Investigating unprompted and prompted diagrams generated by physics majors during problem solving

Michael Vignal and Bethany R. Wilcox

Department of Physics, University of Colorado, 390 UCB, Boulder, Colorado 80309, USA

Diagrams are ubiquitous in physics, especially in physics education and physics problem solving. Physics problem solvers may generate diagrams to orient to a scenario, to organize information, to directly obtain an answer, or as a tool of communication. In this study, we asked 19 undergraduate and graduate physics majors to answer 18 multiple-choice (MC) physics questions and then complete six diagramming tasks of situations similar to six of the MC problems: the MC problems contained no prompting regarding diagrams, while the diagramming tasks explicitly asked participants to carefully generate diagrams. This prompting placed participants in one of two epistemic frames, problem solving or communicating, which allowed us to explore which elements and features the students include (or not) in diagrams generated when students are working within these two frames. By comparing students' spontaneously generated diagrams to their prompted diagrams, we found differences in size, accuracy, and amount of detail in unprompted problem-solving diagrams and prompted communicating diagrams. We also looked at correlations between the presence and features of unprompted diagrams with participants' answer choice. Looking at the different cohorts (e.g., lower-division undergraduate and graduate students), we found that the differences in diagramming between cohorts were generally smaller than the differences within a cohort. We also explore implications for teaching and research.

I. INTRODUCTION AND BACKGROUND

Diagrammatic representations of physical scenarios are ubiquitous in physics. Graphs and figures are used to communicate scientific information, such as in journal articles and textbooks, and graphs, sketches, and more specialized diagrams (e.g., free body diagrams, ray diagrams, etc.) can aid in physics problem solving. Given the variety and importance of diagrams and their uses in physics, it is not surprising that instruction around constructing diagrams is standard in physics education [1–4]. In particular, students are shown and expected to perform both general (e.g., graphs and sketches) and specific (e.g., free body diagrams, ray diagrams) techniques that may aid in solving physics problems.

Many physics education researchers have studied how students interact with diagrams. Most prevalent is research investigating how students interpret and use canonical and professional representations: such work has been done in the contexts of mechanics [5–9], electricity and magnetism [10,11], quantum mechanics [10,12], thermodynamics [13], as well as in chemistry [14], computational physics [15], and education more generally [16–18]. Researchers have also used eye-tracking software to better understand how students refer to diagrams given with the problem prompts [19]. Much of this work specifically looks at the ways in which students interpret and coordinate between multiple representations [5,6,8–14,17,20].

The majority of this work, at least in physics education research, has looked at how students use representations that are provided to them. In these studies, researchers are generally interested in student interpretation, manipulation, translation, and (often) coordination of diagrams. While all of these skills are important for physics problem solvers, by providing students with diagrams, these studies strongly influence when, how, and why students use diagrams, and therefore are limited in what they can tell us about student problem solving through the study of student use of diagrams.

Less common is research into student-generated diagrams [6–8,17,18,21–23]. Cox and Brna found that students interact differently with diagrams provided to them as compared to diagrams they generate themselves [17], and Cox argues (in a later article) that "the distinction between constructing ones own representation versus reasoning with someone elses is extremely important" [24]. Schmidgall et al. lay out three aspects of student-generated diagrams that improve learning (as compared to just reading text): generation, in which information from the problem is used

to produce a new representation; visualization, a particular form of generation in which students often must incorporate contextual information into a representation in order to make a spatially reasonable representation; and externalization, in which some information from the problem can be off loaded onto the representation [18]. While these benefits of generating diagrams have been well studied, this work has focused less on why students generate diagrams. When problem solving, one may generate diagrams for a variety of reasons, including to orient to a situation or problem [25,26]; to aid in the problem-solving process [6,8]; or as a tool of communication once a solution has been devised [7,27]. Since the reason for generating a diagram will influence how the diagram is used, research into student use of self-generated diagrams allows researchers to ask questions that are distinct from (though often related to) questions that can be asked when looking at student use of provided diagrams.

A few researchers have used student-generated diagrams as a lens to better understand student problem solving in physics [6–8,23]. Perhaps most similar to the current study is research done by Heckler comparing success and solution methods on simple mechanics problems for students who either did or did not get an explicit prompt to draw a force diagram. Heckler and others have found that drawing diagrams can increase student success with solving physics problems [6,7]. However, Heckler also found that explicitly prompting students to generate diagrams can lower student success, at least for simple problems in contexts familiar to students. Heckler suggests that the explicit diagramming prompt may direct students away from intuitive solution methods and towards methods with which students have less success (in particular, carefully constructing and then using a force diagram). Heckler also found that students who had received prompting were more likely to have errors in their diagrams than did students who drew diagrams without prompting.

Heckler's work further contributes to the notion that the reason for which students interact with diagrams impacts how they interact with those diagrams. For this reason, it is worth noting that his work, as with much of the work on student-generated diagrams, has looked at diagrams collected as part of homework or exams. Research has found that students select different representations depending on their reason for using them [15], and so it is reasonable to believe that these contexts (graded course work) may influence how (and how often) students create diagrams. In these contexts, students often generate professional-looking diagrams due to either explicit or implicit expectations, which likely influence the frequency and content of the diagrams.

It is worth acknowledging that some work has been done looking at student-generated representations outside of college physics. In particular, diSessa and Cobb found that students had "substantial expertise in inventing,

evaluating, and refining a variety of representational forms," which they refer to as *meta-representational competence* [28]. Much of diSessa's work around representations is in the realm of computational mediums and computer representations [21]. These studies showed that students are capable of inventing and productively using their own representations.

