, and then entropy. As noted, in this investigation, the concepts targeted included the state-variable nature of internal energy and entropy and the path-dependent nature of work and heat transfer, as well as several others; all targeted concepts are enumerated in Table I.

B. STPFaSL-Long instrument and data collection

The Survey of Thermodynamic Processes and First and Second Laws-Long (STPFaSL-Long), a validated survey instrument with 78 problems, was used in this research. This instrument focuses on introductory thermodynamics concepts, e.g., the first and second laws of thermodynamics applied to thermodynamic variables (such as internal energy, work, heat transfer, and entropy) in the setting of different thermodynamic processes (e.g., isothermal, isochoric, isobaric, and adiabatic), as well as on spontaneous and irreversible processes such as free expansion and mixing of gases. The details of the development and validation of the STPFaSL-Long survey instrument can be found in Ref. [70] and the survey itself can be found in the supplemental materials of Ref. [70] and on PhysPort [71]. Most problems on the survey have four possible answer choices, including "not enough information." Problems that deal with entropy and internal energy typically include answer options of *increases*, *decreases*, and *remains the same*, while those dealing with work include instead *positive*, *negative*, and *zero*. Out of 78 total items, 22 are true or false (T/F) questions. We note that there are a total of 19 problem scenarios across the 78 problems, and these 19 scenarios have wording identical to those used in our STPFaSL-Short survey instrument, which contains 33 problems; see Ref. [31]. Students are able to complete either version of the survey (short or long) in one 50-min class period. The difference between the long and short versions is only in the multiple-choice questions. In the long version used here, each problem asks only about one thermodynamic variable at a time (e.g., internal energy, work, or entropy) while in the short version, each problem asks about more than one variable at a time, making it difficult to disentangle and understand student thinking about each thermodynamic variable separately. Some of the problems in the short version also have common alternative conceptions embedded in the answer options provided, unlike the long version. (For example, one "Short" answer option is "W is equal for both processes because the final and initial volumes are the same.")

This investigation of the problem-property dependence of student responses uses data from surveys administered after traditional lecture-based instruction (as post-tests). The written data analyzed here were taken by administering the survey in proctored in-person classes as a post-test after students had studied the relevant concepts but before their final exam in the course. The data are from ten different in-person courses from four different large public institutions. (Two additional classes provided pretest-only data, not discussed in this paper.) All students completed the STPFaSL-Long survey in class on Scantrons in a 50-min class period if they took the full survey. For validation purposes, the survey was split into two parts, and some instructors of the introductory courses administered only one part of the split survey to their students, using only half of the class period. There were 48 questions in one part and 52 questions in the other part; 22 items were common between the two parts. (These 22 common items were the last 22 questions in the part with 48 questions and the first 22 questions in the part with 52 questions.) As discussed in the validation paper [70], the performances of students who were administered the two versions of the common 22 questions were similar (e.g., 58.5% vs 58.8% for calculus based), providing further evidence of the reliability of the survey. Also, the performances of these students on 48 or 52 questions were 57.6% and 55.0%, respectively, showing that the first and second halves of the survey are relatively comparable in difficulty.

Class instruction on relevant topics lasted between 2 and 4 weeks. (We note that prior research has shown that greater amounts of time to cover the material in class using traditional instruction have not translated into better student performance [31], and so minor differences in time-on-topic in the different courses are not considered significant.) Students were given some extra credit for completing the survey. We discuss the analysis of problem-property dependence in the written data from three groups of students: 550 students in the introductory algebra-based (Int-alg) physics course, 492 students in the introductory calculus-based (Int-calc) physics course, and 89 students in the upper-level thermodynamics course. (Due to the use of split halves of the survey, discussed above, not all introductory students answered each survey item; Table II in the Supplemental Material provides the actual sample size for each survey item.) Students in the calculus-based courses typically were engineering majors with some physics, chemistry, and math majors, while students in the algebra-based courses were mainly biological science majors and/or those interested in health-related professions. Students included in the upper-level group were typically physics majors in thermodynamics courses or Ph.D. students in the first semester of their graduate program who had *not* taken any graduate-level thermodynamics course. Since the survey was administered as a pretest to this latter group of graduate students, they were presumed to have taken upper-level undergraduate thermodynamics. When we separately calculated the means and standard deviations of the upper-division students (44 students), and graduate students (45 students), we found that the mean correct response percentages were 72.1% and 78.8%, respectively, and their standard deviations were 13.3% and 13.6%, respectively. Thus, these two groups' performances were very similar, and we combined them into one group of 89 students, represented below as the post-test results of students in "upper-level" thermodynamics.

The interview data are from 11 introductory algebra-based students and 6 upper-level students from one institution who volunteered after an opportunity to participate in this study was announced. Each interview lasted between 1 and 2 h in one sitting depending upon the students' pace. The interviewed students were given $25 for their participation. The interviews used a semistructured think-aloud protocol. Students were asked to "think aloud" as they answered the questions and were not disturbed except for being urged to keep talking if they became quiet.

Only at the end of the interview were they asked for clarifications of points they had not made clear.

C. Statistical measures

To determine whether student performance on problems with different properties was similar or different, statistical testing was done. We compare results for the same survey item across different student groups (e.g., item-33 response rates for the upper-level students and for the introductory algebra-based physics students) as well as results for different survey items within the same student group (e.g., response rates on survey items 33 and 34 for upper-level students). As with all educational research, it is important to distinguish between what is statistically significant and what is pedagogically significant. With large datasets, it is common for performance differences to be statistically significant without being meaningfully different in a pedagogical sense. This is especially true with our data. To avoid this potentially misleading outcome, we offer measures of both statistical and pedagogical significance. We define a pedagogically significant difference in performance as one in which correct-response rates differ by at least 10%, roughly corresponding to one letter grade. We determine statistical significance through the use of both Cohen's h and two-sample z-tests for binomial proportions. In order to call a difference "significant," we require both a minimum 10% difference in response rates *and* specific values on the statistical tests, as discussed below.

To measure effect size, we use Cohen's h, which is a measure of the difference between proportions [72]. We compare (in most cases) the proportions of correct responses on two different individual survey items that bear on the same concept but incorporate different problem properties. Cohen's h does not incorporate sample size in the calculation; instead, it provides a measure of the "nonoverlap" of two data samples consisting of sets of binomial (correct/incorrect) responses, where in our case, the samples are sets of responses on two different problems. (We do pairwise comparisons of all problems within a single problem group, all of which focus on one specific concept.) A larger effect size (larger value of h) represents a larger difference in response rates and thus a larger nonoverlap in the datasets. Below, we show the way in which Cohen's effect-size criterion is generally consistent with a 10% threshold for pedagogical significance.

(We note that an effect size of 0.5 corresponds approximately to a "non-overlap" of one-third of the combined population of the two datasets (Table 2.2.1 in Ref [72]). A loose interpretation in the current context would be that if the difference in response rates of two problems corresponds to an effect size of 0.5, approximately one-third of all students respond to one of the problems in a manner that is incompatible with how they respond to the other problem.)

Cohen's h values range between 0 and 3.14, and Cohen has (somewhat arbitrarily) defined values of h that correspond to "small" ($h = 0.2$), "medium" ($h = 0.5$), and "large" ($h = 0.8$) effect sizes [72]. When calculating Cohen's h for a correct-response rate difference of 10%, h can take on values in the range $0.20 < h < 0.64$, meaning a small-to-medium effect size. (The specific value of h corresponding to a 10% difference in response rates will depend on the absolute values—between 0% and 100%—of those rates themselves.) As another example, when comparing survey items with a 15% difference in response rate, Cohen's h can take on values in the range $0.30 < h < 0.80$, meaning a small-to-large effect size.

We also use z tests of binomial proportions to calculate the p values for differences in response rates; these values do depend on sample sizes. In cases where item response rates differed by 10% or more, we usually found $p < 0.05$, also corresponding (as mentioned above) to $h > 0.2$. A p value of 0.05 or smaller is often considered an indication of a statistically significant difference between response rates. However, a stricter criterion, appropriate when making many comparisons as we do here, would require $p < 0.01$ or even $p < 0.001$ to be considered statistically significant. *For that reason, we require both a minimum response-rate difference of 10% and $p < 0.01$ on the z test to call a difference "significant."* In fact, a substantial majority of the differences we cite reflect $p < 0.001$ and these are indicated explicitly in the text. This criterion is used in place of a Bonferroni correction, which we consider excessively strict in this case in view of our application of two separate significance criteria. We also note that the smaller sample size of the upper-level group makes it less likely that 10% response-rate differences in or with that group would meet the strictest significance criterion. In certain cases of interest where response-rate differences are around 10% and $0.01 < p < 0.05$, we modify the language and specify the precise p values.

To summarize, whenever we compare the performance of students, a 10% difference in correct-response rates corresponds to—at least—a "small" effect size. We generally restrict our discussion of differences in student performance to items (or groups) for which rates differ by at least 10% with $p < 0.01$. (In cases where all response rates in a concept group differ by $< 10\%$, we state that problem-property dependence is weak or nonexistent.) We provide a comprehensive set of Cohen's h and p values in the Supplemental Material for all pairwise rate comparisons under each particular concept heading, that is, when comparing the same survey item across different student groups and when comparing different survey items within the same student group.

III. RESULTS AND DISCUSSION

Problem-property dependence of introductory (calculus-based and algebra-based) and upper-level student responses was investigated for problems involving various concepts related to the change in internal energy of the system (ΔE), heat transfer (Q), work done by a gas (W), and change in entropy (ΔS). As noted in Sec. I, one of our research questions focuses on the extent to which introductory and advanced student responses are dependent on problem properties for each specific concept enumerated in Table I. Our second research question focuses on the extent to which the problem-property dependences of introductory and advanced student responses differ from each other for each of the listed concepts, while our third question addresses the specific nature and mechanism of those dependences. We will discuss the results pertaining to all research questions for each concept below. For convenience, we will refer to the three student groups as Upper-level, Int-calc, and Int-alg.

Each of the histograms below (with one exception that is noted) shows the percentages of correct responses given by upper-level, introductory calculus-based, and algebra-based physics students on various problems with different properties, all related to a single concept. (Complete results, including response rates for all answer options on all problems, can be found in Ref. [70].) We note that only if there is a difference of at least 10% with $p < 0.01$ between two or more problems within a given concept group, or on the same problem between two different student groups, will we consider there to be a significant problem-property dependence. (For example, between problems 33 and 69 for the Int-calc group or on problem 69 between the upper-level and Int-alg group.) Standard errors are represented by error bars on the histograms. We note that for each concept on which we focus, we will only describe student difficulties briefly and only as they are relevant for shedding light on the problem-property dependence of students' thinking. Extensive details regarding student difficulties found using the STPFaSL-Long survey are discussed elsewhere [65,68].

A. Problem-property dependence of student responses for concepts related to ΔE

Problems related to internal energy for which we investigate problem-property dependence are divided into four groups, each focusing on one distinct concept. The four concepts are (i) internal energy E is proportional to temperature T for an ideal monatomic gas; (ii) $\Delta E = 0$ for an isolated system undergoing a spontaneous and irreversible process; (iii) E is a state variable (also implying $\Delta E = 0$ for a system undergoing a cyclic process); (iv) $\Delta E = 0$ for an isothermal expansion; and (v) the sign of ΔE for an ideal gas undergoing isochoric, isobaric, and adiabatic processes is determined by whether pressure or volume is increasing. (Examples of (v) are that $\Delta E > 0$ for an isochoric process in which pressure increases, $\Delta E > 0$ for an isobaric expansion, and $\Delta E < 0$ for an adiabatic expansion.).

1. Internal energy E is proportional to temperature T for an ideal monatomic gas

Problems in the first concept group for which we investigate problem-property dependence can be answered by understanding that internal energy E is proportional to temperature T for an ideal monatomic gas, i.e., $E = (3/2) N k_B T$ (all notations are standard). For these three problems (problems 33, 34, and 69 on the survey), PV diagrams for the processes are shown together with the problem statement, and the initial and final temperatures are explicitly indicated on the diagram. Each problem requires responses regarding ΔE. The fact that ΔE is directly proportional to ΔT can be used to solve these problems correctly; when T increases, ΔE is positive and when T decreases, ΔE is negative, and ΔE is the same for two processes involving the same temperature change. Problem 33 involves an isochoric process in which temperature increases (Process 1 in Fig. 3), while problem 34 involves an isobaric process in which temperature decreases (Process 2 in Fig. 3); answer options for those two processes are that internal energy increases, decreases, remains the same, or there is not enough information. On problem 69 (Fig. 4), students are asked to compare ΔE for two processes shown on a PV diagram that start and end at the same temperatures; these processes are explicitly identified in the problem statement as "adiabatic" (Process 1) and "isochoric" (Process 2). (Since ΔT is the same for both processes, ΔE is the same as well.)

Figure 5 shows the variation in correct-response rates across the three problems in a pattern that is consistent across all three groups and statistically significant for both introductory groups, implying substantial problem-property dependence in student thinking for those groups. In particular, Fig. 5 shows that problem 33 (involving the isochoric temperature-increasing Process 1 in Fig. 3) was the "easiest," having the highest correct-response rate ($p < 0.001$ for the introductory groups), followed by problem 34 (referring to the isobaric temperature-decreasing Process 2 in Fig. 3), while problem 69 (comparing ΔE for the explicitly identified adiabatic and isochoric processes shown in Fig. 4) was the most challenging of the three

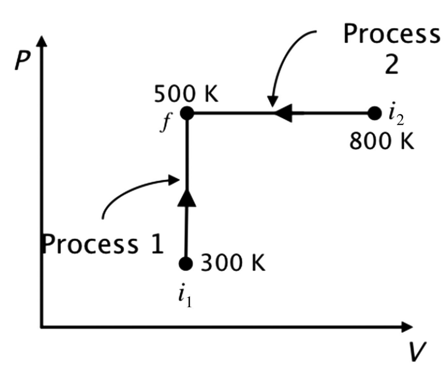

FIG. 3. Diagram provided with problems 33 and 34.

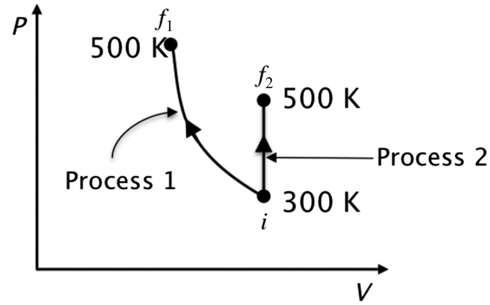

FIG. 4. Diagram provided with problem 69.

problems ($p < 0.001$ for all student groups). Although the upper-level students performed significantly better than the introductory students on items 34 and 69, many students in the upper-level group also failed to consistently apply the concept of proportionality of ΔE and ΔT.

We note that within this problem group, several problem features may be influencing student thinking, e.g., whether temperature *increases* or *decreases*, the type of process (adiabatic, isochoric, isobaric), the direction and shape of the process arrows on the diagram, and whether or not the type of process is *explicitly* identified within the problem statement. However, we have inadequate evidence to pin down the specific influence of particular problem features in these cases, as is discussed further below. Additional research would be needed to disentangle the relative influence of these properties.

On problem 69 (in which ΔE is the same for both processes), we found that, for all student groups, many more students said that internal energy change in the adiabatic Process 1 was *greater* than in the isochoric Process 2, compared to those who said it was less. The difference was highly significant for the introductory groups ($p < 0.001$) and exceeded a 2∶1 ratio for the Int-alg group. This certainly suggests that either the term "adiabatic" or the constant-volume nature of Process 2—or both—had a significant effect on those students' thinking. Based on previous reports, and even on responses to

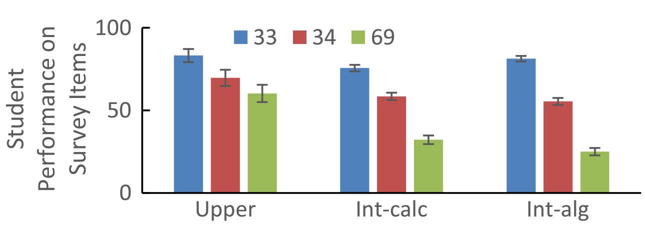

FIG. 5. Percentages of correct responses given by upper-level (Upper), introductory calculus-based (Int-calc), and introductory algebra-based (Int-alg) physics students on problems related to ΔE being proportional to ΔT; the three problems 33, 34, and 69 incorporate different thermodynamic processes. The widely varying correct-response rates for each of the three student groups (particularly both introductory groups) provide evidence for problem-property dependence of student understanding.

another of our survey items (problem 2; see discussion in Sec. III A 5 below), one might have expected more responses consistent with an assumption that internal energy did not change at all in the adiabatic process, rather than that it had the *most* change; however, that was not what we found. Another feature that may influence students is the direction (straight upward or curved) of the process arrow on this problem's PV diagram. Whatever the reasons for the variation in correct-response rates, its mere presence —and substantial magnitude—suggests clearly that many students, both introductory and upper-level, are not able to *consistently* apply the proportionality of ΔE and ΔT for an ideal gas. Although this finding has been hinted at in previous research [26], the clear problem-property dependence shown here provides an unambiguous indication to instructors that extra emphasis may be needed when discussing this particular topic. Moreover, our specific findings imply that further investigation is needed to sort out the relative influence on students' thinking of various problem properties, including those identified here. These results (as do many that follow below) provide a strong caution that adequately assessing student understanding of a specific thermodynamics concept such as E-T proportionality may require multiple diverse problems in which careful attention is paid to the potential influence of *specific problem properties* on students' thinking.