To understand student-generated diagrams, one must obtain and then characterize the diagrams. One challenge to the characterization of student-generated diagrams is trying to avoid placing student work in a deficit framing [29]. It might seem natural to use expert-generated diagrams as the comparison point for defining what a "good" student diagram looks like, and, indeed, this might be appropriate when a student is prompted to produce a high quality representation. However, when a student spontaneously generates a diagram as part of the problem-solving process, they may only externalize the representation to the point that it is helpful to them. For example, labeling the axes of a graph may or may not be necessary for a student sketching the qualitative behavior of a function to check the limits of their answer. An expert looking at that same sketch later might be inclined to say the sketch was "poor" or conclude the student had poor sketching skills since it was not to scale or lacked certain elements of an expertlike representation when, in fact, the student simply chose not to include those elements because they were not necessary or useful for them in the moment. diSessa points out that representations include "trade-offs among multiple things that may be desirable in a representation," and that "judging a representation should always be relative to the task…unless one understands the function that a structure serves, one is in a weak position to judge it." [30].

Our work is focused on capturing spontaneously generated diagrams from physics problem solving and attempting to characterize and understand why and how they were generated and used. This work aims to inform a wide array of research and instructional practices, including the following: informing further study of student understanding of professional representations, including multiple representations; providing a foundation for more targeted investigations of student use of specific types of diagrams during problem solving; and informing instructional practices regarding learning goals and assessment around student diagramming. However, as alluded to previously, without a clear reference point or point of comparison, characterization of unprompted diagrams becomes tenuous at best. In this study, we compare unprompted and prompted diagrams generated by the same set of students, using the prompted diagrams, rather than our notions of 'ideal' or 'complete' diagrams, as a foundation for interpreting diagrams produced spontaneously during problem solving. This method allows us to identify diagramming elements that were intentionally included or excluded in the unprompted situations, and we have a separate indicator

of which elements might be challenging or unfamiliar to students.

The method suggested above is predicated on the hypothesis that students' prompted and unprompted diagrams would differ in meaningful ways. In a previous paper [23], we tested this hypothesis through the creation of a set of interviews designed to capture both unprompted and prompted student-generated diagrams. These diagrams were generated by 19 undergraduate and graduate physics majors during 1-on-1 problem solving interviews with author MV. A subset of the interview consisted of 6 multiple-choice problems (with no diagramming prompts) and paired diagramming tasks (with explicit diagramming prompts). By looking at both the unprompted and prompted diagrams generated by the same set of students on these six problem-task pairs, we attempted to answer the following research question:

RQ1 How do spontaneously generated student diagrams used in problem solving compare with similar, prompted student diagrams?

Given the variety of reasons for which a student might generate a diagram, a direct comparison between such diagrams and prompted diagrams generated by the same student can help us understand how diagrams generated without prompting, where students are generating diagrams for themselves, may differ from easier-to-capture prompted diagrams, where the student is generating the diagram for someone else.

In this initial study, we found that students' unprompted diagrams were generally smaller, messier, and contained fewer details that their prompted diagrams [23]. Furthermore, we found instances in which producing unprompted diagrams seemed to be correlated with higher performance on the problem (these were primarily problems that contained vectors), instances in which an unprompted diagram seemed unnecessary for the students to be able to correctly solve the problem, and instances in which unprompted diagrams did not seem to help students solve difficult problems. These findings suggest that diagramming can certainly be beneficial for students in some situations, but that there are many instances in which requiring students to diagram a problem may not be beneficial for them (at least in terms of getting the correct answer), which is consistent with and expands on previous findings that require students to draw free body diagrams does not necessarily help them and may indeed impede students [7].

We now expand on this work by both revisiting RQ1 to expand our analysis (to look at, for example, differences between lower-division, upper-division, and graduate students) and also by looking at other aspects of student diagramming in unprompted situations. While the previous study only looked at 6 of the problem-task pairs, we now look at diagrams and data from all of the 18 multiple-choice problems that students worked through in the interviews

described above. With this expanded dataset, we also aim to explore two additional research questions:

RQ2 When and how do students use diagrams during problem solving, such as to orient to the problem, organize information, or directly obtain an answer?

RQ3 How does generation and use of diagrams differ between lower-division undergraduate, upper-division undergraduate, and graduate physics majors, if at all?

To answer **RQ2**, we look at what elements students include in their diagrams as well as when in the problem-solving process the diagram occurs. We identify instances when student-generated diagrams appear to be used primarily or entirely to orient the student to the scenario, as a tool that is manipulated to help solve the problem, or both.

To answer **RQ3**, we look at the various elements of our analysis (correct answers, accurate prompted diagrams, properties of unprompted diagrams, timing) across student cohorts ranging from 1st and 2nd year undergraduate physics majors to physics graduate students. Our findings with regards to this research question are informing ongoing research about the development of diagramming habits throughout a physicist's career.

Development of our interview prompts is discussed in Sec. II. Primarily qualitative analysis from our interviews are described in Sec. III, whereas quantitative analyses can be found in Sec. IV. Finally, Sec. V contains implications for teaching and research and also limitations of this study.

II. METHODS

To study and characterize diagrams generated by students, we conducted 19 problem-solving interviews in which physics majors worked on multiple-choice (MC) physics problems followed by explicit diagramming tasks. We note that while the students were recruited to, and working in, an interview setting (i.e., outside the context of a course, working in a room with a researcher present and available to answer clarifying questions), the students were not prompted to talk through their reasoning and the interviewer did not ask them to discuss their problem-solving process. This was done to avoid prompting students to switch to a mode of "communicating their process to the interviewer" as that would have defeated the goal of observing spontaneously generated diagrams for the purpose of problem solving only. Students were allowed to use calculators (provided if necessary) or other tools that they thought would help them: some students used rulers, and one used a protractor, in the diagramming tasks at the end of the interviews (discussed in the following paragraph). We interviewed 4 lower-division undergraduates (1st and 2nd year), 5 juniors (3rd year undergraduate), 5 seniors (4th year undergraduate), and 5 graduate students: this population of students was selected so we could explore diagramming throughout novice-to-expert development, and because graduate students are an understudied group in physics education research. Interview participants

TABLE I.   Topics covered in the interview. These topics were covered at the introductory level (with the exception of delta functions, which was included to investigate student difficulties observed in previous work [31]). Thermodynamics was abandoned due to the interview being too long and difficulty generating problems for which students were likely to spontaneously generate diagrams.

Included topics	Vector addition	Moment of inertia
	Kinematics	Oscillations
	Rotational kinematics	Mirrors
	Forces	Lenses
	Conservation of energy	Snell's law
	Conservation of momentum	Delta functions
	Center of mass	Electric potential
	Torque	Electric field
Abandoned topics	Thermodynamics	

responded to email solicitations sent to physics majors at the University of Colorado Boulder and were financially compensated for their time.

The MC problems were primarily introductory-physics level content, and most could be solved entirely or in part with a diagram. One problem, with an electric charge distribution written in terms of delta functions, was not introductory level but was included as part of a potential expansion of previous findings of student difficulty with charge distributions described using delta functions [31]. Problem development began with the authors generating a list of content areas appropriate for physics majors who had taken introductory physics courses (Table I): this process was guided by skimming through several introductory texts [32,33]. Special attention was paid to topics with common or canonical diagrams. From this list, author MV developed 20 MC problems and 8 diagramming tasks that closely resembled 8 of the 20 MC problems. A wide range of physics contexts was desirable in that it allowed us to look at student-generated diagramming more holistically than would be possible in just a few contexts. While most of the MC problems were introductory level, they were intentionally designed to be challenging enough that we were confident students would be engaging in authentic problem solving.

To avoid inadvertently cuing students to draw diagrams during the interview, multiple-choice problem statements were text only and did not ask for explanations or illustrations. Distractors were developed by intentionally making simple errors during the problem-solving process and listing the resulting answer as one of the MC options.

The diagramming tasks at the end of the interview asked students to carefully sketch, draw, or graph and then label a scenario resembling one of the MC problems, giving us problem-task pairs to compare unprompted and prompted

diagrams of similar situations. These 28 items were piloted first by author B. R. W., with follow-up pilots by an additional physics faculty member and a physics graduate student, both at the University of Colorado Boulder. The faculty member and graduate student were told that the purpose of the study was to investigate student problem solving, but we did not inform them that we were specifically interested in capturing and studying student-generated diagrams. Two thermodynamics problem-task pairs were abandoned after these pilot interviews over concerns of both interview length and the challenge of developing thermodynamic problems that were likely to illicit spontaneous diagrams from students without prompting. The remaining 18 MC problems and 6 diagramming tasks were tweaked following the pilot interviews, and the final wording of all problems and diagrams can be found in Tables II and III. Solutions to all of the final 18 items could, in principle, include a diagram; however, we intentionally included a spectrum of items, from those that were very likely to illicit a diagram to those less likely to do so. This was done to test the hypothesis that physics students may have been enculturated to always produce a diagram of some sort, regardless of its potential value.

As with the pilot interviews, students participating in the interviews were only told that we were investigating student problem solving; we gave them no indication we were interested in studying student diagrams. As a result, the diagrams that students generated while working through the MC problem were generated only if a student chose to draw the unprompted diagram as part of their problem-solving process. During the interview, all students first worked through the 18 MC problems followed by the 6 diagramming tasks; all students saw the questions in the same order. By asking students to generate prompted diagrams at the end of the interview, we were then able to compare student diagrams across epistemic frames [34,35]. An epistemic frame is the perspective a learner has for an educational setting or task. Working to learn and working to get the right answer are two common (and sometimes opposing) epistemic frames a student might adopt during problem solving. Another example of epistemic frames might be the idea that a task should be done independently or in collaboration with others. Studies of student diagrams that gather diagramming data from homework and exams, therefore, must be considered in the context of those assignments, where student diagrams are generally encouraged and often explicitly assessed.

By telling our interview participants their goal was to get the correct answer, but by not asking them to show their work or draw diagrams, we aimed to place students in a problem-solving epistemic frame. This means that any diagram they generated would be generated because the student thought it would help them solve the problem, not because it would be expected or assessed.

TABLE II.   The six problem (left) and task (right) prompts that comprised the problem-task pairs for our interview. The problems in this table are listed in the same order as they were presented to students with the exception of the E-field and delta problems, which were switched for students (though interspersed between these problems were the other 12 multiple-choice problems). The tasks were the final 6 items in the interview, presented in the same order as they were presented to students. Students were informed at the beginning of the interview that they could ask the interviewer for formulas and use a calculator during the interview. For the interviews, these problems were given numerical rather than descriptive names.

Problem-Task Pairs

Maps

A car drives 2.5 miles South, 4 miles Southeast, 1 mile East, then 2 miles North. How far is the car from its starting point? (a) 4.1 miles (b) 5.1 miles (c) 6.3 miles (d) 9.5 miles	Carefully draw and label a 'map' where a person travels 200 m North, 200 m Southeast, 500 m South, then 400 m East.