One might ask whether this analysis has identified any "salient distracting features" among this set of problems. We cannot definitively answer this question. The fact that, for example, a significantly larger proportion of students answered problem 33 correctly than they did problem 34 may be due to one or both of two specific features: (i) temperature is *increasing* in problem 33 while it is *decreasing* in problem 34; (ii) problem 33 involves an *upward* pointing ("zero work") arrow, while problem 34 involves a *sideways-pointing* ("negative work") arrow. Our interview data suggest that some students were distracted by trying (unsuccessfully) to analyze the zero-work and negative-work conditions to arrive at a solution. (Heat transfer was consistently ignored.) With the limited evidence available to us, the cause of the response-rate difference remains uncertain.

In problem 69, it is reasonable to speculate that the need to *compare* two processes (rather than merely analyze a single process) is itself a property that tends to cognitively overburden students and decrease their correct-response rates. However, we also observed an enormous difference between those introductory students who (incorrectly) stated that the adiabatic process had greater internal energy change compared to those who (incorrectly) stated that it had the smaller energy change, a result that is somewhat surprising in view of previous reports that students often associate adiabatic processes with zero energy change. It is unclear whether the term "adiabatic" or the different shapes of the process arrows—or both—acted as salient distracting features in this problem. However, most of the students who gave incorrect answers in the interviews were clearly distracted by attempting to employ work considerations (while ignoring heat transfer) to arrive at an answer, prompted by the specific shapes of the process arrows; the volume compression of the adiabatic process was associated with an energy change (either positive or negative) while the isochoric process lacked that volume-change feature. The term "adiabatic" did not appear to play much role in their thinking. This observation, consistent with the interview data for problems 33 and 34, does seem to suggest that the shapes and directions of the process arrows may be significant salient distracting features, simply because they seem to cue fruitless attempts to employ work considerations, often ignoring heat transfer.

Note: We present here a more detailed statistical analysis of Fig. 5 as an example of the analysis we performed on all other figures discussed in this paper. We use Cohen's h for effect sizes and a test of binomial proportions for p values. For further details, see the Supplemental Material. Note that the figure itself shows that (i) the mean correct-response rate for each of the three items within each of the three groups is separated by around two standard errors, or more, from the other two means and (ii) the correct-response rates for the three items differ from each other within each group by at least 10%.

Int-alg: Correct-response rates on items 33, 34, and 69 are all significantly different from each other, with the three effect sizes for (i) 34 compared to 33, (ii) 34 compared to 69, and (iii) 33 compared to 69 ranging from 0.57 (medium) to 1.20 (large) with $p < 0.001$ in all three cases. *Int-calc*: Correct-response rates on items 33, 34, and 69 are all significantly different from each other, with the three effect sizes ranging from 0.37 to 0.90 and $p < 0.001$ in all three cases. *Upper*: Effect sizes range from 0.20 to 0.52, with $p < 0.05$ for the 33/34 and $p < 0.001$ for the 33/69 comparison. *Between-group comparison*: The Upper group performance on items 34 and 69 was 12% or more above that of both introductory groups (effect size range 0.23–0.73, $p < 0.05$ on 34 and $p < 0.001$ on 69).

2. $\Delta E = 0$ *for an isolated system undergoing a spontaneous and irreversible process*

The second problem group in Fig. 6 involves the concept that $\Delta E = 0$ for an isolated system undergoing spontaneous and irreversible processes; this group includes problems 13, 23, 50, and 72. In an isolated system, there is no outside environment with which to interact, so $W = 0$, $Q = 0$, and $\Delta E = 0$. Within each of the three student groups, correct-response rates varied substantially from problem to problem. The performances of the two introductory groups were comparable to each other, while the upper-level students performed somewhat better.

Although all four of these problems involve isolated systems undergoing spontaneous processes within containers described as "insulated," their surface features differ

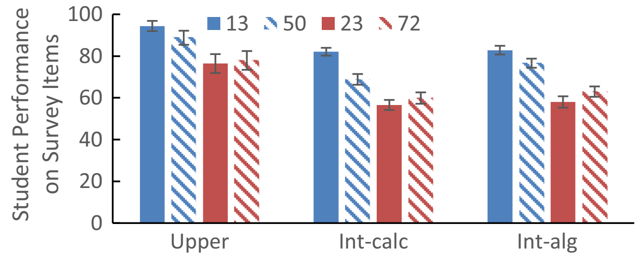

FIG. 6. Percentages of correct responses given by upper-level (Upper), introductory calculus-based (Int-calc), and introductory algebra-based (Int-alg) physics students on problems related to ΔE for an isolated system; error bars represent standard errors. The four problems (13, 50, 23, and 72) involve four different processes each in a different physical setting. Problems involving heat transfer between different-temperature solids or gases in contact (problems 13 and 50) had the highest correct-response rates, while both free expansion (problem 23) and mixing of two non-interacting gases (problem 72) had lower correct-response rates.

substantially from each other. Problem 23 involves a free expansion process; an ideal gas initially in one chamber expands into a second chamber containing a vacuum when a stopcock is opened. Problem 72 involves two different noninteracting ideal gases initially in separate chambers, mixing freely with each other when a stopcock is opened. The surface features of problems 23 and 72 can be viewed as similar to each other in that the mixing of two non-interacting gases can be viewed as somewhat analogous to a free expansion of a single gas. Indeed, the correct-response rates for these two problems did not differ significantly from each other for any group (5% or less). Problems 13 and 50 appear similar to each other in that each involves two subsystems in thermal contact with each other, allowing thermal energy transfer between them. Problem 13 involves two solids at different temperatures in contact with each other, while problem 50 involves two gases at different temperatures in similar contact. The magnitude (and significance) of the response-rate differences between those problems varied from group to group. Problems 23 and 72 (free expansion/mixing) had similar correct-response rates to each other but were lower than those of problems 13 and 50 for each student group. The correct-response rate for problem 13 was significantly higher ($p < 0.001$) than for both problems 23 and 72 for the introductory students. The problem-50 rate was higher than that of problems 23 and 72 for all student groups, with varying significance ($p < 0.01$ for Int-alg, $p < 0.05$ for upper-level and Int-calc), suggesting at least marginal significance.

It is notable that students seemed to struggle more on problems involving free expansion and mixing of gases (problems 23 and 72) than on problems involving heat transfer between subsystems at different temperatures (problems 13 and 50). Our findings thus suggest that students find it *easier* in heat-transfer settings, and *harder* in expansion and mixing settings, to apply the concept that $\Delta E = 0$ for an isolated system undergoing a spontaneous and irreversible process. (It is interesting to compare this finding to results on entropy questions involving these very same settings, to be discussed later in Sec. III D 1, as the entropy questions show a distinctly different pattern of relative difficulty for the identical set of physical scenarios.)

Some evidence regarding the specific reasoning processes followed by the students is provided through the interviews. One interviewee who thought, incorrectly, that internal energy would decrease in problem 72, gave as the reason "because the work is positive, the internal energy has decreased," thus associating positive work with expansion and mixing even when not appropriate to do so due to the nonquasistatic nature of the process. Another student stated that there was not enough information to answer problem 23, saying "Ok, so the internal energy is heat minus work, so the work is positive. I don't know what the heat situation is." This student did not realize that the freely expanding gas constituted an isolated system for which $\Delta E = 0$, and that both heat and work are zero in this case. We find that the temptation to apply concepts and formulas that have become familiar to students from settings involving quasistatic, reversible processes is strong and frequently misleading when attempting to analyze processes that are spontaneous and irreversible. This misapplication then leads to a substantial problem-property dependence in responses to questions involving irreversible processes, with surface features in those problems therefore playing a large and inappropriate role in students' thinking. In particular, both mixing and free expansion seem to be particularly challenging settings in which to apply this concept, apparently because they mislead students into thinking that work is nonzero in those cases. This particular distraction does not seem to apply in heat-transfer settings, suggesting that free expansion and mixing represent salient distracting features in these situations.

3. E is a state variable (also implying $\Delta E = 0$ for a system undergoing a cyclic process)

Solution of items in the third problem group, shown in Fig. 7, requires understanding and applying the concept that internal energy E is a state variable. Specifically, to answer these problems correctly, students must realize that E of a system represented by a specific point on a PV diagram does not depend on the process that led to it. This group includes problems 6, 25, and 47. Problems 6 and 25 both ask whether ΔE is greater than, less than, or equal to zero for a system undergoing a complete cyclic process that is shown on an accompanying PV diagram. (The diagrams show a counterclockwise process for Problem 6, and a clockwise process for Problem 25). Question 47 is a true/false question that asks explicitly whether the value of the internal energy of a system is determined by the state of the system or by the process that led to that state. (We note that

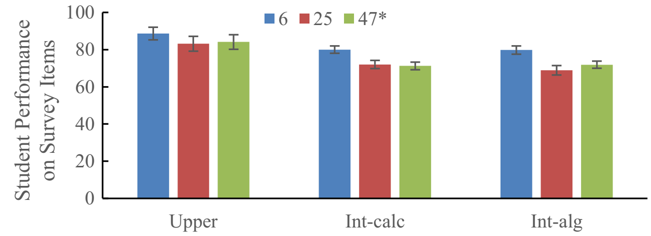

FIG. 7. Percentages of correct responses given by upper-level (Upper), introductory calculus-based (Int-calc), and introductory algebra-based (Int-alg) physics students on problems related to the state-variable property of E; the three problems 6, 25, and 47 all have different properties. The correct-response rates for the three problems are similar to each other, thus showing weak problem-property dependence of student thinking for each student group. Problem 47, marked by an asterisk (*), is a T/F problem but shows comparable student performance to problems 6 and 25 which have four choices apiece.

since there are only two choices for a T/F question compared to the other two problems which each have four choices, correct-response rates cannot directly be compared with those two problems, even though—as it turns out—all three percentages are comparable in this case.) Figure 7 shows that each student group was reasonably consistent in providing correct responses across the different problems. In particular, Fig. 7 shows that correct-response rates of students within each student group do not vary significantly from problem to problem, with a maximum variation ~10%. (We note that responses on incorrect-answer options differ by the same small amount, indicating little influence of surface features on students' thinking regarding these problems.) The upper-level students performed somewhat better than the two introductory student groups, both of which performed comparably to each other. Specifically, 83–89% of upper-level students gave a correct response on these problems, while 71%–80% of Int-calc and 69%–80% of Int-alg students did so.

4. $\Delta E = 0$ for an isothermal expansion

The first problem group in Fig. 8 includes two problems that focus on the internal energy of an ideal gas undergoing an isothermal expansion. Problems 61 and 65 both ask whether E increases, decreases, or does not change for a gas undergoing an isothermal expansion, but problem 65 included an additional explicit stipulation that the process was reversible. (No PV diagram is provided in either case.) The knowledge that needs to be invoked and applied to solve this problem is an understanding that E is proportional to T for an ideal gas, so E can't change if T doesn't change. Figure 8 shows that students at all levels performed relatively poorly on these two problems, with less than half (28%–49%) of the introductory students providing a correct response; upper-level students did best, and the Int-alg group did worst. The data also show that explicit

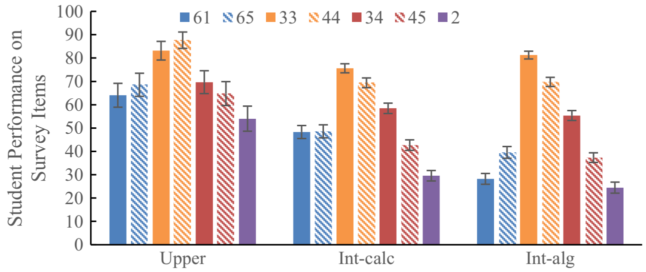

FIG. 8. Percentages of correct responses given by upper-level, introductory calculus-based (Int-calc), and introductory algebra-based (Int-alg) physics students on problems related to ΔE for various thermodynamic processes, including problems 61 and 65 (isothermal), 33 and 44 (isochoric), 34 and 45 (isobaric) and 2 (adiabatic). The widely varying correct-response rates for each student group provide evidence for problem-property dependence of student performance.

mention in problem 65 that the isothermal process is reversible did not appear to significantly impact student performance for upper-level and Int-calc groups but may have done so to some extent for the Int-alg group (12% higher correct-response rate on problem 65; $p < 0.001$). Both $\Delta E > 0$ and $\Delta E < 0$ were common incorrect responses for both problems, but both the upper-level group and the Int-alg group gave significantly more "increases" than "decreases" responses on *both* problems ($p \leq 0.002$); this result strongly suggests that for some student groups at least, "expansion" tends to cue "increase" even in cases where a thermodynamic variable is unchanging. In interviews, students used work arguments to support both E increases *and* E decreases as incorrect responses, generally a consequence of neglecting to consider heat transfer.

Interview data suggest that a common pathway toward incorrect reasoning lay in attempting to explicitly solve for ΔE using the first law of thermodynamics $\Delta E = Q - W$, even though the problem statements did not include enough information to do so. For example, on problem 61, a student who arrived at the incorrect response $\Delta E < 0$ said, "Ok, so for isothermal…if [the gas] is expanding then I think the gas is doing work, which means [work] is going to be positive. And then if there is heat being transferred away from the gas then that would make the internal energy probably decrease." Although the student correctly stated that $W > 0$, they mistakenly assumed that there would be net heat transfer *away* from the gas. Another student who gave the $\Delta E > 0$ response for problem 61 said, "Ok, so isothermal expansion. Internal energy… if the gas is expanding, then I think so the volume would increase, so I think the internal energy increases." These responses suggest that students often were so focused on attempting to apply the first law of thermodynamics that they failed to realize that the proportionality between E and T immediately implied that $\Delta E = 0$ for the isothermal process setting.

Although high error rates on isothermal-process problems have been noted in prior student difficulties research, the "first-law focus" explanation for students' thinking regarding ΔE in isothermal processes does not appear to have been reported. The apparent tendency to favor an (incorrect) "increases" response in the case of an isothermal expansion also appears to be a new result.

5. The sign of ΔE for an ideal gas undergoing isochoric, isobaric, and adiabatic processes is determined by whether pressure and/or volume are increasing

We consider the three types of processes individually.

(i) *$\Delta E > 0$ for an isochoric process in which pressure increases:* The second problem group in Fig. 8 focuses on the internal energy of an ideal gas undergoing an isochoric process in which pressure increases; this group includes problems 33 and 44. We note that the processes in problems 33 (Fig. 3, temperatures shown explicitly) and 44 (Fig. 9, no temperatures shown) were both depicted on PV diagrams. Both problems 33 and 44 refer to isochoric processes represented by upward-pointing arrows, thus implying (from $PV = Nk_BT$) that $\Delta T > 0$ and therefore $\Delta E > 0$. The PV diagram for problem 33 explicitly shows the temperatures associated with the initial and final states of the process, information that is absent from the problem 44 diagram; compare Figs. 3 and 9. Figure 8 shows that Int-alg students found problem 33 (temperatures shown) to be easier than problem 44 (temperatures absent). It is notable that the Int-alg students did significantly worse on problem 44 ($p < 0.001$), suggesting that the lack of temperature information was particularly troublesome for them (in contrast to the Int-calc and upper-level students). For both problems, $\Delta E = 0$ was the most common incorrect response. In the interviews, some students focused only on the fact that $W = 0$ in an isochoric process and concluded that this implies $\Delta E = 0$, thus ignoring the role of heat transfer Q. The focus on W while ignoring Q is similar to the reasoning patterns we noted above for other ΔE problems. Similar observations of students restricting their reasoning to only two of three relevant thermodynamic variables have

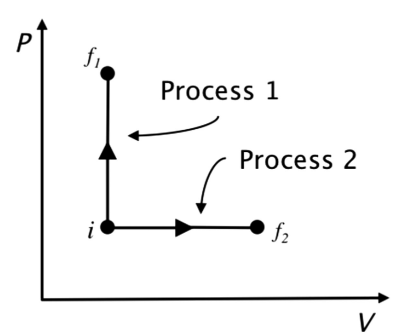

FIG. 9. Diagram provided with problems 40–45.

been reported in prior student difficulties research [29,33].