Blocks

A 1 kg block sits on top of a 2 kg block, which sits on the floor. The coefficient of both static and kinetic friction is 0.4 between the two boxes and 0.3 between the bottom (2 kg) box and the floor. If a 100 N force is applied horizontally to the top (1 kg) box, will the bottom box slide along the floor? (a) Yes (b) No (c) Not enough information	Carefully draw and label a free body diagram for a block sliding down a slope of angle θ with coefficient of kinetic friction μ.

Decay

A mass hanging from a spring is displaced 20 cm and oscillates up and down when released. Because of friction and air resistance, the amplitude of oscillation is halved every 10 seconds. What is the amplitude of oscillation after 15 seconds? (a) 4.6 cm (b) 5.0 cm (c) 7.1 cm (d) None of the above	Carefully sketch and label the graph of $10\cos(x) \cdot 2^{-\frac{x}{2\pi}}$ through at least 3 full periods.

Mirrors

Two flat, square mirrors are placed edge to edge with a 60° angle between their surfaces. Light comes in, bounces off of each mirror exactly once, and then leaves the system of mirrors. What is the angle between the incoming and outgoing light? (a) 30° (b) 60° (c) 120° (d) Not enough information	Carefully draw and label a ray-diagram for a ray of light that bounces off of two mirrors with an angle of 135° between them.

E-field

A charge of $-q$ sits at $(\ell, 0, 0)$ and a second charge $2q$ sits at $(0, \ell, 0)$. What is the electric field at $(0, 0, \ell)$? (a) $\frac{kq}{\ell^2}(-1, 2, 1)$ (b) $\frac{kq}{\ell^2}(1, -2, 1)$ (c) $\frac{kq}{\ell^2}\frac{(-1,2,1)}{2^{\frac{3}{2}}}$ (d) $\frac{kq}{\ell^2}\frac{(1,-2,1)}{2^{\frac{3}{2}}}$	Carefully draw and label 3 points: A at $(\ell, 0, 0)$, B at $(0, -\ell, 0)$ and C at $(0, 0, \ell)$. Then, if a charge $-q$ sits at A and a charge $3q$ sits at B, sketch the electric field at the point C.

Delta

Consider the 2-dimensional charge distribution: $\sigma(x,y) = A\delta(x-1)\delta(y+1) + B\delta(x+1)\delta(y-1) + C\delta(x+2)$, and assume A, B, and C have the appropriate units to make all of the dimensions work out. How much total charge exists in the space defined below: $-3 \le x \le 3$ $0 \le y \le 3$? (a) $A + B + C$ (b) $A + B + 3C$ (c) $B + 3C$ (d) None of the above	Carefully draw and label the following charge distribution: $\sigma(x,y) = A\delta(x+2)\delta(y-1) + B\delta(x+1)\delta(y+2) + C\delta(x-1)$.

In order to observe diagrams generated in an explicitly communicative epistemic frame, the final 6 items of the interview asked students to carefully generate and label a diagram: there was no actual problem to solve. By comparing these diagrams to students' unprompted diagrams, we could argue that differences in these diagrams suggest we were successful in capturing diagrams generated while students were in two distinct epistemic frames, and we could identify features of diagrams that appeared differentially in these different frames.

In tandem with the epistemic frames framework, we used the distributed cognition framework [26,27,36,37] to help us understand differences between students' unprompted and prompted diagrams. The distributed cognition

TABLE III.   The 12 multiple-choice problems that were not paired with a diagramming task. These problems are listed in the same order (when read left to right, top to bottom) as they were presented to students, though they were interspersed between the problems from the problem-task pairs (Table II). For the interviews, these problems were given numerical rather than descriptive names.

Unpaired multiple-choice problem prompts	
Two cars	**Projectile**
Two cars drive from a house to a park. Both cars leave at the same time but take different routes. Car A travels 500 m at 10 m/s, then 3000 m at 30 m/s, and finally 1500 m at 20 m/s. Car B travels 2000 m at 10 m/s, then 1500 m at 20 m/s. Which car arrives at the park first? (a) Car A (b) Car B (c) Car A and Car B arrive at the same time	The position of a projectile is given by: $y(x) = 5 \text{ m} + 20x - 2.5 \text{ m}^{-1} \cdot x^2$. At what position x does the projectile start moving in the negative y-direction? (a) 2 m (b) 4 m (c) 6 m (d) 8 m
Rolling disk	**Stage ramps**
A solid disk of mass 1 kg and radius 0.1 m rolls along the ground at a speed of 10 m/s. The disk then rolls up a smooth hill. How high up the hill does the disk roll before stopping and rolling back down? (a) 5.0 m (b) 7.5 m (c) 10.0 m (d) None of the above (e) Not enough information	An elevated stage has a steep ramp leading up to it on one side and a shallow ramp leading up to it on the other (the ramps are the same total height but different lengths). You want to slide a 100 kg box onto the stage: using which ramp requires you to expend the least energy to slide the box up and onto the stage? (a) The steep ramp (b) The shallow ramp (c) The amount of energy required is the same for both ramps (d) Not enough information
Center of mass	**Collision**
A uniform, flat tray (1 kg) centered at the origin holds three items: A book (2.5 kg) sitting at (10 cm, −10 cm); a cup (0.5 kg) sitting at (−10 cm, −10 cm); and a plate (1 kg) sitting at (0 cm, 10 cm). What is the center of mass of the system (i.e., of the tray and the three objects)? (a) (0 cm, 0 cm) (b) (0 cm, −2.5 cm) (c) (4 cm, −4 cm) (d) (5 cm, −5 cm)	A 500 g basketball (at rest) is hit by a 50 g tennis ball moving at 8 m/s. If the tennis ball bounces off of the basketball at 2 m/s (back in the direction it came from), what is the speed of the basketball after the collision? (a) 2.4 m/s (b) 1.0 m/s (c) 0.8 m/s (d) 0.6 m/s
Circular track	**Torque**
A car drives around a circular track ($r = 20$ m). The car starts from rest and accelerates at 1 m/s^2. What is the magnitude of centripetal (radial) acceleration after 10 seconds? (a) $\frac{1}{2}$ m/s^2 (b) 1 m/s^2 (c) 2 m/s^2 (d) 5 m/s^2	A tree branch of uniform mass density ($L = 2$ m and $m = 4$ kg) lies horizontally across a creek with only its ends touching the ground. A squirrel ($m = 500$ g) uses this branch as a bridge to cross over the creek. When the squirrel is one-third of the way across the branch, what is the upward force of the ground on the **far** end of the branch? (a) 20.83 N (b) 21.66 N (c) 22.50 N (d) 23.33 N
Moment of interia	**Thin lens**
What is the moment of inertia for baseball bat of length L (with a linear mass density $\lambda(x) = \lambda_0 + \alpha x$) rotating about $x = 0$? (a) $\frac{\lambda_0 L^3}{3}$ (b) $\frac{\lambda_0 L^3}{6}$ (c) $\frac{\lambda_0 L^3}{3} + \frac{\alpha L^4}{4}$ (d) $\frac{\lambda_0 L^3}{6} + \frac{\alpha L^4}{8}$	A converging lens with a focal length of 1 m is placed 2 m from an object. What is the magnification M of the image of the object? (a) 1 (b) −1 (c) $\frac{1}{3}$ (d) $-\frac{1}{3}$
Snell's law	**E-potential**
Light travelling in a vacuum strikes the window ($n = 1.3$) of a spaceship at 10° from the normal. By what angle does the direction of the light change as it moves from the vacuum into the window? (a) 13° (b) 7.7° (c) 2.3° (d) None of the above	What is the electric potential at a point 1 m from a +2 C charge, 2 m from a −3 C charge, and 3 m from a +1 C charge? (a) $0k$ C/m (b) $\frac{5}{6}k$ C/m (c) $\frac{5}{36}k$ C/m (d) Not enough information

framework holds that people will externalize (e.g., generate a diagram) for two main reasons: to reduce their cognitive load or to communicate information [26,36]. It is reasonable to expect externalizations would differ when done for these two reasons. Furthermore, this framework helps us avoid a deficit framing of student diagrams: rather than view these diagrams as lacking when compared to expert diagrams, we can view the features that are present as
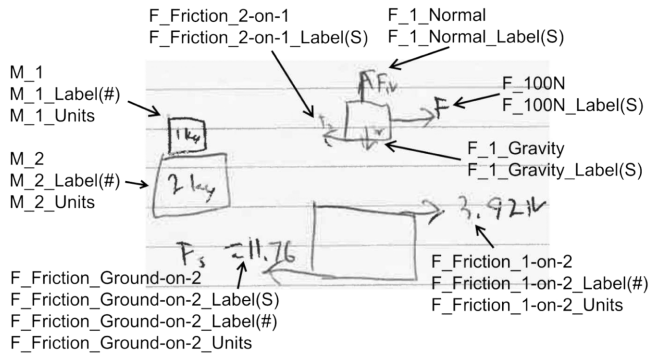
FIG. 1.   An illustration of the coding performed on all student diagrams. Actual coding was done in a spreadsheet and not as a markup of each diagram. We also time stamped each detail (excluding units) for the unprompted diagrams.

signposts that help us understand students' cognitive processes.

After conducting all 19 interviews, we reviewed the video recordings to code every mark and label students added to their unprompted and prompted diagrams (see Fig. 1). We did this by first generating a list of expected diagram elements (e.g., particular objects, arrows, labels, etc.), then iteratively coding these elements and identifying other marks and labels not yet accounted for in our coding [38,39]. We revised our coding until it accounted for virtually every mark a student made as part of a diagram. Interrater reliability was established by author B. R. W. independently coding one-sixth of the diagrams, selected randomly, with agreement between this coding and the primary coding reaching 80% after a few general clarifications and reaching 96% agreement with the original coding after discussion.

Coded elements were also time stamped based on the videos for the 18 MC problems. Time stamps were recorded using a custom computer script that allowed a researcher (author M. V.) to watch and control the video recordings manually while also capturing and recording time stamps into a spreadsheet, producing precise and accurate time stamps for more than 1700 diagram elements. Using the time stamps of coded diagram elements, we identified when during problem solving students generated diagrams and if students updated those diagrams. This added another lens through which to compare student-generated diagrams, investigating how diagram timing correlates with selecting the correct answer and how it differs across student cohort.

In addition, every problem was coded along multiple axes: Student, problem, answer (correct, incorrect, blank), problem-solving duration, diagram (diagram, no diagram), etc. We also added a code for the cohort the student was in: lower-division (undergraduate), junior, senior, or graduate student. If a student drew an unprompted diagram while solving a problem, in addition to every detail, we also coded if these details were given in the problem statement

or not, and whether something was written before the student started the diagram (see Sec. IV). We did not code or time stamp algebra or other actions taken by students (e.g., using a calculator, speaking, etc.), other than the first nondiagram action taken after the diagram was started.

With regards to student answers to questions, we note that there are only 7 instances for which students selected an answer without showing work, and 4 of these instances were for problem 6 which required no math: because of this low frequency of an answer without any work, we do not attempt to adjust for student guessing.