(ii) *The sign of* ΔE *for an isobaric process is determined by whether the volume is increasing or decreasing:* The third problem group in Fig. 8 focuses on the internal energy of an ideal gas undergoing an isobaric process; this group includes problems 34 and 45. The PV diagram in problem 34 shows a volume compression and explicitly indicates both final and initial temperatures, the temperature decrease thus implying that $\Delta E < 0$. By contrast, the diagram for problem 45 depicts an expansion, thus implying (from $PV = Nk_BT$) that $\Delta T > 0$ and therefore $\Delta E > 0$, although it omits any *explicit* indications of temperature. For both problems, the use of the ideal gas proportionality between E and T leads to the correct answer. Figure 8 shows that problem 34 (temperatures shown) is significantly less challenging than problem 45 for the introductory student groups ($p < 0.001$). Interviews suggested that faulty reasoning regarding work was more prevalent on problem 45, but it is not clear whether this was due to the temperatures not being shown or because the volume was *increasing*. It is notable that upper-level performance on the two problems was essentially identical. Generally, students were more likely to get the sign of ΔE wrong than state that $\Delta E = 0$ (although $\Delta E = 0$ was also a common response).

Among the students interviewed, most of those who provided incorrect responses on problem 34 ignored the fact that the final temperature on the PV diagram was lower than the initial temperature. Instead, they claimed that since the compression implied that the work done by the system was negative, ΔE must be greater than zero. They simply did not account for heat transfer, another example of ignoring a crucial variable. Similarly, for problem 45, seven of the nine interviewees providing incorrect responses stated that $\Delta E < 0$ because the work done by the system was positive, again ignoring the role of heat transfer. In that sense, most of the interviewees who provided incorrect responses to these two problems had consistent incorrect reasoning that ignored the essential variable Q. As noted, correct-response rates were higher on problem 34. Since the temperatures of the initial and final states were provided explicitly in problem 34, fewer students may have been misled to ignore heat transfer in that problem. However, those who focused on the first law in the interviews indeed were misled, as they often only considered W and ignored Q.

(iii) $\Delta E < 0$ *for an adiabatic expansion:* Problem 2 asks about the internal energy of a gas undergoing "a reversible adiabatic expansion process"; no diagram is provided. The correct response is $\Delta E < 0$ because positive work is done by the system and $Q = 0$, so $\Delta E = Q - W < 0$. Figure 8 shows that only 54% of upper-level students, and 30% and 24% of the Int-calc and Int-alg students, respectively, provided correct responses. This problem differs from the others discussed above in that the proportionality between E and T does not lead directly to an answer *unless one already knows* that temperature decreases in an adiabatic expansion; instead, the first law of thermodynamics and the sign of work can be applied to find a solution. However, this complication should not prevent students from giving *some* $\Delta E \neq 0$ answer, regardless of sign. Nonetheless, the most common incorrect response was indeed $\Delta E = 0$ (comprising 39%–47% of the responses by the introductory students), followed by $\Delta E > 0$. (We note that the $\Delta E = 0$ response appears to be in conflict with the most popular response to problem 69, which was that E *would* change in an adiabatic process; see discussion in Sec. III A 1 above.) In the interviews, three of seven students who provided an incorrect response stated that $Q = 0$ implies $\Delta T = 0$ and $\Delta E = 0$, consistent with the most popular response. Other students who provided the incorrect response $\Delta E > 0$ used incorrect reasoning involving the first law, e.g., stating that if $Q = 0$ and the gas expands, $\Delta E > 0$. Often this latter set of students did not reason conceptually to see that expansion work done by the system implies a decrease in the system's internal energy; instead, they algorithmically applied the first-law equation $\Delta E = Q - W$ but made a sign error. Another possible explanation for incorrect responses on this problem is that many students simply do not realize that heat transfer Q in any adiabatic process is, by definition, zero. We know this to be the case because problem 31—not included in the figures in this paper—is a simple true or false question regarding the statement "there is NO net heat transfer between the system and the surroundings"; the question is whether this is true or false for "any adiabatic process." The correct response rate for both the Int-calc and Int-alg group on this question was only 69% (and 87% for the upper-level group), indicating significant confusion regarding the definition of adiabatic *even though the introductory section of the survey explicitly defines an adiabatic process* as "one in which there is no heat transfer between a system and its surroundings."

Comparison of responses on problems 44 and 45: Problems 44 (isochoric) and 45 (isobaric) both relate to processes depicted on a PV diagram in which internal energy *E increases*; in neither case is the temperature explicitly indicated on the diagram. Despite these similarities, the correct-response rate (that E increases) is *dramatically* higher ($p < 0.001$) for the isochoric process (problem 44) compared to the isobaric process (problem 45) for all student groups. (Upper: 88% and 65%; Int-calc: 69% and 43%; Int-alg: 70% and 37%.) The problem-property dependence of this problem pair is as large as any that we observed. However, it is not clear whether the greater success rate on the isochoric process is due to the zero-work nature of that process or, more simplistically, merely due to the apparent cueing provided by an *upward*-pointing arrow representing that process. An answer is suggested by the seven of the nine interviewees providing incorrect responses on problem 45 who stated that $\Delta E < 0$ because the work done by the system was positive. They clearly ignored the role of heat transfer as well as the higher temperature implied by the increased value of PV. It seems that on the "vertical line" isochoric process, the relatively obvious fact that work done is zero in that case removed a key salient distracting feature (that is, an expansion process), although it is still not clear from the interviews exactly how students arrived at their correct answer on that problem.

Summary of Sec. III A: We find that an inconsistent and unreliable application of the E-T proportionality and the state function property of E (i.e., that a specific point on a PV diagram is linked to a specific value of E) often allows superficial problem features to dominate students' thinking on many of the E-related concepts that we investigated. The problems in which temperatures are explicitly provided are associated with higher correct response rates, but not dramatically so. Interviews showed that flawed reasoning regarding work was the most common obstacle to correctly applying the E-T proportionality; students' thinking was consistently distracted by specific shape features of the process arrows on PV diagrams. Explicitly identified process characteristics such as "reversible" and "expansion" may also significantly distract some students' thinking and, in certain cases, make it less likely that they invoke and apply the state function property correctly. The term "isothermal" does *not* often cue the expected correct response $\Delta E = 0$ in part because many students fail to associate temperature with internal energy.

We noted that students were more likely to struggle with the reversible adiabatic expansion (problem 2) than with any other problem related to ΔE, as almost half of the introductory students thought incorrectly that $\Delta E = 0$ in that case. Although the $\Delta T = 0$ (isothermal) expansion process did *not* tend to cue a (correct) response that $\Delta E = 0$, it is notable—and somewhat ironic—that the adiabatic $Q = 0$ expansion did often cue an *incorrect* $\Delta E = 0$ response. Interviews confirmed that students often associated "adiabatic" with "no energy change." We note that previous reports show that students often associate thermal insulation with constancy of temperature [47], and that association may well extend to processes described as adiabatic. We note that a "cognitive load" argument that would connect the difficulty of problem 2 primarily to the need to invoke the first law of thermodynamics should probably also imply a larger number of sign errors than $\Delta E = 0$ errors, which is the opposite of what we observed. Thus, it appears that the bare term "adiabatic" itself often functions as a salient distracting feature, here and perhaps in other settings as well; this conclusion seems to have been borne out in the interviews.

In marked contrast to problem 2—for which $\Delta E = 0$ was the most popular response—most students considering the adiabatic process in problem 69 felt that not only did it have $\Delta E \neq 0$ (correct), but that it had a *larger* change in internal energy than an isochoric process with identical initial and final temperatures (incorrect). Interviews indicated that the low correct-response rate on problem 69 seems more closely associated with flawed reasoning regarding work than with the adiabatic nature of the process. In fact, the interviews strongly suggest that the salience of *work* outweighed the salience of *adiabatic* in this problem group, with both features ultimately distracting students into flawed and unsuccessful reasoning pathways.

B. Problem-property dependence of student responses on problems related to work

The three problem groups in Figs. 10, 11, and 14 focus on concepts related to work W done by the system in various thermodynamic processes. The concepts related to work for which we investigate problem-property dependence include (i) $W = 0$ when there is no net system expansion against a movable boundary wall, (ii) W may be interpreted as the (signed) area under the curve in a PV diagram, and (iii) W is positive for an expansion, whether isothermal or adiabatic.

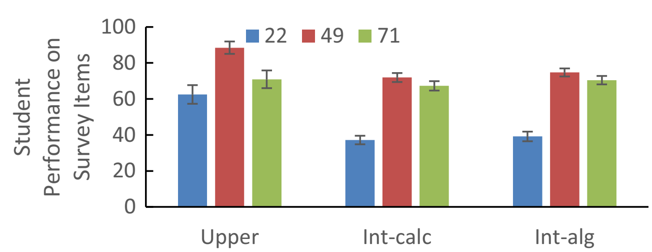

FIG. 10. Correct-response rates on problems involving $W = 0$ for various spontaneous processes in isolated systems. Percentages of correct responses given by upper-level (Upper), introductory calculus-based (Int-calc), and introductory algebra-based (Int-alg) physics students are shown for problems 22 (free expansion), 49 (two gases in thermal contact at different temperatures), and 71 (mixing of two noninteracting gases). The widely varying correct-response rates for each student group provide evidence for problem-property dependence of student understanding related to this concept.

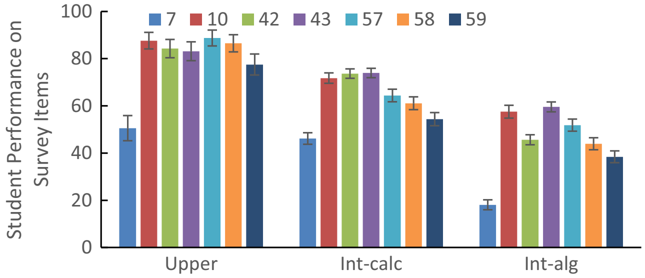

FIG. 11. Correct-response rates on problems involving work as the area under the curve in a PV diagram. Percentages of correct responses given by upper-level (Upper), introductory calculus-based (Int-calc), and introductory algebra-based (Int-alg) physics students are shown for problems 7, 10, 42, 43, 57, 58, and 59, all of which involve problem properties that differ from each other. The widely varying correct-response rates for each student group provide evidence for problem-property dependence of student performance related to this concept.

1. $W = 0$ when there is no net system expansion against a movable boundary wall

The first group, shown in Fig. 10, includes problems 22, 49, and 71, and deals with W for various spontaneous nonequilibrium processes in isolated systems. Problem 22 refers to a free expansion process, problem 49 is about spontaneous heat transfer between two gases at $T_C \ll T_H$ when they are in contact, and problem 71 is about the spontaneous mixing of two noninteracting gases initially in two different adjacent chambers. Correct-response rates show considerable problem-property dependence and were highest on problem 49 for all student groups, followed by problem 71; they were lowest by far on problem 22. (Only the upper-level group had a *significant* difference between 49 and 71.) Arguably, the biggest challenge for students on these problems is to recognize that the formula $W = \int PdV$ (or its algebra equivalent) that applies to quasistatic near-equilibrium processes is simply inapplicable here. Since there is no net system expansion against a movable boundary wall, no work is done by the system in any of these scenarios. Nonetheless, many students did not treat all three scenarios as equivalent regarding work.

Problem 49 specifically states that gas molecules are confined to their original chambers and only heat transfer can take place freely between the two chambers; the setting is therefore clearly "constant volume." Thus, it is not surprising that it is easiest for students to recognize in this situation that there is no change in volume and so $W = 0$. On the other hand, problem 22 involving a free expansion process describes a gas that was initially confined to only one chamber, eventually occupying both chambers at the end of the process. Similarly, problem 71 about the mixing of noninteracting gases describes each type of gas initially present only in its own chamber, then mixing when the stopcock is opened. The setting of problem 22 may be interpreted as "changing volume" and, while that of problem 71 is similar, it is arguably more ambiguous due to the presence of gas in each chamber before mixing.

Most students who provided incorrect responses to problems 22 (free expansion) and 71 (mixing) thought that $W \neq 0$, with many claiming that the work done by the system is positive. Among those interviewed, nine of ten students who provided incorrect responses claimed that the work done by the gas would be positive ($W > 0$) because the gas expands. By contrast, only two of the interviewed students provided incorrect responses to problem 71; they both claimed $W > 0$. Since the focus was on mixing in problem 71 as opposed to free expansion or volume change in problem 22, more students thought that $W \neq 0$ on problem 22. During the interviews, one student showed some confusion about their response to problem 22 by saying, "I think there's no work done because the gas doesn't act on anything, it's just expanding…but expansion of a gas is also positive work, so that's a little confusing. So, I'm just going to say that positive work is done." Although the upper-level students performed better than the introductory students (see Fig. 10), their difficulties with W also persisted across diverse settings since 11%–37% of them did not realize that $W = 0$ in these processes.

We can summarize the findings for this group of problems as strongly supporting previous reports that free-expansions are particularly challenging settings for applying the work concept while adding that—in contrast—students are better able to apply that concept to spontaneous processes in which container volume does not change.

2. W may be interpreted as the (signed) area under the curve in a PV diagram

The group in Fig. 11 includes the seven problems 7, 10, 42, 43, 57, 58, and 59. They all require comprehension and comparisons of work done by the system in different processes or assessments of whether work done in a particular process is positive, negative, or zero (or if there is not enough information to decide). The approach normally taught in introductory thermodynamics classes for solving this type of problem makes use of the "work equals area under the curve" interpretation of (curved or straight) paths on PV diagrams that represent reversible processes; work done by the system is positive for expansion processes and negative for compressions. (The relationship $W = P\Delta V$ may also be directly applicable in certain cases, but it can also be misleading when applied too broadly.) Figure 11 shows a great deal of variation in correct-response rates for these problems among all student groups, suggesting that students' ability to correctly apply this interpretation of work is highly dependent on problem properties.

Problem 7 provides students with a PV diagram showing a counterclockwise cyclic process and asks whether the net work done by the system in one complete cycle is positive, negative, zero, or whether there is not enough information

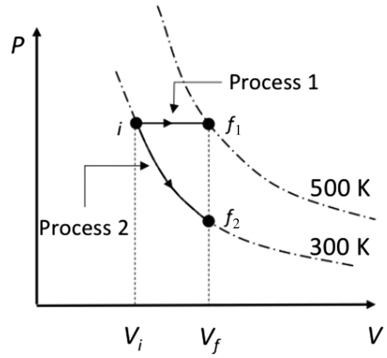

FIG. 12. Diagram provided with problem 10.

to decide; this is the only cyclic process problem in this group and performance was the poorest by far on this problem. In problem 10, there were two expansion processes represented on a PV diagram (Fig. 12): a constant pressure process (Process 1) and an isothermal process (Process 2), and students were asked to determine the process in which the amount of work done by the gas is largest. Problems 42 and 43, both shown on a PV diagram (Fig. 9), involve an ideal monatomic gas undergoing an isochoric process in which the pressure increases (problem 42) and an isobaric expansion (problem 43); students were asked whether W was positive, negative, or zero. In problems 57–59 (Fig. 13), students were shown two expansion processes on a PV diagram: one being an isothermal process and the other an adiabatic process: $W > 0$ for both expansions. In problem 57, students were asked whether work done by the system in the isothermal process was positive, negative, or zero; in problem 58, the same question was posed in regard to the adiabatic process. In problem 59, similar to problem 10, students were asked to compare the magnitudes of the work done in the two processes shown on the diagram. By applying the area under the curve interpretation to the PV diagram in Fig. 13, it is easy to infer that the work done is greater in the isothermal process since both processes share the same initial and final volume.

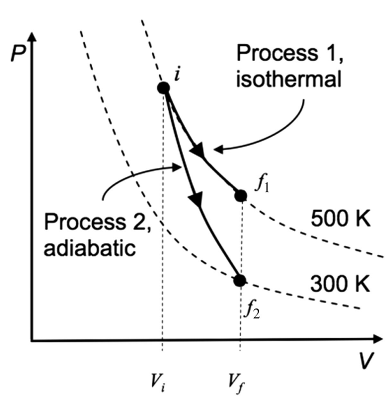

FIG. 13. Diagram provided with problems 57–59.

Overview of results. Figure 11 shows that the upper-level students performed reasonably well (> 77% correct responses) on all problems *except* for the cyclic process in problem 7 (51%). However, Int-calc students did substantially worse than upper-level students, and the Int-alg group performed poorly on all problems in this group, with a high score of 60% on problem 43. Scores of the Int-calc group ranged widely from 46% to 74%, while those of the Int-alg group ranged from 18% to 60%, suggesting a strong problem-property dependence in students' ability to solve these problems.