III. QUALITATIVE RESULTS AND INTERPRETATIONS

In this section, we will reiterate and elaborate on some of our previous findings [23] regarding the 6 problem-task pairs (Sec. III A) and categorize the different ways that students used diagrams for each of the other 12 multiple-choice problems (Sec. III B). In the final section (Sec. IV), we look at the timing aspect of unprompted student-generated diagrams (i.e., when students drew) and other patterns that emerged in our analyses.

A. Problem-task pairs

Table IV shows student performance on problem-task pairs across three axes: answer correctness, unprompted diagram detail, and prompted diagram accuracy. For unprompted diagrams, we identify diagrams depicting only information given in the prompt (G), diagrams depicting additional information, such as calculated values or simplifications (G+), when no diagram was drawn (ND), and when the work area was left blank (B). Distinguishing between (G+) and (G) helped us determine if a diagram was used for organizing information, discussed more in Sec. IV. We now discuss student answers and diagrams for each problem-task pair.

1. Maps

The maps problem-task pair (described in Table II) required students to consider the addition of four spatial vectors. As measured by the high number of correct student answers (Table IV), this problem was one of the easiest in the interview and the easiest of the 6 problems that were paired with a diagramming task. Every student drew an unprompted diagram for the multiple-choice problem. For the diagramming task, every student completed the task and only one student drew an inaccurate diagram. Only 3 students did not select the correct answer: 2 because of algebra mistakes and 1 who only drew 3 of the 4 segments of the path. No students attempted to answer the problem by drawing a diagram to scale and measuring the desired distance.

As can be seen in Fig. 2—and this holds for all 18 problems the students were asked to complete—the level of detail in the prompted diagram could vary wildly. Of the

TABLE IV. Overall performance on problem-task pairs. Numbers are number of students. If either part of a problem-task pair was skipped due to time constraints, the pair is omitted from the table. Columns indicate correctness of student answers to problems. Subcolumns categorize *unprompted* diagram content: more than given information (G+); only given information (G); no diagram (ND); or blank (B) if no work was shown. Rows describe accuracy of *prompted* student diagrams: Accurate or containing small errors (e.g., a 400 m length drawn longer than a 500 m length), inaccurate (e.g., forces missing or going in the wrong direction), or no diagram if the student chose not to diagram the scenario.

Maps ($n = 19$) — Maps MC problem

Paired task	Correct				Incorrect				No answer			
	G+	G	ND	B	G+	G	ND	B	G+	G	ND	B
↳ Accurate	11	4	⋯	⋯	2	⋯	⋯	⋯	1	⋯	⋯	⋯
↳ Inaccurate	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
↳ No diagram	1	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯

Blocks ($n = 19$) — Blocks MC problem

Paired task	Correct				Incorrect				No answer			
	G+	G	ND	B	G+	G	ND	B	G+	G	ND	B
↳ Accurate	9	4	⋯	⋯	1	1	⋯	⋯	1	1	⋯	⋯
↳ Inaccurate	⋯	⋯	⋯	⋯	⋯	1	⋯	⋯	⋯	1	⋯	⋯
↳ No diagram	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯

Decay ($n = 17$) — Decay MC problem

Paired task	Correct				Incorrect				No answer			
	G+	G	ND	B	G+	G	ND	B	G+	G	ND	B
↳ Accurate	1	4	2	3	⋯	1	1	⋯	⋯	⋯	⋯	1
↳ Inaccurate	1	⋯	1	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
↳ No diagram	⋯	1	⋯	⋯	⋯	⋯	⋯	1	⋯	⋯	⋯	⋯

Mirrors ($n = 19$) — Mirrors MC problem

Paired task	Correct				Incorrect				No answer			
	G+	G	ND	B	G+	G	ND	B	G+	G	ND	B
↳ Accurate	4	1	⋯	⋯	3	3	⋯	⋯	1	⋯	⋯	⋯
↳ Inaccurate	⋯	1	⋯	⋯	1	⋯	⋯	⋯	1	⋯	⋯	⋯
↳ No diagram	2	⋯	⋯	⋯	⋯	1	⋯	⋯	1	⋯	⋯	⋯

E-field ($n = 15$) — E-field MC problem

Paired task	Correct				Incorrect				No answer			
	G+	G	ND	B	G+	G	ND	B	G+	G	ND	B
↳ Accurate	4	1	⋯	⋯	3	1	1	⋯	⋯	⋯	1	⋯
↳ Inaccurate	1	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
↳ No diagram	⋯	⋯	⋯	⋯	⋯	1	⋯	⋯	2	⋯	⋯	⋯

delta ($n = 18$) — Delta MC problem

Paired task	Correct				Incorrect				No answer			
	G+	G	ND	B	G+	G	ND	B	G+	G	ND	B
↳ Accurate	⋯	2	⋯	⋯	⋯	1	⋯	⋯	⋯	⋯	⋯	⋯
↳ Inaccurate	⋯	⋯	⋯	⋯	⋯	2	2	1	⋯	⋯	⋯	⋯
↳ No diagram	⋯	⋯	1	⋯	⋯	2	2	1	⋯	1	1	2

problems discussed in this paper, the maps problem is the only one where every student clearly referred to their unprompted diagram when answering the question (specifically, we observed all students refer back to their diagram, generally to retrieve numeric values, while setting up their algebraic solutions). So while the level of detail in these unprompted diagrams varied greatly, all of the students used the diagrams to answer the question, and almost every student (including the two whose diagrams are shown in Fig. 2) answered the problem correctly. This could indicate that the drawn segments are all students needed to externalize to answer the question, or that different students needed to externalize different amounts in order to succeed (and that they only drew what they needed). In reality, we believe the truth is somewhere in between: that the segments and potentially other features needed to be externalized for these students to succeed, but that some of the features sometimes drawn (including labels and units) likely did not have to be externalized in order to solve the problem.
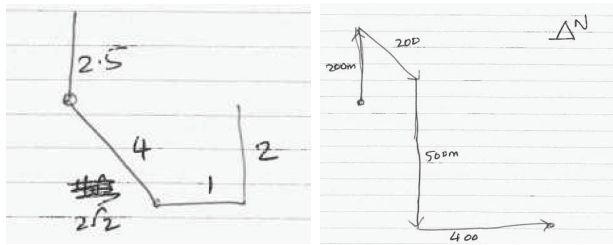
(a) Maps diagrams from one student.
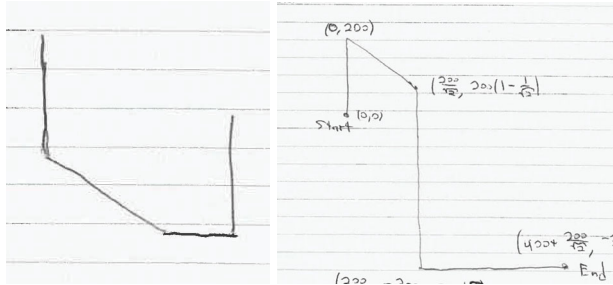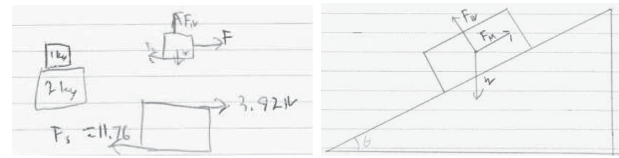

(b) Maps diagrams from another student.

FIG. 2. Examples of unprompted (left) and prompted (right) student diagrams for the maps problem-task pair, with each pair of images being from one student. Detail of unprompted diagrams (left) varied widely.
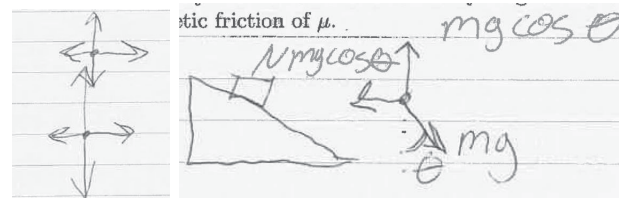

(a) Blocks diagrams from one student.
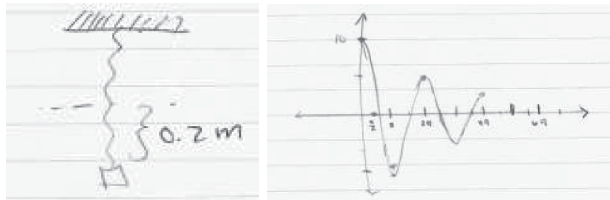

(b) Blocks diagrams from another student.

FIG. 3. Examples of unprompted (left) and prompted (right) student diagrams for the blocks problem-task pair, with each pair of images being from one student.

2. Blocks

The blocks problem-task pair asked students to consider mechanical forces (including friction) acting on massive blocks. The problem is a canonical but challenging force problem that requires students to reason about and correctly calculate friction forces, and for this problem, every student drew an unprompted diagram, though 8 students only included information that was given to them in the problem statement (i.e., they did not add any force pairs, calculated values, or other features not described in the problem statement). The presence of (and detail in) unprompted diagrams is not strongly correlated with the student selecting the correct answer, as can be seen in Table IV and discussed more in Sec. IV, with the exception that 5 of 6 students who drew the (nongiven) forces that were acting on the block in question (the bottom block) got the correct answer.

In the unprompted diagrams, 11 of 14 students who represented forces with arrows drew those arrows at the locations where the forces act (Fig. 3), rather than at the center of mass, and 5 students also did this during the prompted diagramming task. We note this is consistent with student examples shown in Heckler [7], though this feature is not discussed at length in that text.

As exemplified in Fig. 3, the amount of detail in unprompted diagrams varied greatly. For this problem, diagrams that include only information given in the problem statement do not seem to improve the chances a student will answer the question correctly. The only details that appear to increase the likelihood a student will select

the correct answer are the nongiven forces on the bottom block. Many students were able to draw accurate free-body diagrams for forces acting on blocks when they were prompted to do so but chose not to for the MC problem. This suggests that students do not always draw diagrams (or elements of diagrams) that could help them correctly solve problems, even if they are capable of drawing such diagrams.
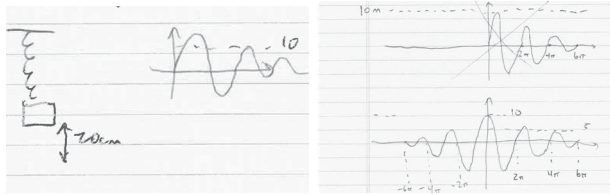
3. Decay

The Decay scenarios asked students to consider the amplitude (over time) of a damped oscillator. While it is possible to plot the displacement of the oscillator over time (an exponentially decaying cosine curve) or depict the amplitude of the oscillator over time (a decaying exponential), only one student drew an unprompted graph depicting the oscillator [Fig. 4(b) left] for the multiple choice problem. Eleven students, including the one who drew a graph, drew sketches of the oscillator [e.g., Fig. 4(a) left and 4(b) left]. Only two students included any features in their diagrams that were not stated in the problem statement.