Cyclic process problem. We found that the counterclockwise cyclic process in problem 7 was the most challenging for all student groups. Previous research has indicated substantial confusion in students' interpretations of work done in a cyclic process [26]. (In our present study, the three student groups differed among themselves on the preferred incorrect response; see Ref. [65].) In an interview, one student who incorrectly claimed that $W > 0$ said, "I'm thinking that it is positive. I am thinking about the equation, Work = PV and then yeah, it would be like positive…". For the upper-level and Int-calc students, sign errors were more common than claims that $W = 0$, although the latter error category was also well represented (16% and 21%). For the Int-alg students, $W = 0$ was by far the most popular response, given by 51% of that group; this indicates quite a large difference between the Int-alg and Int-calc groups on this particular error. Based both on our results here and on previous reports in the literature, it seems quite reasonable to conclude that for the algebra-based students, the cyclic nature of the process represents a salient distracting feature that obstructs students' attempts to apply either the "area under the curve" mnemonic or some other principle that might help them arrive at the correct answer. It also seems quite likely that it acts as an SDF for the other student groups as well, but for them, it may be overshadowed by difficulties in applying the mnemonic correctly so as to avoid sign errors.

Introductory groups. Since upper-level students performed reasonably well on all problems except the cyclic process, here we mainly focus on the performance of the introductory groups. The most challenging of the noncyclic-process problems was 59 (comparing W for adiabatic and isothermal processes depicted in a PV diagram), while no problem was easier than problem 43 (isobaric expansion, depicted by a horizontal line on a PV diagram). It is notable that in problem 10, the only other problem requiring a comparison of work done in two different processes, correct-response rates were almost identical to those of the "easy" problem 43 and not the much lower rates of the work-comparison problem 59. For the introductory groups, the difference between their problem-10 and problem-59 performance was highly significant ($p < 0.001$), implying that problem properties played a significant role. In problem 10, one of the processes being compared was isobaric, represented by a horizontal line on the PV diagram, just as in problem 43. This may suggest that for isobaric processes represented by horizontal lines on PV diagrams, students may be more easily able to invoke and correctly apply the work equals area under the curve interpretation. Interviews indicated that the "area" approach was indeed students' primary road to success on problem 10. However, it is also possible that the inclusion of an *adiabatic* process in problem 59 served to confuse students, since—as we have previously discussed—adiabatic processes, in particular, seem to pose additional difficulties for many students. The interview data in this case were not enlightening, so evidence is insufficient to determine which of the problem properties might have been most salient in this instance.

Other notable features of the Int-alg data include poor performance (46% correct) on the isochoric process problem 42, represented by a vertical line on the PV diagram, despite what may appear to be—to a physics instructor—a straightforward argument that $W = 0$ since $\Delta V = 0$. (The Int-calc group scored 74% correct on this problem.) In striking contrast, the Int-alg group scored a significantly ($p < 0.001$) higher 60% correct on problem 43 which involved an isobaric expansion represented by a horizontal line, while the Int-calc group score on this problem was 74%, identical to their problem-42 score. For the algebra-based group, it was significantly easier to determine that positive work had been done in the isobaric expansion than that zero work had been done in the isochoric process. The Int-alg group also had slightly poorer performance ($p < 0.05$) on problem 58 (adiabatic process, 44%) compared to problem 57 (isothermal process, 52%), even though both processes are represented by curved lines directed downward on the PV diagram. Our findings suggest that specific problem properties have a significant influence on algebra-based students' performance on work problems. In fact, we propose that isobaric "horizontal-line" processes are a feature that specifically *helps* algebra-based students reason through work-related problems, whereas those processes do not appear to play that same helping role for calculus-based or upper-level students, at least not to the same degree.

Interviews suggest that some students who performed poorly on these problems knew that work done by the gas was related to pressure and ΔV but were unable to apply the area under the curve interpretation. For example, on problem 10 (Fig. 12), one student who thought that W is equal for both processes said, "I think the work done, if I remember correctly, I'm thinking work equals P delta V. They both start at the same P and then both end at the same V_f, so I'm thinking that the work done would be equal." The student correctly said that the two processes start at the same pressure and end at the same volume but overlooked the fact that pressure has decreased in one of the processes, thus implying a smaller W. In this particular problem, due to the isobaric or horizontal line process, one could successfully argue using $W = P\Delta V$. However, in problem 59 (comparing adiabatic and isothermal processes), the *nonlinear* covariation of pressure and volume suggests the need for an area under the curve interpretation, deriving from the "true" work equation $W = \int P dV$. It would be very understandable if algebra-based students, in particular, find this argument difficult to follow as they are the group with the least experience with integrals. However, although scores on the "comparison" problem 59 were lower for both Int-alg and Int-calc than those on the adiabatic process problem 58 (for which only the sign was required), the difference was not significant. Still by attempting to apply $W = P\Delta V$ in settings where it is not appropriate—instead of using "area under the curve"—students may be unable to arrive at correct conclusions, e.g., as in problem 59. Although some previous reports have examined the mathematical-context origins of student difficulties with this concept [29], our investigation of physics-specific problem-property dependences provides additional perspective on this issue.

3. W is positive for an expansion, whether isothermal or adiabatic

The group in Fig. 14 includes only problems 3 and 62; both problems use words only and lack a PV diagram. Students are asked whether work done by the system "must be" positive, negative, or zero (or if there is insufficient information to decide) in a reversible adiabatic expansion (problem 3) and an isothermal expansion (problem 62). Since both processes are expansions, work done by the system is positive in both cases. Nonetheless, correct-response rates for the introductory students were significantly higher on the isothermal expansion problem (Int-calc: 11%, $p < 0.01$; Int-alg: 13%, $p < 0.001$). This was primarily due to a higher probability of a "work done must be zero" response on the adiabatic process for these students, while sign-error rates on the two problems were

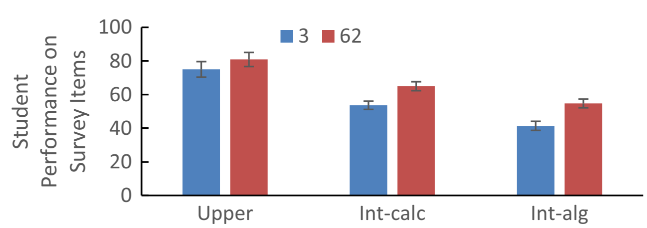

FIG. 14. Correct-response rates on problems involving the sign of work done in adiabatic (problem 3) and isothermal (problem 62) processes. Percentages of correct responses given by upper-level (Upper), introductory calculus-based (Int-calc), and introductory algebra-based (Int-alg) physics students on problems 3 and 62 are shown. The significantly different correct-response rates on the two problems for the introductory groups provide evidence for problem-property dependence of student responses among those groups.

quite comparable. (In the interviews, one of the upper-level students explicitly gave the "work is zero" response for problem 3. Other interviews suggested that the sign errors were often due to an overfocus on formula-based reasoning, rather than simply reasoning through the implications of the processes being expansions.) The additional confusion apparently introduced by the term "adiabatic" is consistent with similar findings discussed above in Sec. III A, underscoring the need for further research on student thinking related to this term. These findings increase the likelihood that the term "adiabatic" *by itself* may serve as a salient distracting feature.

Summary of Sec. III B: We find that students' ability to invoke and apply work concepts in thermodynamics tends to be problem-property dependent, in part perhaps because the work concept itself is subtle and insufficiently practiced in multiple settings and scenarios. For example, the application of work concepts to solids (that are obviously constant volume) appears to be relatively easy, whereas free-expansion processes (that obviously involve volume changes) are particularly challenging. We find that many students clearly recognize the absence of heat transfer in an adiabatic process but then tend to ignore (or misconstrue) the effects of work in that process. However, when faced with problems involving isobaric processes—represented by horizontal lines on PV diagrams—students appear to be able to recognize the presence and effects of work more easily than in other settings, yet are more prone to ignore the presence and effects of heat transfer. We hypothesize that the term "adiabatic" and the presence of horizontal lines on PV diagrams are themselves key cues for student thinking that, in many settings, may draw excessive attention and thus lead students to ignore other important factors that are essential to reasoning correctly. (These are precisely the characteristics of salient distracting features.) Cyclic processes offer special challenges for invoking and applying the work concept (as they also do for heat; see below). This is in part because it is more difficult to apply the "area under the curve" algorithm to cyclic processes on PV diagrams (involving, as they do, both negative and positive work), but also because it is particularly easy to misapply the $W = P\Delta V$ formulation in cases where the initial and final states have identical V *and* identical P. (The integral formulation of work may be easier to apply in principle, but more mathematically challenging; see Ref. [29].) Thus, the presence or absence of PV diagrams can be a factor that influences student thinking, but likely to a lesser extent than a physics instructor might anticipate because students are not able to utilize those diagrams as efficiently as an expert would.

C. Problem-property dependence of student responses to heat transfer problems

The problem groups found in Figs. 15 and 16 focus on concepts related to heat transfer, Q: (i) determining the correct sign of Q in isothermal processes, and (ii) recognizing that Q is a path-dependent variable.

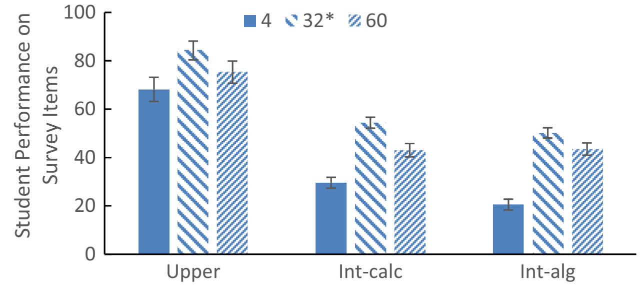

FIG. 15. Correct-response rates on problems involving determination of the sign of Q for isothermal processes. Percentages of correct responses given by upper-level (Upper), introductory calculus-based (Int-calc), and introductory algebra-based (Int-alg) physics students on problems 4, 32, and 60 are shown. The widely varying correct-response rates for each student group provide evidence for problem-property dependence of student understanding related to this concept. Problem 32, marked by an asterisk (*), is a T/F problem and cannot be compared directly with problems having four choices.

1. *In a reversible isothermal process, $Q \neq 0$ and the sign of Q is determined by whether volume is increasing or decreasing*

Figure 15 includes problems 4, 60, and the true or false problem 32. Problems 4 and 60 describe processes in words without providing any diagrams, including a reversible isothermal compression (problem 4) and an isothermal expansion (problem 60, which omits the word "reversible"). The answer options are that there is heat transfer to the gas, heat transfer away from the gas, no net heat transfer, or not enough information. Since both processes are isothermal, neither T nor E can change; thus, $Q = W$ since $Q - W = \Delta E = \Delta T = 0$. The compression in problem 4 implies $W < 0$ and thus $Q < 0$ (heat transfer *away*

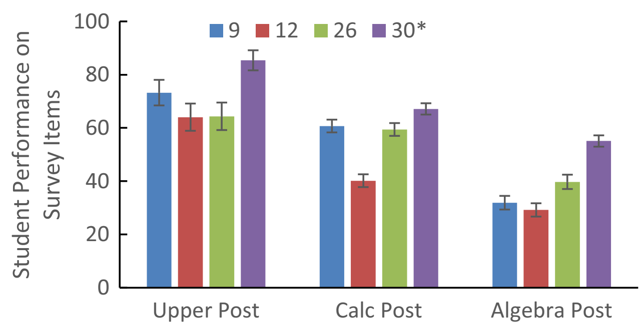

FIG. 16. Percentage of responses consistent with Q being path dependent given by upper-level (Upper), introductory calculus-based (Int-calc), and introductory algebra-based (Int-alg) physics students on problems 9, 12, 26, and 30. Problem 30, marked by an asterisk (*) is a T/F problem and cannot be compared directly with other problems with four choices.

from the gas), while the expansion in problem 60 implies $W > 0$ and $Q > 0$ (heat transfer *to* the gas). Problem 32 is a true or false problem that explicitly asks whether it is true that $Q = 0$ for "any" isothermal process.

All of these problems require some understanding of the first law of thermodynamics to answer correctly: $\Delta E = Q - W$ and $\Delta E = 0$ so $Q = W$. Correct responses were provided by no more than 75% of the upper-level students across these problems, but the correct-response rates for the introductory students on problems 4 and 60 are very low indeed: well below 50% on both problems and approaching the "random guessing" level in some cases. It is notable that in problem 4, a compression process ($W < 0$ so $Q < 0$), correct-response rates for the introductory students were significantly lower ($p < 0.001$) than in the expansion process, problem 60 ($Q > 0$), with a difference ranging from 13% (Int-calc) to 23% (Int-alg). The most common incorrect response by far was $Q = 0$, and this error was much more prevalent for the compression process among the introductory students than for the expansion (50% vs 30% for Int-calc, and 59% vs 37% for Int-alg); the difference is highly significant ($p < 0.001$). Sign errors were also common on both problems among the introductory students. Results on the true/false problem 32, as simple as the problem may appear to a physics instructor, are also quite notable, since correct-response rates for both introductory groups were at or near 50%, a rate consistent with random guessing. These low correct-response rates are consistent with previous student difficulties research [29], however, our finding that incorrect $Q = 0$ responses are significantly more common on a "reversible compression" problem is a new and potentially important result.

Interview data leave no doubt that the term "isothermal" is a powerful cue for the idea "no heat transfer." However, we have to speculate as to the reasons for the difference in correct-response rates for the introductory students on problems 4 and 60. (Since the differences for the upper-level students were not significant, we do not comment further on those.) It will be up to future research to test these conjectures. First, we observe that the difference is almost entirely due to the greater proportion of $Q = 0$ responses on problem 4 (compression), and *not* due to sign errors as one might naively have expected. It is not clear why a $Q = 0$ response would be more common for a problem involving a compression instead of an expansion. It may simply be that the term "expansion" is a more salient cue for introductory students in prompting consideration of heat transfer processes. There is some evidence from the interviews that "expansion" cued students into consideration of work more frequently than did "compression." Perhaps the term "expansion" itself cues the idea of *increase* (since volume is increasing) more readily than the term "compression" cues *decrease*. The other obvious difference is that problem 4 explicitly states that the isothermal process is *reversible*. Whether that single additional word is sufficient to trigger such a large difference in response rates across these settings requires further investigation. In the interviews, some of the students who gave correct answers on problem 60 reverted to a "no heat transfer in isothermal processes" argument for problem 4; one of them added that "reversible also means no net heat transfer." (Strictly speaking, omission of the term "reversible" and absence of a PV diagram could justify a "not enough information" response in problem 60, although the process was described as quasi-static in the introductory section of the survey. However, the tiny difference between problems 4 and 60 on this response indicates that its influence was negligible.)

Future studies could isolate the compression/expansion and the "reversible"-present/"reversible"-absent conditions to try and resolve this issue, so that instructors may become aware of which particular surface features pose special challenges in these settings and researchers developing curricula/pedagogies can account for these.

Previous investigations have noted one specific specious argument that is often observed regarding isothermal processes. It is illustrated by a response to problem 4 from one of the students in our interview sample, who said, "Isothermal is no change in temperature, so I think there is no net heat transfer because $Q = mc\Delta T$ and ΔT is zero." This confusion involving calorimetric equations—completely inappropriate in this problem—has been pointed out in the literature previously [26]. However, it was only rarely cited by the students we interviewed.

2. *Q is a path-dependent variable*

The problem group in Fig. 16 includes problems 9, 12, and 26, and the true/false cyclic process problem 30. Problems 9 and 26 show PV diagrams representing a complete cyclic process (9: counterclockwise; 26: clockwise), while the PV diagram for problem 12 (Fig. 17) shows two different reversible expansion processes (Processes 1 and 2) that share the same initial and final states but have distinctly different areas under the curve (this problem was drawn directly from Ref. [26]). Answer options for 9 and 26 are net heat transfer to the gas, net heat transfer away from the gas, no (or "zero") net heat transfer, and not enough information. Problem 12 asks students to

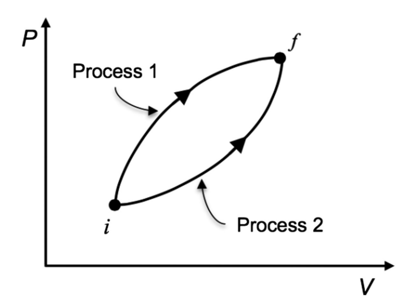

FIG. 17. Diagram provided with problem 12.

determine which of the two processes has the greater or smaller value of Q or whether instead they are equal. Here we focus *not* on correct-response rates but instead on *all* responses that were consistent with Q being a path-dependent variable. For example, on problem 9 (a cyclic process), both responses "to the gas" and "away from the gas" are consistent with Q being a path-dependent variable. For problem 12, both $Q_1 > Q_2$ *and* $Q_2 > Q_1$ meet that criterion. Even so, the total rates of such "Q is path dependent" responses on all three of the multiple-choice problems are low for both of the introductory groups (29%–40% for Int-alg, 40%–61% for Int-calc). The upper-level students scored 64% on the "path-dependent" criterion on both problems 12 and 26. For all three student groups there was little difference in response rates for problems 9 and 26, but the Int-calc group performed *much* worse on problem 12 (40% on the path-dependent criterion compared to 59-61% on the other two problems; $p < 0.001$).

Among the introductory students, the most popular incorrect responses on the three multiple-choice problems—and, indeed, very nearly the most popular overall responses—were those that were consistent with incorrectly treating Q as a state variable; these responses were $Q = 0$ for both cyclic process problems and $Q_1 = Q_2$ for problem 12. (Only for the Int-calc students on problem 9 was the correct answer ($Q < 0$) marginally more popular than the $Q = 0$ answer.) Response rates consistent with this idea ranged from 33-56% for Int-calc and 55-66% for Int-alg, highest in both cases on problem 12. Although these "state variable" errors were also common among the upper-level students, sign errors appeared frequently among them as well. It is notable that correct-response rates and $Q = 0$ response rates on the two cyclic process questions were very similar for the introductory students, indicating that whether the cycle was clockwise or counterclockwise had little influence on these students' thinking.

Problem 30 is a true/false problem that asks whether it is true that there is "NO net heat transfer between the system and the surroundings" for any cyclic process. It is interesting that correct-response rates on this problem were significantly above 50% for both Int-calc and upper-level students, but only marginally so for Int-alg students, suggesting that the latter group was largely guessing.

We note that while the cyclic process diagrams often lead introductory students directly to conclude that net heat transfer is zero because "the final state is the same as the initial state," we also found that sign errors on those problems were much more frequent than on problem 12. We found in the interviews that students frequently struggled to analyze the segment-by-segment details of the (clockwise or counterclockwise) processes, potentially making them much more liable to commit sign errors along the way. Our results on problem 12, consistent with findings on student difficulties previously reported [26], suggest that the diagram showing two non-cyclic processes sharing the same initial and final states is itself a particularly powerful cue for the "Q is independent of process" line of thinking, thus limiting the number of sign errors observed on that problem.

In an interview, one student who answered $Q = 0$ for the cycle in problem 26 explained, "But you're returning to the same state, so there would be no change [no heat transfer] because you are finishing where you started." This student treated heat transfer as though it is a state variable. On problem 12 (Fig. 17), another student stated, "I think that heat transfer should be equal because they have the same starting and ending point and then it's just a difference of pressure and volume." (Another interviewee echoed the "same starting and ending point" argument.) This student was able to recognize that the processes had different paths but did not realize that heat transfer is dependent on the path and so must differ in the two cases. We note that the interviews reported in [26] indicated that explanations for $Q_1 = Q_2$ responses on Problem 12 were almost universally consistent with "heat is independent of path," increasing the likelihood that this idea dominates students' thinking on this problem. Sign errors, by contrast, could occur even when students employed an otherwise sound reasoning strategy. For example, a student who incorrectly said that there was net heat transfer *to* the gas ($Q > 0$) on problem 9 explained, "Net heat transfer to the gas…the work is being done on the system. If it's positive, then the net heat transfer is the opposite sign from the work." This student appears to have attempted to use the first law of thermodynamics, reasoning from $\Delta E = Q - W = 0$, but arrived at $Q = -W$ instead of $Q = W$.

It was well known from previous research that both Problem 12 and cyclic process problems strongly cued "Q is independent of path" responses. This suggests that processes with the same initial and final state, or two processes sharing identical initial and final states, are both salient distracting features cuing incorrect responses consistent with Q being independent of process. However, our finding that Problem 12 cued this response far more strongly for Int-calc students than did cyclic process problems is new and merits additional investigation.

Summary of Sec. III C: We found significant problem-property dependence in student responses on problems related to heat transfer concepts, with some hints of the possible origins of this dependence but no conclusive evidence. For example, we found a much higher likelihood of introductory students claiming that heat transfer would be zero in a "reversible isothermal compression" than they did for an "isothermal expansion." This suggests that either or both terms "reversible" and "compression" (when contrasted with "expansion") may act as salient distracting features and cue unproductive thinking related to the first law of thermodynamics in these problems. We found very similar response rates on problems involving heat transfer in cyclic processes represented as both clockwise and *counter*clockwise on a PV diagram, but—at least for the

Int-calc group—much poorer performance on a problem involving two noncyclic processes represented on a PV diagram as sharing identical initial and final states. The latter (problem 12) scenario appears to be a particularly salient cue among the calculus-based students for the incorrect idea that "Q is independent of path," previously identified in the literature as a common student difficulty [26], while the cyclic process scenario appears to induce a large number of sign errors, in addition to the very common $Q = 0$ response. In any case, it is evident that "same initial and final states" is itself a primary cue for thinking that Q is path independent, as has been previously reported.

D. Problem-property dependence of student responses to entropy problems

The problem groups found in Figs. 18–20 focus on concepts related to entropy, S, in different settings: (i) entropy changes "of the universe" during various spontaneous and irreversible processes (for the universe, $\Delta S > 0$), (ii) entropy changes in cyclic processes (entropy is a state variable so for the system, $\Delta S = 0$ in a full cycle), and (iii) entropy changes for a system undergoing a reversible isothermal expansion or compression (for the system, $\Delta S > 0$ and $\Delta S < 0$, respectively).

1. Entropy of the universe increases in a spontaneous and irreversible process

Figure 18 focuses on student responses to problems 17, 21, 53, 67, and 75, all relating to the concept that for the

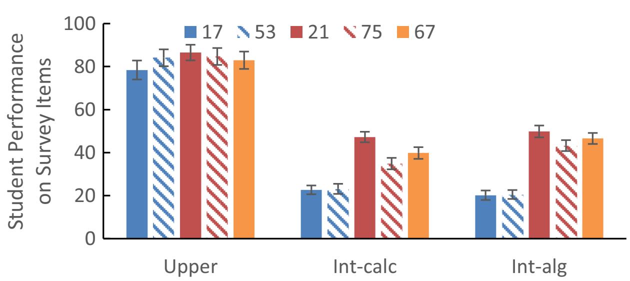

FIG. 18. Percentages of correct responses given by upper-level (Upper), introductory calculus-based (Int-calc), and introductory algebra-based (Int-alg) physics students on problems related to changes in entropy of the universe for a spontaneous and irreversible process, including problems 17 (heat transfer between two solids in contact), 53 (heat transfer between two gases in contact), 21 (free expansion), 75 (mixing of two non-interacting gases) and 67 (irreversible isochoric process). Responses for the upper-level students are highly consistent with each other, indicating very little problem-property dependence, consistent with the uniformly high correct-response rates. Correct-response rates for the Int-calc and Int-alg student groups show significant problem-property dependence, as the heat transfer problems 17 and 53 were significantly more challenging than the other three problems, although correct-response rates were 50% or lower on all items for these groups.

universe, $\Delta S > 0$ for a spontaneous and irreversible process. Problem 67 involves an isochoric process with net heat transfer to an ideal gas in contact with a thermal reservoir; it is explicitly identified as an "irreversible" process in the problem statement, the only one of the five so identified, and so the entropy of the system + reservoir must increase. The other four problems involve, in effect, combined two-component systems that are themselves isolated from the rest of the universe, undergoing spontaneous and irreversible processes. Students are asked whether the entropy of the combined system increases, decreases, remains the same, or whether there is not enough information. Since irreversible processes occur in all of the combined systems and since the systems are isolated from the universe, the net entropy for the systems must increase. An incorrect "entropy remains the same" response would be consistent with the idea that $\Delta S = 0$ for the universe in a spontaneous and irreversible process.

Although four of these problems (17, 21, 53, and 75) relate to spontaneous and irreversible processes occurring within rigid, thermally isolated containers, they have different surface features that can act as distractors for students. The settings for problems 21 and 75 can be viewed as similar to each other, but different from another "similar" pair, problems 17 and 53. Problem 21 involves a free expansion with an ideal gas initially in one chamber expanding into a vacuum when the stopcock is opened, while problem 75 involves two different noninteracting ideal gases initially in separate chambers, then allowed to mix with each other. The mixing setting of problem 75 could be viewed as somewhat analogous to the free expansion in problem 21, thus making 21 and 75 similar; the results indicate that both upper-level and Int-alg students did in fact view them that way; see discussion below. Both processes are irreversible because there is no way to reverse the process by making an infinitesimal change in system parameters.

Problems 17 and 53 can also be viewed as similar (even more so than 21 and 75). Problems 17 and 53 both involve two subsystems at very different temperatures in contact with each other, with heat transfer occurring between them (two solids at different temperatures in thermal contact in problem 17 and two gases at different temperatures in thermal contact in problem 53). Therefore, problems 17 and 53 share an identical element—heat transfer between the hot and cold subsystems—and thus could be seen as even more similar to each other than is the 21/75 pair. (Both processes are irreversible because the temperature difference between the interacting subsystems is large—indicated as $T_C \ll T_H$—and not infinitesimal.)

The performance of the three student groups on the five problems is represented in Fig. 18 by the percentage of students who correctly answered $\Delta S > 0$. Figure 18 shows that most of the upper-level students provided the correct response regardless of the setting of the problem, while there appears to be a significant problem-property dependence in responses for both introductory groups, with a large fraction of the introductory students not providing correct responses for each of the problems. Specifically, upper-level students are fairly consistent in their correct-response rates for all five problems, with rates ranging between 78% and 87%. This suggests that most upper-level students can consistently and correctly invoke and apply their knowledge regarding $\Delta S_{\mathrm{universe}} > 0$ across different spontaneous and irreversible process settings. Thus, we will now focus mainly on introductory student responses to these different problems.

Figure 18 shows that introductory students struggled the most on problems 53 and 17, as less than 25% of them provided correct responses to either of those problems; this was at the level of random guessing. (As noted above, these two problems both involve spontaneous heat transfer from a hot substance to a cold substance; the substances were solids in problem 17 and gases in problem 53.) In fact, the most popular response on these two problems—given by two-thirds of the introductory students—was that the entropy of the combined system does not change, implying $\Delta S = 0$ for the universe. Interviews corroborate these findings. For example, on problem 17, one student who thought $\Delta S = 0$ said, "Since there was no loss [of heat] to the environment, we will assume that the entropy has not changed." On the same problem, another student said, "The entropy of the combined system of the two solids has, I feel like no change, kinda like conservation. Cause while one's increasing, the other one's decreasing." This conservation argument appeared frequently in the interviews. It seems that the heat-transfer setting offers special challenges to introductory students in that the inherent nature of heat transfer as an *entropy-increasing* process is often weakly or not at all understood. Indeed, in this setting, most students did not realize that entropy would change at all, let alone that it would increase.

(It may be worth pointing out here that introductory textbooks often underemphasize the distinction between *reversible* processes involving heat transfer to or from a thermal reservoir due to infinitesimal temperature differences—such as isothermal expansions and compressions depicted on PV diagrams—and *irreversible* heat-transfer processes involving large temperature differences between interacting subsystems. We return to this idea in the Summary section below.)

While there are apparent "volume change" similarities linking the expansion or mixing problems in the 21/75 pair to each other and a "heat transfer" element connecting the two problems in the 17/53 pair, the only feature that the 17/53 pair truly shares with the 21/75 pair is that all four problems deal with irreversible processes. That is, all four problems involve isolated systems undergoing spontaneous and irreversible processes leading to $\Delta S > 0$ for the universe. One might then expect similar student outcomes for the 21/75 pair on the one hand, and for the 17/53 pair on the other, but not necessarily similar results when comparing the two pairs. In fact, that is exactly what is found: correct-response rates (as well as rates for the *incorrect* "increased" and "not changed" responses) for the 17/53 pair are almost identical, but very different from the 21/75 pair, which had higher correct response rates. While the 21/75 pair had lower error rates than 17/53, the results for 21 and 75 were not as similar to each other as those in the 17/53 pair that share the "heat transfer" element. As for why error rates in 17/53 were higher than the other pair, interview evidence suggests that when contemplating entropy changes, introductory students may view free expansion as a more salient cue than heat flow.

Figure 18 shows that for both introductory groups, the correct response rates on problem 21 (free expansion) are only slightly higher than those on problem 75 (gas mixing), although both are far higher than the corresponding rates for the 17/53 pair. However, there is a large difference between these two problems in the popularity of the incorrect $\Delta S = 0$ response. While 47% and 43% of the Int-calc and Int-alg groups, respectively, responded $\Delta S = 0$ for the mixing process (problem 75), the corresponding percentages for the free expansion process (problem 21) are only 29% and 27%. On that free-expansion problem, the response that entropy would *decrease* was nearly as popular as that it would remain unchanged. Thus, the free-expansion setting is seen to trigger incorrect "entropy decreases" and "entropy is constant" responses with almost equal frequency, while the other problem settings preferentially yield the "entropy is constant" response.

Correct-response rates for problem 67, the isochoric process explicitly identified as irreversible, fell between those for problems 21 and 75 for the introductory groups, indicating that even this problem was quite challenging for them. That is noteworthy since the irreversible nature of this process was not in question, while students did have to make that inference on their own from the problem descriptions provided for the other problems.

As mentioned above, the upper-level students performed significantly better overall than the introductory groups and were more consistent in their responses. Introductory students' responses were dependent on problem properties, with mixing and free-expansion problems having higher correct-response rates than heat-transfer problems. The correct-response rates of the two introductory groups for all problems related to entropy changes in irreversible processes are similar, and all are in the 20%–50% range. While student difficulties in some of these problem settings have not been investigated before, our other findings are consistent with previous student difficulties research that focused on only one or two particular problem settings (not several at a time) [28,29].

The difference in correct-response rates for entropy problems posed in different irreversible-process settings suggests that many students did not discern the deep similarity between these settings, that is, (i) that they were irreversible because infinitesimal changes in system parameters could not force the processes to run in reverse, and (ii) that $\Delta S_{\mathrm{universe}} > 0$ for all such spontaneous and irreversible processes

2. $\Delta S = 0$ for a system undergoing a cyclic process

The group in Fig. 19 includes problems 8 and 24, both of which incorporate PV diagrams showing gases undergoing a complete cyclic process (8: counterclockwise; 24: clockwise, also described as "reversible") and ask whether the final entropy "of the gas" is greater than, equal to, or less than the initial entropy, or whether there is not enough information. During a thermodynamic process, regardless of how a gas gets from its initial state to its final state, only the two end points determine the change in entropy since entropy is a state variable. In a cyclic process after one complete cycle, the gas ends up in the same state that it started in, $\Delta S = 0$. Correct-response rates on both of these cyclic problems were relatively low for all three student groups—no higher than 67%—and it is notable that upper-level students had lower correct-response rates than introductory students, although the difference was only significant for the Int-alg group on problem 8. The Int-calc group did better on problem 24 than on problem 8 ($p < 0.01$), perhaps due to the presence of the word "reversible," but the differences for the other two groups were small.

The fact that upper-level performances on both these cyclic problem settings are no better than—and occasionally actually worse than—some of those of the introductory students suggests that significant confusion persists in the upper-level courses regarding the state-function property of entropy. This is a cause for concern and one that upper-level thermodynamics instructors should take into consideration.

The most common incorrect response for all groups was that system entropy increases after a complete cycle; this response was substantially more common among the upper-level students on both problems 8 and 24 (36% and 38%) than it was in the introductory groups (25% and 27% for Int-calc, $p < 0.05$; 19% and 22% for Int-alg, $p < 0.01$). This was echoed in interviews with upper-level students. For example, in problem 24, one upper-level student said, "is heat in or heat out related to the entropy?" and then tried to determine the heat transferred into the system in each step of the cycle to find the change in entropy. It is true that in a reversible clockwise cycle, there is a net heat transfer to the system, but stepwise application of $\Delta S_{\mathrm{cycle}} = \Sigma Q_{\mathrm{rev}}/T$ still necessarily yields $\Delta S_{\mathrm{cycle}} = 0$, as is more easily determined by the identity of final and initial states and the state-variable nature of entropy. This upper-level student appeared to be distracted by the details of the individual subprocesses, not recognizing that entropy is a state variable and is therefore unchanged in a cyclic process. The fact that the upper-level students found both cyclic process problems to be almost equally challenging suggests that in this case, it was not, apparently, a problem feature that was distracting them, but instead the strength of the "entropy increases" idea. (Note that the problems very specifically asked about the entropy *of the gas*, not of the universe or some combined system.)

3. The sign of ΔS for a system undergoing a reversible isothermal process is determined by whether the volume is increasing or decreasing

The problem group in Fig. 20 includes problems 5 and 63; both of these deal with ideal monatomic gas processes. The setting of problem 5 is described in the problem statement as a "reversible isothermal compression," while that of problem 63 is a "reversible isothermal expansion"; neither problem includes a diagram. Students are asked in both problems whether the entropy of the gas will increase, decrease, or remain the same. Here one may apply $Q = W$ (since $Q - W = \Delta E = 0$) and $\Delta S = Q_{\mathrm{rev}}/T$ to realize that the entropy of the gas will increase when Q is positive

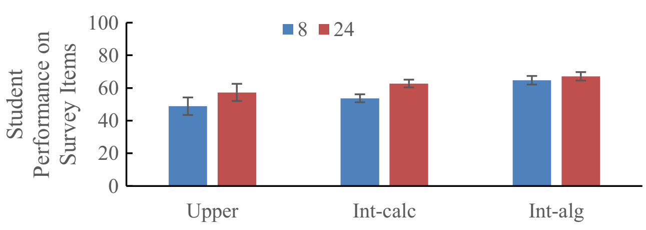

FIG. 19. Percentages of correct responses given by upper-level (Upper), introductory calculus-based (Int-calc), and introductory algebra-based (Int-alg) physics students on problems related to entropy changes in a cyclic process. Response rates for problems 8 ("counterclockwise") and 24 ("clockwise") are shown. The correct-response rates for these two problems are similar within each student group, indicating only weak problem-property dependence. Notably, correct-response rates of upper-level students were slightly lower than those of the introductory students.

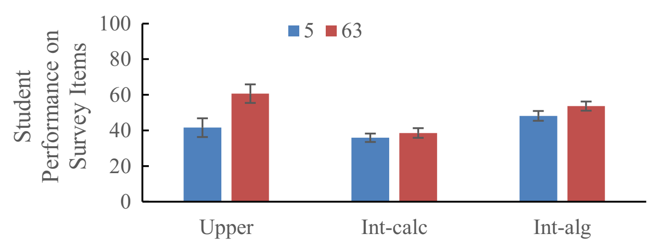

FIG. 20. Percentages of correct responses given by upper-level (Upper), introductory calculus-based (Int-calc), and introductory algebra-based (Int-alg) physics students on problems related to the changes in entropy of a system undergoing a reversible isothermal expansion (63) or compression (5). Correct-response rates for introductory students on these two problems are similar, thus showing only weak problem-property dependence. Although upper-level students were significantly more successful on the expansion problem setting, this may be fortuitous; see text.

(expansion) and decrease when Q is negative (compression). Alternatively, if one knows that the entropy of an ideal gas increases either with increasing volume or increasing internal energy—and internal energy does not change in an isothermal process—one can arrive at the same results. Once again, Fig. 20 shows that the correct-response rates for all student groups were quite low (maximum 61%), with sign errors and $\Delta S = 0$ errors both appearing frequently for all student groups. More than a third of the Int-calc students gave $\Delta S = 0$ responses on both problems 5 and 63 (33% and 38%), and nearly a quarter of the Int-alg students did so as well (22% and 23%). This response was also quite popular with the upper-level students (31% and 35%). During the interviews, three students explicitly cited the "reversible" nature of the process as the reason that entropy would not change, evidently applying to a *system* something that is only true for the *universe* as a whole.

There were small ($< 10\%$) differences in the proportions of $\Delta S = 0$ responses and sign-error responses for the introductory students between the two settings (expansion and compression); however, only the sign-error difference for the Int-calc students was statistically significant (fewer sign errors on problem 63, the expansion). By contrast, for the upper-level students, there was a *very* low 4% rate of sign errors on problem 63 (compared to 26% on problem 5). We suspect that this low rate of sign errors may merely be attributable to the popularity among the upper-level students of the incorrect idea that "entropy always increases," regardless of the thermodynamic process or whether the entropy of the *system* or the *universe* is being considered [29]. For the isothermal expansion in problem 63, "entropy increases" happens to be the correct answer, but it is incorrect for the compression in problem 5. It is interesting that Bucy, Thompson, and Mountcastle reported asking a small ($N = 7$) group of upper-level students about entropy changes in an isothermal expansion and found that all were able to give the correct positive sign [36]. However, to our knowledge, there is no prior research regarding student thinking about ΔS in a reversible isothermal *compression*.

Summary of Sec. III D: We found several primary problem-property dependences for introductory student groups' responses to problems involving entropy. For example, in irreversible processes, students were much more likely to struggle in recognizing that entropy would increase in a heat-transfer setting (transfer of thermal energy from hotter to colder objects) than they were for processes involving expansion or mixing of gases. Introductory students were frequently inclined to incorrectly assert a "conservation of entropy" argument (that is, that entropy of the system or universe would not change) even in settings that involved an entropy increase, while upper-level students tended to do the opposite, that is, assert that entropy would increase even in settings in which that was not true (e.g., entropy of the *system* in a complete cyclic process or in a reversible isothermal compression). It appears that this latter finding is indicative of increasing awareness of entropy increase (in the universe) as students pursue their physics studies, an awareness that, however, takes on a life of its own and tends to mislead many students into thinking that entropy increases even in cases in which it does not. The upper-level student response after traditional instruction appears to represent an overcorrection of a difficulty identified in previous studies [28], that is, the tendency for introductory students to resort inappropriately to "entropy conservation" lines of reasoning. Our findings appear to represent the first solid evidence of this overcorrection phenomenon, although it certainly was alluded to by Loverude [73] and noted by Crossette *et al.* [74].

We found it striking that in interviews both introductory and upper-level students, with very few exceptions, essentially ignored the critical distinction between entropy changes of the *universe* and entropy changes of a *system* contained within that universe, consistent with findings reported in Ref. [28]. This fact in itself helps explain the generally low correct-response rates on several of the problems and may also be responsible for some of the problem-property dependence that we observed. It is difficult to confirm that latter conjecture since the interview subjects so rarely addressed the system or universe distinction in any manner at all.

On problems involving ΔS for a system undergoing reversible isothermal expansion or compression, problem-property dependence of responses was observed primarily for upper-level students (Fig. 20), who performed better on the reversible isothermal expansion for which $\Delta S > 0$ than on the compression for which $\Delta S < 0$. However, we suspect strongly that this higher correct-response rate is not due to better reasoning but is instead a by-product of upper-level students' excessive inclination to adopt an "entropy always increases" line of reasoning.

IV. SUMMARY AND CONCLUSIONS

A. Overview

A deep and detailed understanding of the problem-property dependence of students' responses to diverse problems involving the same physics concept is an important aspect of improving the teaching and learning process in physics. To help students develop expertise, instructors must guide students to apply basic concepts to a wide variety of problems in diverse physical settings incorporating a multitude of surface features. Although physics principles and concepts may be easy to state succinctly, their practical application in realistic and varied problem situations is invariably challenging for most students. Unlike experts who have learned to discern deep features of problems and developed expertise in applying the same concept across different settings and scenarios, students must still travel a long road to reach that same point.

Practical application of important physics concepts in diverse physical settings can be enormously challenging to students at all levels. A well-known example from mechanics is Newton's third law, that interacting objects exert equal magnitude but oppositely directed forces on each other. Research shows that a wide variety of surface features in a problem may distract students from correctly applying this law, e.g., the relative masses or velocities of the objects, whether their acceleration or their velocity is zero or nonzero, etc. For most of the thermodynamic concepts and in most of the physical settings addressed in this investigation, a physics expert would find that invoking and applying appropriate knowledge to solve the problems would be quite straightforward. For many students at all levels investigated, however, this was not the case. In particular, we find that a variety of surface features in different settings can easily derail students' reasoning processes at both introductory and advanced levels, primarily because students are still developing expertise and have only a relatively limited grasp of the concepts involved. To inform instruction and to help develop effective curricular materials, it can be extremely helpful to acquire an understanding of which settings are more or less challenging and of how various problem properties may affect student thinking. The type of analysis conducted here is a substantial step toward that goal.

This study is, to our knowledge, the first systematic examination of problem-property dependence of both introductory and upper-level student responses to problems involving thermodynamics concepts. This examination has been made possible by the use of a validated survey instrument in which the same concepts are explored in various diverse settings across multiple problems. Of course, previous investigations have investigated specific issues related to problem properties, mostly in one or at most two situations, including their apparent influence on student difficulties. For the most part, however, these issues of problem-property dependence have at most only been addressed in an *ad hoc* manner, as they arise within the particular cases under investigation. For example, in the recent comprehensive review by Loverude of thermodynamics education research [29], specific issues of problem-property dependence are addressed only occasionally and mostly in passing. In our study, issues of problem-property dependence are addressed in a systematic manner for many of the problem types and physical processes most frequently employed in introductory undergraduate thermodynamics instruction. Our findings thus provide a valuable initial guide for instructors and curriculum developers, as well as for future researchers to investigate these issues further. However, we note that our findings are necessarily limited by the range of problems that we were able to incorporate in our survey; it will be up to future researchers to further sort out, confirm, and apply the findings discussed here in other settings for the same or similar concepts.

B. Findings on research questions

Regarding our RQ1, we find that the extent of problem-property dependence of students' responses to thermodynamics problems is often highly varied, occurring to different degrees for problems involving different variables or different processes. In general, based upon our findings, we state with confidence that the response patterns of students at all levels on thermodynamics problems can vary substantially from problem to problem depending on problem properties, even on problems whose solution pathways may all appear extremely similar to the expert. We observed in this investigation that slight differences in diagrams or wording can be associated with large variations in student responses, not all of which may be easily explainable without, for example, further interviews or a broader range of survey items. For some concepts, when error rates were high and students' responses were consistent across different problems, this very consistency suggested the presence of widespread common difficulties (e.g., the idea among upper-level students that entropy always increases).

Regarding RQ2 overall, we find that for most concepts, introductory students' responses to thermodynamics problems depended more heavily on the problem properties than the responses of upper-level students on the same problems. We would attribute this latter finding to the relatively higher level of expertise and a stronger grasp of concepts that upper-level students would have after instruction in comparison to introductory students. Our findings are consistent with the hypothesis that a better understanding of key concepts should lead to less variation in correct-response rates when problems are posed in diverse settings and scenarios with varied tasks. However, we also note that upper-level students often showed a clear problem-property dependence in their problem responses, indicating that increased content knowledge is not always adequate to guard against reasoning errors cued by salient distracting features. This suggests that a specific instructional focus on improving reasoning skills may be required even for upper-level physics students to fully address many of the difficulties carried over from their introductory studies.

Regarding our RQ3, we present below a summary of our specific findings, along with a discussion of some of their implications. As a reminder, here is RQ3:

RQ3: In cases where problem-property dependence exists, what specific problem properties can be identified either as "distracting" or "potentially productive" for one or more student groups and what is the nature of their influence?

1. Students were much more successful in determining the sign of ΔE (positive or negative) on a temperature-increasing isochoric process than on a temperature-decreasing isobaric process, even when provided with

a PV diagram on which initial and final temperatures for each process were explicitly shown. Some interview evidence suggests that the horizontal line representing the isobaric process distracted students into engaging in unproductive analyses of work instead of focusing on the relevant temperature variable.

2. When comparing magnitudes of ΔE for temperature-increasing adiabatic and isochoric processes shown on a PV diagram with identical initial and final temperatures explicitly indicated, incorrect responses that ΔE was greater for the adiabatic process were twice as frequent among the introductory students compared to (also incorrect) responses that ΔE was greater for the isochoric process. Interview evidence suggests that the relative shapes of the process arrows played a major role in students' thinking, distracting students into engaging in unproductive analyses of work.
3. Students found it *easier* in heat-transfer settings, and *harder* in free-expansion and mixing settings, to apply the concept that $\Delta E = 0$ for an isolated system undergoing a spontaneous and irreversible process. This seems to be a consequence of the expansion and mixing settings misleading students into thinking that net work had been performed in those cases.
4. For a system undergoing a cyclic process, students' correct responses that $\Delta E = 0$ as well as their *incorrect* responses that $Q = 0$ were unaffected by whether the process was represented on a PV diagram as clockwise or counterclockwise.
5. Explicitly indicating that an isothermal expansion process was "reversible" seemed to somewhat increase algebra-based students' low success rate in recognizing that $\Delta E = 0$ for that process. It is conceivable that the term "reversible" somehow cues an idea of "no change" in some students.
6. Students used work arguments that generally neglected heat transfer to support incorrect responses that internal energy would increase *or* decrease in an isothermal expansion.
7. For an isothermal (constant energy) expansion process, incorrect answers that internal energy "increases" were much more common than "decreases." It may be that volume expansions preferentially cue "increases" even in cases when "no change" is correct.
8. Evidence from interviews and performance data on multiple problems regarding ΔE that use PV diagrams suggests that flawed reasoning regarding work—while ignoring heat transfer—is a common route to arriving at incorrect answers regarding internal energy.
9. Isochoric "vertical line" (zero-work) processes appear to act in some problem settings to reduce, and in other settings to enhance, the degree of unproductive distraction due to flawed reasoning regarding work. Some students (such as the Int-alg group) appeared to find the vertical-line representation helpful for some tasks (e.g., evaluating internal energy change) and harmful for others (e.g., evaluating work done by the system).
10. The tendency for the term "adiabatic" to function as a salient distractor *by itself* is exacerbated by the confusion many students have regarding the actual meaning of the term. For example, many of them associate adiabatic with "no energy change."
11. We confirmed previous reports that free expansions are particularly challenging settings for applying the work concept to spontaneous processes, while finding that—in contrast—students are better able to apply that concept to spontaneous processes in which container volume does not change.
12. Isobaric "horizontal-line" processes are a diagram feature that specifically seems to *help* algebra-based students reason through some work-related problems, whereas those processes do not appear to play that same helping role for calculus-based or upper-level students, at least not to the same degree.
13. Introductory students were much more likely to give a correct $W > 0$ response for an isothermal expansion problem than they were for an adiabatic expansion. This was primarily due to a higher probability of $W = 0$ responses on the adiabatic process problem. This finding is consistent with the hypothesis that the use of the term "adiabatic" can be, in itself, a salient distracting feature.
14. We find that many students clearly recognize the absence of heat transfer in an adiabatic process but then tend to ignore (or misconstrue) the effects of work in that process. However, when faced with problems involving isobaric processes—represented by horizontal lines on PV diagrams—students appear able to recognize the presence and effects of work more easily than in other settings yet are *more* prone to ignore the presence and effects of heat transfer.
15. Introductory students were far more likely to give the correct sign of heat transfer Q for an isothermal expansion process than for a "reversible isothermal compression." This was due to the far greater prevalence of $Q = 0$ responses on the compression problem. Both the presence of the word "reversible" and/or the compression vs expansion difference could be suspected of acting as salient distracting features in this setting.
16. We confirmed previous reports that students frequently assert $Q = 0$ for processes with the same initial and final state (that is, cyclic processes) and $Q_1 = Q_2$ for two different processes sharing identical initial and final states, However, our finding that

the two-process problem cued the "path-independent" response far more strongly for calculus-based students than did cyclic process problems is new and merits additional investigation.

17. Introductory students were much more likely to recognize that net entropy increases in gas-expansion or mixing processes than in heat-transfer processes. It seems that most introductory students do not recognize heat transfer as an inherently entropy-increasing process. It is notable that precisely the opposite trend in student responses appears to be the case for problems that focus on changes in internal energy rather than entropy; students had greater success on the heat-transfer processes for problems involving ΔE.
18. Upper-level students performed no better than—and occasionally actually worse than—introductory students on cyclic process problems asking whether system entropy increases, decreases, or remains constant. This suggests that significant confusion persists in the upper-level courses regarding the state-function property of entropy. These consistent between-group differences suggest that instructors should carefully contemplate the way that the state-function property of entropy is taught in courses at all levels.
19. Introductory students frequently asserted a "conservation of entropy" argument (that is, that entropy of the system or universe would not change) even in settings that involved an entropy increase, while upper-level students tended to do the opposite, that is, assert that entropy would increase even in settings in which that was not true (e.g., entropy of the *system* in a complete cyclic process or in a reversible isothermal compression). The upper-level student response after traditional instruction appears to represent an overcorrection of a difficulty identified in previous studies [28], that is, a tendency for introductory students to resort inappropriately to "entropy conservation" lines of reasoning. Both introductory and upper-level students tended to disregard or confuse the crucial issues of whether the *system* entropy or the entropy of the *universe* is under discussion, and whether the process is reversible or irreversible. (We note that analogous observations have been reported or referenced previously [68,73,74].)
20. On problems involving ΔS for a system undergoing reversible isothermal expansion or compression, upper-level students performed better on the reversible isothermal expansion (for which $\Delta S > 0$) than on the compression (for which $\Delta S < 0$). We suspect strongly that this higher correct-response rate is not due to better reasoning but is instead a by-product of upper-level students' excessive inclination to adopt an "entropy always increases" line of reasoning.
21. We find that cyclic processes in a variety of settings offer a special set of challenges to students' understanding. Although students were generally successful on cyclic process problems involving the concept of internal energy (a state variable), that success did not necessarily extend to cyclic process problems involving heat, work, or the state variable entropy. For example, our findings—consistent across diverse settings, as well as with some prior research on student difficulties—confirm that many students are firmly attached to the idea that net work done and net heat transferred must be zero in processes that begin and end in the same state. And yet, those same processes frequently resulted in incorrect $\Delta S \neq 0$ responses when asking about net entropy change. These processes were also often found to be difficult to interpret using PV diagrams, particularly when work calculations were involved.
22. It was extremely common for interview subjects to invoke and apply one or another technical term in a haphazard manner depending upon the problem setting when trying to explain their thinking on a specific problem related to a broader concept. Although it was often difficult to identify patterns in the students' use of terminology, it seemed clear that relatively minor problem properties often had a disproportionate influence on students' verbal explanations. This illustrates the pitfalls that poorly understood terminology can create in students' thermodynamic thinking and may often be ascribed to vague, imprecise, or overly succinct definitions provided in introductory courses.

Our observational evidence is too limited in most cases for us to propose precise cognitive mechanisms for the operation of the various problem properties identified above. Our findings are consistent with dual-process theories of reasoning in that many specific problem features appear to cue students along certain lines of "intuitive" thinking based on immediate impressions rather than deep deliberation. In cases when performance data and interview evidence allow us, we have noted these features; the term "adiabatic," and processes with common initial and final states are examples in this category. However, it seems equally clear from the interviews that certain other problem features—principally process arrows on PV diagrams—do not necessarily cue rapid, intuitive thinking, whether productive or unproductive. Instead, and to the contrary, those features tend to lead students along a lengthy path of unproductive attempts to analyze work and then to incorporate those analyses into first-law considerations. Extensive additional investigation would be necessary to categorize appropriately the various problem properties noted above.

C. Implications for instruction

Although many implications for instruction could be drawn from the list of findings above, we want to emphasize three of the more obvious ones:

1. In a broad array of problems involving PV diagrams, the precise shapes of the process arrows often triggered inappropriate and unproductive lines of reasoning regarding work, distracting students from the features of the problems most relevant to finding a solution. Instructors should be cognizant of this potential distraction and consider addressing it by having students explore problems employing a variety of PV diagrams and guiding students to focus on the most relevant features while consciously avoiding dead-end pathways specifically involving work.
2. Instructors should be aware that certain terms may simply by their very presence trigger students into unanticipated and potentially unproductive lines of thinking. These terms include, among others, "adiabatic" and "reversible." When those terms are not directly relevant to a problem, instructors and curriculum developers should exercise caution when including them in problem descriptions. By contrast, terms such as "expansion" might in some problem settings cue productive lines of reasoning, and as such may sometimes be useful in helping students learn to apply the concepts they have studied.
3. In thermodynamics, students are particularly prone to focus undue attention on the "most salient" variable (such as heat in an adiabatic process or work in an isobaric process) while ignoring other variables that are essential to solving the problem. Instructors might address this tendency by explicitly requiring students to consider the potential relevance of multiple thermodynamic variables in each problem.

D. Conclusion

A lack of opportunity to develop a robust knowledge structure can prevent students at all levels—even upper-level students—from successfully and consistently solving conceptual thermodynamics problems in diverse settings. Students' responses on the survey problems discussed here show a strong problem-property dependence for some concepts and consistently poor performance across problems for other concepts, suggesting inadequate mastery of some basic thermodynamic concepts taught in introductory courses. One major goal of both introductory and upper-level thermodynamics instruction should be to help students construct robust schemas, *explicitly* connecting key concepts to each other and to the varied ways they are manifested in different thermodynamic processes and settings. This type of well-organized knowledge structure can assist students in correctly invoking and applying different thermodynamic concepts learned in a variety of problem settings, including different types of processes and different patterns of variation of thermodynamic variables, in diverse physical scenarios represented with various types of diagrams. It is important for instructors at both introductory and advanced levels to assess their students' conceptual knowledge and build on it appropriately, to help students organize, extend, and repair their knowledge structure so that surface features and contextual cues of the problems (such as those discussed above) do not derail their problem solving. This requires, however, both knowledge of and attention to the specific problem features (including diagrammatic characteristics and details of wording) that can influence students' thinking. The significance of this study for instructors and researchers is that it documents in detail how different specific features in particular thermodynamics problems can cue very different responses from students at all levels, even when the underlying concepts are similar or identical. This is a particular issue for thermodynamics since it involves a wide variety of systems, processes, and variables.

In conclusion, the development of a robust knowledge structure is necessary to enable students to invoke and apply their thermodynamics knowledge appropriately in diverse settings. Knowledge of how students think about specific thermodynamic variables and processes, how consistently they do so after instruction at each level, and the specific problem features that can influence students' thinking can all be valuable in developing effective curricula and assessments for this topic. Effective assessment relies upon a realistic understanding of how—and how well—responses on specific assessment items reflect students' conceptual understanding; effective instruction depends on addressing the issues identified by the assessments. Research-based and research-validated assessments, curricula, and pedagogical approaches are therefore needed to help students more efficiently develop a functional understanding of basic concepts. Such materials should be sensitive to students' confusion arising from specific problem features and explicitly address that confusion, to enable students to recognize similarities between different problems probing the same concept across different settings and scenarios. The ultimate goal is for students to be able to invoke and apply their knowledge appropriately across diverse settings in a consistent and reliable manner.

ACKNOWLEDGMENTS

We are grateful to one of our reviewers whose exceptionally thoughtful and detailed comments have helped us make substantial improvements to the manuscript. We thank all of the students who helped in this research.

[1] A. Newell and H. A. Simon, *Human Problem Solving* (Prentice-Hall, Englewood Cliffs, NJ, 1972).
[2] F. Reif and J. I. Heller, Knowledge structure and problem solving in physics, Educ. Psychol. **17**, 102 (1982).
[3] C. Singh, Assessing student expertise in introductory physics with isomorphic problems. I. Performance on nonintuitive problem pair from introductory physics, Phys. Rev. ST Phys. Educ. Res. **4**, 010104 (2008).
[4] C. Singh, Assessing student expertise in introductory physics with isomorphic problems. II. Effect of some potential factors on problem solving and transfer, Phys. Rev. ST Phys. Educ. Res. **4**, 010105 (2008).
[5] H. A. Simon and J. R. Hayes, The understanding process: Problem isomorphs, Cogn. Psychol. **8**, 165 (1976).
[6] L. R. Novick, Analogical transfer, problem similarity, and expertise, J. Exp. Psychol. **14**, 510 (1988).
[7] J. Lobato, Alternative perspectives on the transfer of learning: History, issues, and challenges for future research, J. Learn. Sci. **15**, 431 (2006).
[8] D. Schwartz, J. Bransford, and D. Sears, Efficiency and innovation in transfer, in *Transfer of Learning from a Modern Multidisciplinary Perspective*, edited by J. Mestre (Information Age Publishing, Greenwich, CT, 2005), pp. 1–51.
[9] T. Sinha and M. Kapur, When problem solving followed by instruction works: Evidence for productive failure, Rev. Educ. Res. **91**, 761 (2021).
[10] M. T. H. Chi, P. J. Feltovich, and R. Glaser, Categorization and representation of physics problems by experts and novices, Cogn. Sci. **5**, 121 (1981).
[11] S.-Y. Lin and C. Singh, Categorization of quantum mechanics problems by professors and students, Eur. J. Phys. **31**, 57 (2010).
[12] D. L. McBride, D. Zollman, and N. S. Rebello, Method for analyzing students' utilization of prior physics learning in new contexts, Phys. Rev. ST Phys. Educ. Res. **6**, 020101 (2010).
[13] N. S. Rebello *et al.*, Transfer of learning in problem solving in the context of mathematics and physics, in *Learning to Solve Complex Scientific Problems*, edited by D. H. Jonassen (Routledge, 2017), pp. 223–246.
[14] S.-Y. Lin and C. Singh, Using an isomorphic problem pair to learn introductory physics: Transferring from a two-step problem to a three-step problem, Phys. Rev. ST Phys. Educ. Res. **9**, 020114 (2013).
[15] S.-Y. Lin and C. Singh, Using isomorphic problems to learn introductory physics, Phys. Rev. ST Phys. Educ. Res. **7**, 020104 (2011).
[16] K. Kotovsky, J. R. Hayes, and H. A. Simon, Why are some problems hard? Evidence from Tower of Hanoi, Cogn. Psychol. **17**, 248 (1985).
[17] R. J. Dufresne, W. J. Gerace, P. Thibodeau Hardiman, and J. P. Mestre, Constraining novices to perform expertlike problem analyses: Effects on schema acquisition, J. Learn. Sci. **2**, 307 (1992).
[18] B. Eylon and F. Reif, Effects of knowledge organization on task performance, Cognit. Instr. **1**, 5 (1984).
[19] F. Reif, Millikan Lecture 1994: Understanding and teaching important scientific thought processes, Am. J. Phys. **63**, 17 (1995).
[20] S.-Y. Lin and C. Singh, Effect of scaffolding on helping introductory physics students solve quantitative problems involving strong alternative conceptions, Phys. Rev. ST Phys. Educ. Res. **11**, 020105 (2015).
[21] S.-Y. Lin and C. Singh, Challenges in using analogies, Phys. Teach. **49**, 512 (2011).
[22] M. Bassok and K. J. Holyoak, Interdomain transfer between isomorphic topics in algebra and physics, J. Exp. Psychol. **15**, 153 (1989).
[23] K. J. Holyoak and P. Thagard, *Mental Leaps: Analogy in Creative Thought* (The MIT Press, Cambridge, MA 1995).
[24] D. Gentner and C. Toupin, Systematicity and surface similarity in the development of analogy, Cogn. Sci. **10**, 277 (1986).
[25] T. I. Smith, W. M. Christensen, D. B. Mountcastle, and J. R. Thompson, Identifying student difficulties with entropy, heat engines, and the Carnot cycle, Phys. Rev. ST Phys. Educ. Res. **11**, 020116 (2015).
[26] D. E. Meltzer, Investigation of students' reasoning regarding heat, work, and the first law of thermodynamics in an introductory calculus-based general physics course, Am. J. Phys. **72**, 1432 (2004).
[27] B. W. Dreyfus, B. D. Geller, D. E. Meltzer, and V. Sawtelle, Resource Letter TTSM-1: Teaching thermodynamics and statistical mechanics in introductory physics, chemistry, and biology, Am. J. Phys. **83**, 5 (2015).
[28] W. M. Christensen, D. E. Meltzer, and C. A. Ogilvie, Student ideas regarding entropy and the second law of thermodynamics in an introductory physics course, Am. J. Phys. **77**, 907 (2009).
[29] M. E. Loverude, Student understanding of thermal physics, in *The International Handbook of Physics Education Research: Learning Physics*, edited by M. F. Taşar and P. R. L. Heron (AIP Publishing (online), Melville, NY, 2023), pp. 3.1–3.38.
[30] M. J. Cochran and P. R. L. Heron, Development and assessment of research-based tutorials on heat engines and the second law of thermodynamics, Am. J. Phys. **74**, 734 (2006).
[31] B. Brown and C. Singh, Development and validation of a conceptual survey instrument to evaluate students' understanding of thermodynamics, Phys. Rev. Phys. Educ. Res. **17**, 010104 (2021).
[32] J. M. Bennett and M. Sözbilir, A study of Turkish chemistry undergraduates' understanding of entropy, J. Chem. Educ. **84**, 1204 (2007).
[33] B. Brown and C. Singh, Student understanding of the first law and second law of thermodynamics, Eur. J. Phys. **42**, 065702 (2021).
[34] H. Georgiou and M. D. Sharma, Does using active learning in thermodynamics lectures improve students' conceptual understanding and learning experiences?, Eur. J. Phys. **36**, 015020 (2014).
[35] K. Bain, A. Moon, M. R. Mack, and M. H. Towns, A review of research on the teaching and learning of thermodynamics at the university level, Chem. Educ. Res. Pract. **15**, 320 (2014).
[36] B. R. Bucy, J. R. Thompson, and D. B. Mountcastle, What is entropy? Advanced undergraduate performance comparing ideal gas processes, AIP Conf. Proc. **818**, 81 (2006).

[37] H.-E. Chu, D. F. Treagust, S. Yeo, and M. Zadnik, Evaluation of students' understanding of thermal concepts in everyday contexts, Int. J. Sci. Educ. **34**, 1509 (2012).

[38] J. W. Clark, J. R. Thompson, and D. B. Mountcastle, Comparing student conceptual understanding of thermodynamics in physics and engineering, AIP Conf. Proc. **1513**, 102 (2013).

[39] H. Goldring and J. Osborne, Students' difficulties with energy and related concepts, Phys. Educ. **29**, 26 (1994).

[40] M. F. Granville, Student misconceptions in thermodynamics, J. Chem. Educ. **62**, 847 (1985).

[41] C. H. Kautz, P. R. L. Heron, M. E. Loverude, and L. C. McDermott, Student understanding of the ideal gas law, Part I: A macroscopic perspective, Am. J. Phys. **73**, 1055 (2005).

[42] C. H. Kautz and G. Schmitz, Probing student understanding of basic concepts and principles in introductory engineering thermodynamics, in *Proceedings of the ASME 2007 International Mechanical Engineering Congress and Exposition. Volume 6: Energy Systems: Analysis, Thermodynamics and Sustainability* (2007), pp. 473–480, 10.1115/IMECE2007-41863.

[43] E. Langbeheim, S. A. Safran, S. Livne, and E. Yerushalmi, Evolution in students' understanding of thermal physics with increasing complexity, Phys. Rev. ST Phys. Educ. Res. **9**, 020117 (2013).

[44] R. Leinonen, M. A. Asikainen, and P. E. Hirvonen, University students explaining adiabatic compression of an ideal gas—a new phenomenon in introductory thermal physics, Res. Sci. Educ. **42**, 1165 (2012).

[45] R. Leinonen, M. A. Asikainen, and P. E. Hirvonen, Overcoming students' misconceptions concerning thermal physics with the aid of hints and peer interaction during a lecture course, Phys. Rev. ST Phys. Educ. Res. **9**, 020112 (2013).

[46] R. Leinonen, E. Räsänen, M. Asikainen, and P. E Hirvonen, Students' pre-knowledge as a guideline in the teaching of introductory thermal physics at university, Eur. J. Phys. **30**, 593 (2009).

[47] M. E. Loverude, C. H. Kautz, and P. R. L. Heron, Student understanding of the first law of thermodynamics: Relating work to the adiabatic compression of an ideal gas, Am. J. Phys. **70**, 137 (2002).

[48] M. Malgieri, P. Onorato, A. Valentini, and A. De Ambrosis, Improving the connection between the microscopic and macroscopic approaches to thermodynamics in high school, Phys. Educ. **51**, 065010 (2016).

[49] D. E. Meltzer, Investigation of student learning in thermodynamics and implications for instruction in chemistry and engineering, AIP Conf. Proc. **883**, 38 (2007).

[50] D. E. Meltzer, Observations of general learning patterns in an upper-level thermal physics course, AIP Conf. Proc. **1179**, 31 (2009).

[51] K. C. Midkiff, T. A. Litzinger, and D. L. Evans, Development of engineering thermodynamics concept inventory instruments, in *Proceedings of the 31st Annual Frontiers in Education Conference. Impact on Engineering and Science Education, (Cat. No.01CH37193), Reno, NV, USA* (2001), Vol. 2, pp. F2A–F23, https://ieeexplore.ieee.org/document/963691.

[52] T. Nilsson and H. Niedderer, An analytical tool to determine undergraduate students' use of volume and pressure when describing expansion work and technical work, Chem. Educ. Res. Pract. **13**, 348 (2012).

[53] K. D. Rainey, M. Vignal, and B. R. Wilcox, Validation of a coupled, multiple response assessment for upper-division thermal physics, Phys. Rev. Phys. Educ. Res. **18**, 020116 (2022).

[54] T. I. Smith, D. Mountcastle, and J. Thompson, Student understanding of the Boltzmann factor, Phys. Rev. ST Phys. Educ. Res. **11**, 020123 (2015).

[55] T. I. Smith, D. B. Mountcastle, and J. R. Thompson, Identifying student difficulties with conflicting ideas in statistical mechanics, AIP Conf. Proc. **1513**, 386 (2013).

[56] R. Streveler *et al.*, Rigorous methodology for concept inventory development: Using the assessment triangle to develop and test the Thermal and Transport Science Concept Inventory (TTCI), Int. J. Eng. Educ. **27**, 968 (2011), https://www.ijee.ie/articles/Vol27-5/06_ijee2480ns.pdf.

[57] P. L. Thomas and R. W. Schwenz, College physical chemistry students' conceptions of equilibrium and fundamental thermodynamics, J. Res. Sci. Teach. **35**, 1151 (1998).

[58] A. Tongchai *et al.*, Consistency of students' conceptions of wave propagation: Findings from a conceptual survey in mechanical waves, Phys. Rev. ST Phys. Educ. Res. **7**, 020101 (2011).

[59] P. Wattanakasiwich *et al.*, Development and implementation of a conceptual survey in thermodynamics, Int. J. Innov. Sci. Math. Educ. **21**, 29 (2013), https://openjournals.library.sydney.edu.au/CAL/article/view/6459.

[60] S. Yeo and M. Zadnik, Introductory thermal concept evaluation: Assessing students' understanding, Phys. Teach. **39**, 496 (2001).

[61] B. Brown, Developing and assessing research-based tools for teaching quantum mechanics and thermodynamics, Ph.D. Dissertation, University of Pittsburgh, 2015.

[62] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.21.010127 for the STPFaSL-Long instrument.

[63] P. V. Engelhardt, An introduction to Classical Test Theory as applied to conceptual multiple-choice tests, in *Getting Started in PER*, edited by C. Henderson and K. Harper, Reviews in PER Vol. 2 (AAPT, College Park, MD, 2009).

[64] B. Brown and C. Singh, Student understanding of thermodynamic processes, variables and systems, Eur. J. Phys. **43**, 055705 (2022).

[65] M. J. Brundage, D. E. Meltzer, and C. Singh, Investigating introductory and advanced students' difficulties with change in internal energy, work, and heat transfer using a validated instrument, Phys. Rev. Phys. Educ. Res. **20**, 010115 (2024).

[66] C. R. Gette, M. Kryjevskaia, M. R. Stetzer, and P. R. L. Heron, Probing student reasoning approaches through the lens of dual-process theories: A case study in buoyancy, Phys. Rev. Phys. Educ. Res. **14**, 010113 (2018).

[67] J. C. Speirs, M. R. Stetzer, B. A. Lindsey, and M. Kryjevskaia, Exploring and supporting student reasoning in physics by leveraging dual-process theories of reasoning and decision making, Phys. Rev. Phys. Educ. Res. **17**, 020137 (2021).

[68] M. J. Brundage, D. E. Meltzer, and C. Singh, Investigating introductory and advanced students' difficulties with entropy and the second law of thermodynamics using a validated instrument, Phys. Rev. Phys. Educ. Res. **20**, 020110 (2024).

[69] M. J. Brundage, D. E. Meltzer, and C. Singh, Investigating context dependence of introductory and advanced student responses to introductory thermodynamics conceptual problems, presented at PER Conf. 2023, Sacramento, CA, edited by D. L. Jones, Q. X. Ryan, and A. Pawl, 10.1119/perc.2023.pr.Brundage, pp. 70–75.

[70] M. J. Brundage and C. Singh, Development and validation of a conceptual multiple-choice survey instrument to assess student understanding of introductory thermodynamics, Phys. Rev. Phys. Educ. Res. **19**, 020112 (2023).

[71] PhysPort, https://www.physport.org/.

[72] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*, (Lawrence Erlbaum Associates, Hillsdale, NJ, 1988).

[73] M. Loverude, Identifying student resources in reasoning about entropy and the approach to thermal equilibrium, Phys. Rev. ST Phys. Educ. Res. **11**, 020118 (2015).

[74] N. Crossette, M. Vignal, and B. R. Wilcox, Investigating graduate student reasoning on a conceptual energy questionnaire, Phys. Rev. Phys. Educ. Res. **17**, 020119 (2021).

SUPPLEMENTAL MATERIAL:
SAMPLE SIZES, P-VALUES, AND COHEN'S h VALUES FOR PERFORMANCE COMPARISONS

The following table shows the sample size of each of the three groups used in the statistical analysis of the questions discussed in the text.

Table II. The sample size for selected items on the STPFaSL-Long. For the survey items discussed in this paper, the total number of student responses is shown (N) for each group of students.

Item #	N	Level	Item #	N	Level	Item #	N	Level	Item #	N	Level
2	87	Upper	17	88	Upper	42	89	Upper	60	89	Upper
	416	Int-calc		419	Int-calc		490	Int-calc		323	Int-calc
	331	Int-alg		332	Int-alg		550	Int-alg		377	Int-alg
3	88	Upper	21	89	Upper	43	89	Upper	61	89	Upper
	414	Int-calc		419	Int-calc		491	Int-calc		323	Int-calc
	331	Int-alg		331	Int-alg		549	Int-alg		379	Int-alg
4	88	Upper	22	88	Upper	44	89	Upper	62	89	Upper
	416	Int-calc		417	Int-calc		490	Int-calc		323	Int-calc
	331	Int-alg		332	Int-alg		549	Int-alg		380	Int-alg
5	89	Upper	23	89	Upper	45	88	Upper	63	89	Upper
	415	Int-calc		419	Int-calc		487	Int-calc		324	Int-calc
	332	Int-alg		331	Int-alg		549	Int-alg		380	Int-alg
6	88	Upper	24	89	Upper	47	88	Upper	65	89	Upper
	415	Int-calc		416	Int-calc		490	Int-calc		321	Int-calc
	331	Int-alg		332	Int-alg		548	Int-alg		379	Int-alg
7	89	Upper	25	89	Upper	49	87	Upper	67	88	Upper
	418	Int-calc		418	Int-calc		324	Int-calc		324	Int-calc
	332	Int-alg		331	Int-alg		380	Int-alg		380	Int-alg
8	88	Upper	26	87	Upper	50	89	Upper	69	88	Upper
	417	Int-calc		419	Int-calc		324	Int-calc		323	Int-calc
	332	Int-alg		332	Int-alg		380	Int-alg		380	Int-alg
9	86	Upper	30	89	Upper	53	88	Upper	71	86	Upper
	415	Int-calc		490	Int-calc		324	Int-calc		324	Int-calc
	332	Int-alg		550	Int-alg		380	Int-alg		379	Int-alg
10	89	Upper	32	89	Upper	57	89	Upper	72	86	Upper
	418	Int-calc		489	Int-calc		323	Int-calc		324	Int-calc
	332	Int-alg		550	Int-alg		380	Int-alg		378	Int-alg
12	89	Upper	33	89	Upper	58	89	Upper	75	85	Upper
	418	Int-calc		492	Int-calc		324	Int-calc		321	Int-calc
	332	Int-alg		550	Int-alg		380	Int-alg		379	Int-alg
13	89	Upper	34	89	Upper	59	89	Upper			
	418	Int-calc		491	Int-calc		320	Int-calc			
	332	Int-alg		549	Int-alg		380	Int-alg			

The following tables show statistical comparisons for survey items on the STPFaSL-Long, specifically, the effect size and the p-value for differences in response rates on problem pairs. Each table shows comparisons for a figure in the main text. Comparisons are made between different items for the same group of students and for different groups of students on the same item. The three groups are upper-level physics students (U), calculus-based introductory physics students (C), and algebra-based introductory physics students (A). The number that appears after the group letter is the survey item number. For each comparison, the effect size in terms of Cohen's h is given along with the significance level of the p-value from a binomial proportions z-test (equivalent to a 2×2 Chi-square test). A z-test resulting in $p < 0.001$ is indicated by ***, while $0.001 \leq p < 0.01$ is indicated by ** and $0.01 \leq p < 0.05$ is indicated by *.

Table III: Comparisons for STPFaSL-Long items in Figure 5.

	U 33	U 34	U 69	C 33	C 34	C 69	A 33	A 34	A 69
U 33	…								
U 34	0.32 *	…							
U 69	0.52 ***	0.20	…						
C 33	0.19			…					
C 34		0.23 *		0.37 ***	…				
C 69			0.57 ***	0.90 ***	0.53 ***	…			
A 33	0.05			0.14 *			…		
A 34		0.30 **			0.06		0.57 ***	…	
A 69			0.73 ***			0.16 *	1.20 ***	0.63 ***	…

Table IV: Comparisons for STPFaSL-Long items in Figure 6.

	U13	U50	U23	U72	C13	C50	C23	C72	A13	A50	A23	A72
U13	…											
U50	0.21	…										
U23	0.54 ***	0.33 *	…									
U72	0.50 **	0.30 *	0.04	…								
C13	0.40 **				…							
C50		0.50 ***			0.31 ***	…						
C23			0.42 ***		0.56 ***	0.25 ***	…					
C72				0.39 **	0.50 ***	0.19 *	0.07	…				
A13	0.38 **				0.02				…			
A50		0.33 *				0.17 *			0.16 *	…		
A23			0.40 **				0.03		0.56 ***	0.40 ***	…	
A72				0.33 **				0.06	0.45 ***	0.30 ***	0.10	…

Table V: Comparisons for STPFaSL-Long items in Figure 7.

	U6	U25	U47	C6	C25	C47	A6	A25	A47
U6	…								
U25	0.16	…							
U47	0.13	0.03	…						
C6	0.24 *			…					
C25		0.27 *		0.19 **	…				
C47			0.31 *	0.21 **	0.02	…			
A6	0.25			0.01			…		
A25		0.34 **			0.07		0.25 **	…	
A47			0.30 *			0.01	0.18 **	0.07	…

Table VI: Comparisons for STPFaSL-Long items in Figure 8.

	U61	U65	U33	U44	U34	U45	U2	C61	C65	C33	C44	C34	C45	C2	A61	A65	A33	A44	A34	A45	A2
U61	…																				
U65	0.10	…																			
U33	0.44 **	0.34 *	…																		
U44	0.57 ***	0.47 **	0.13	…																	
U34	0.12	0.02	0.32 *	0.45 **	…																
U45	0.02	0.08	0.42 **	0.55 ***	0.10	…															
U2	0.20	0.30 *	0.64 ***	1.39 ***	0.32 *	0.22	…														
C61	0.32 **							…													
C65		0.41 ***						0.01	…												
C33			0.19					0.57 ***	0.57 ***	…											
C44				0.45 ***				0.43 ***	0.43 ***	0.14 *	…										
C34					0.23 *			0.20 **	0.20 *	0.37 ***	0.23 ***	…									
C45						0.45 ***		0.11	0.12	0.68 ***	0.54 ***	0.32 ***	…								
C2							0.50 ***	0.39 ***	0.39 ***	0.96 ***	0.82 ***	0.59 ***	0.27 ***	…							
A61	0.74 ***							0.42 ***							…						
A65		0.59 ***							0.18 *						0.24 **	…					
A33			0.05							0.14 *					1.13 ***	0.89 ***	…				
A44				0.45 ***							0.01				0.86 ***	0.62 ***	0.27 ***	…			
A34					0.30 **							0.06			0.56 ***	0.32 ***	0.57 ***	0.30 ***	…		
A45						0.56 ***							0.11 *		0.19 **	0.05	0.93 ***	0.66 ***	0.36 ***	…	
A2							0.62 ***							0.11	0.09	0.33 ***	1.21 ***	0.94 ***	0.64 ***	0.28 ***	…

Table VII: Comparisons for STPFaSL-Long items in Figure 10.

	U22	U49	U71	C22	C49	C71	A22	A49	A71
U22	…								
U49	0.63 ***	…							
U71	0.18	0.45 **	…						
C22	0.51 ***			…					
C49		0.43 **		0.71 ***	…				
C71			0.08	0.61 ***	0.10	…			
A22	0.47 ***			0.04			…		
A49		0.36 **			0.06		0.74 ***	…	
A71			0.01			0.07	0.64 ***	0.10	…

Table VIII: Comparisons for STPFaSL-Long items in Figure 11.

	U7	U10	U42	U43	U57	U58	U59	C7	C10	C42	C43	C57	C58	C59	A7	A10	A42	A43	A57	A58	A59
U7	…																				
U10	0.84 ***	…																			
U42	0.74 ***	0.10	…																		
U43	0.71 ***	0.13	0.03	…																	
U57	0.88 ***	0.03	0.13	0.16	…																
U58	0.81 ***	0.03	0.06	0.09	0.07	…															
U59	0.57 ***	0.27	0.17	0.14	0.30 *	0.24	…														
C7	0.09							…													
C10		0.40 **						0.53 ***	…												
C42			0.26 *					0.57 ***	0.04	…											
C43				0.23				0.58 ***	0.05	0.01	…										
C57					0.60 ***			0.37 ***	0.16 *	0.20 **	0.21 **	…									
C58						0.59 ***		0.30 ***	0.23 **	0.27 ***	0.27 ***	0.07	…								
C59							0.50 ***	0.16 *	0.36 ***	0.41 ***	0.41 ***	0.20 **	0.14	…							
A7	0.70 ***							0.62 ***							…						
A10		0.70 ***							0.30 ***						0.84 ***	…					
A42			0.84 ***							0.58 ***					0.61 ***	0.24 ***	…				
A43				0.53 ***							0.31 ***				0.89 ***	0.04	0.28 ***	…			
A57					0.85 ***							0.26 **			0.73 ***	0.11	0.12	0.16 *	…		
A58						0.94 ***							0.35 ***		0.57 ***	0.27 ***	0.03	0.31 ***	0.16 *	…	
A59							0.82 ***							0.32 ***	0.46 ***	0.38 ***	0.15 *	0.43 ***	0.27 ***	0.11	…

Table IX: Comparisons for STPFaSL-Long items in Figure 14.

	U3	U62	C3	C62	A3	A62
U3	…					
U62	0.14	…				
C3	0.45 ***		…			
C62		0.36 **	0.23 **	…		
A3	0.70 ***		0.25 ***		…	
A62		0.57 ***		0.21 **	0.27 ***	…

Table X: Comparisons for STPFaSL-Long items in Figure 15.

	U4	U32	U60	C4	C32	C60	A4	A32	A60
U4	…								
U32	0.38 *	…							
U60	0.16	0.23	…						
C4	0.79 ***			…					
C32		0.67 ***		0.51 ***	…				
C60			0.67 ***	0.28 ***	0.23 **	…			
A4	1.00 ***			0.21 **			…		
A32		0.75 ***			0.08		0.63 ***	…	
A60			0.66 ***			0.01	0.50 ***	0.13	…

Table XI: Comparisons for STPFaSL-Long items in Figure 16.

	U9	U12	U26	U30	C9	C12	C26	C30	A9	A12	A26	A30
U9	…											
U12	0.20	…										
U26	0.19	0.01	…									
U30	0.30	0.50 **	0.50 **	…								
C9	0.27 *				…							
C12		0.48 ***			0.41 ***	…						
C26			0.10		0.03	0.39 ***	…					
C30				0.44 ***	0.13	0.55 ***	0.16 *	…				
A9	0.85 ***				0.59 ***				…			
A12		0.71 ***				0.23 **			0.06	…		
A26			0.50 ***				0.40 ***		0.16 *	0.22 **	…	
A30				0.68 ***				0.25 ***	0.47 ***	0.53 ***	0.31 ***	…

Table XII: Comparisons for STPFaSL-Long items in Figure 18.

	U17	U53	U21	U75	U67	C17	C53	C21	C75	C67	A17	A53	A21	A75	A67
U17	…														
U53	0.15	…													
U21	0.21	0.07	…												
U75	0.16	0.02	0.05	…											
U67	0.12	0.03	0.10	0.05	…										
C17	1.18 ***					…									
C53		1.32 ***				0.01	…								
C21			0.87 ***			0.52 ***	0.51 ***	…							
C75				1.07 ***		0.27 ***	0.26 ***	0.25 **	…						
C67					0.92 ***	0.37 ***	0.36 ***	0.15	0.10	…					
A17	1.24 ***					0.06					…				
A53		1.38 ***					0.06				0.01	…			
A21			0.82 ***					0.05			0.64 ***	0.63 ***	…		
A75				0.90 ***					0.17 *		0.50 ***	0.50 ***	0.13	…	
A67					0.79 ***					0.14	0.57 ***	0.56 ***	0.07	0.07	…

Table XIII: Comparisons for STPFaSL-Long items in Figure 19.

	U8	U24	C8	C24	A8	A24
U8	…					
U24	0.17	…				
C8	0.10		…			
C24		0.11	0.18 **	…		
A8	0.32 **		0.23 **		…	
A24		0.20		0.09	0.05	…

Table XIV: Comparisons for STPFaSL-Long items in Figure 20.

	U5	U63	C5	C63	A5	A63
U5	…					
U63	0.38 *	…				
C5	0.12		…			
C63		0.45 ***	0.06	…		
A5	0.13		0.25 ***		…	
A63		0.14		0.30 ***	0.11	…