This problem-task pair provides strong evidence of diagrams being used by students almost exclusively to orient them to the physical situation, as a single sketch of the oscillator does not capture the time evolution of this system and thus cannot directly help students answer the question. This is not to say that the sketches were not useful or that orienting a student to a problem is not a productive use of a diagram, only that these diagrams did not directly lead the student to a solution. As 13 of 17 students who answered the question selected the correct answer, and as students were largely successful in drawing a decaying trig function when prompted [e.g., Fig. 3(a) right and 3(b) right], we believe students productively chose tools other than diagramming (e.g., algebra, sense making about the

(a) Decay diagrams from one student.

(b) Decay diagrams from another student.

FIG. 4. Examples of unprompted (left) and prompted (right) student diagrams for the decay problem-task pair, with each pair of images being from one student.
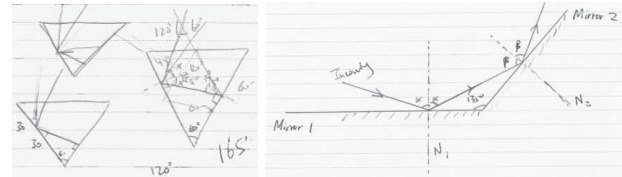

(a) Mirrors diagrams from one student.

(b) Mirrors diagrams from another student.

FIG. 5. Examples of unprompted (left) and prompted (right) student diagrams for the mirrors problem-task pair, with each pair of images being from one student. As can be seen in these examples, many students drew the mirrors multiple times for both the MC problem and the diagramming task.

functional behavior of the oscillator) to solve this problem, and thus graphs of the oscillator were unnecessary.

Furthermore, while students' prompted diagrams of the decaying oscillator generally had the correct shape, many students struggled with these diagrams. As exemplified in Fig. 3(b), some students struggled with knowing whether decaying sine or cosine curves were appropriate or with knowing what the graph would look like for $t < 0$. As this region of the graph would likely be irrelevant in solving a problem, this suggests the possibility that difficulties with these diagrams, which were almost never drawn spontaneously for the MC problems, could negatively impact a student's ability to answer a question correctly.
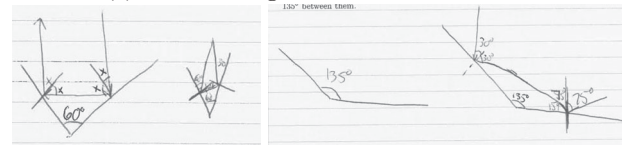
4. Mirrors

The mirrors problem-task pair asked students to consider a single beam of light that reflects off two adjacent plane-mirrors with a given angle between the mirrors. Every student drew unprompted diagrams for this problem, with many students electing to sketch the mirrors multiple times [e.g., Fig. 5(a) left and 5(b) left]. The incoming and outgoing light rays cross in the MC problem, with the problem asking students to find at what angle they intersect. However, as the prompt is ambiguous as to which of these supplementary angles (60° or 120°) students were supposed to find, we categorized students' answers correct if they successfully identified either of these angles. While the problem statement did not give the angle at which the incoming ray hit the first mirror, no student asked for this information or even commented on it during problem solving.

Eight students answered the problem correctly, another 8 answered incorrectly, and 3 students drew unprompted diagrams without selecting an answer. For the most part, there seemed to be little correlation between the amount of

detail in the unprompted diagrams and getting the correct answer, with the exception that the 4 students who selected the incorrect answer "Not enough information" drew very little, perhaps because they believed that adding further detail would not be productive. Interestingly, the 3 students who did not select an answer drew some of the most detailed unprompted diagrams.

We believe most students who drew the situation multiple times first started with a sketch to orient themselves to the problem, and once they realized more fully what the problem was asking, they created new, often larger, often neater sketches that they developed into ray diagrams. Additionally, we found no correlation between students who selected the right or wrong answer with students who did or did not struggle to draw the similar situation when prompted in the diagramming task. This indicates that drawing this particular ray diagram is both challenging to do and not sufficient (by itself) to answer the mirrors problem correctly.

5. E-field

The E-field scenarios asked students to consider the electric field (both magnitude and direction) at a point noncollinear with two point charges (the charges and point of interest all lie a distance ℓ from the origin, each on a different axis). Six students selected the correct answer for the electric field, and 5 students selected an answer with the correct magnitude and form but a sign error. None of the 5 students with a sign error labeled their axes or in anyway depicted positive or negative directions on their diagrams, whereas 4 of 7 students who did not have a sign error (the six who selected the correct answer and 1 who had only a magnitude error) labeled axes or directions. Two of the 5 students with a sign error had an accurate diagram but without direction labels, and 3 did not have a diagram or did not draw the components of the electric field. Three

(a) E-field diagrams from one student.

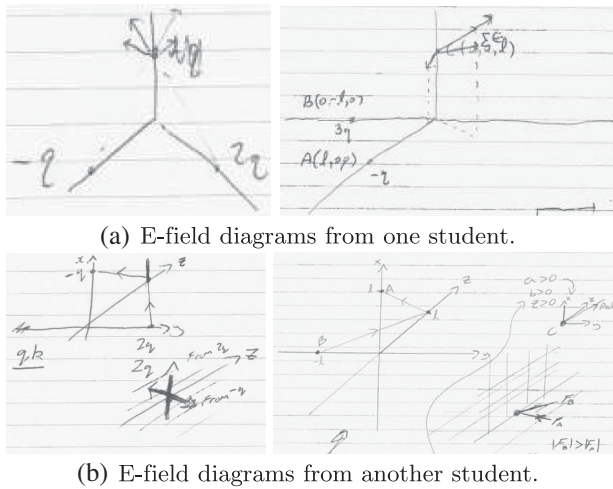(b) E-field diagrams from another student.

FIG. 6. Examples of unprompted (left) and prompted (right) student diagrams for the maps problem-task pair, with each pair of images being from one student.

(a) Deltas diagrams from one student.

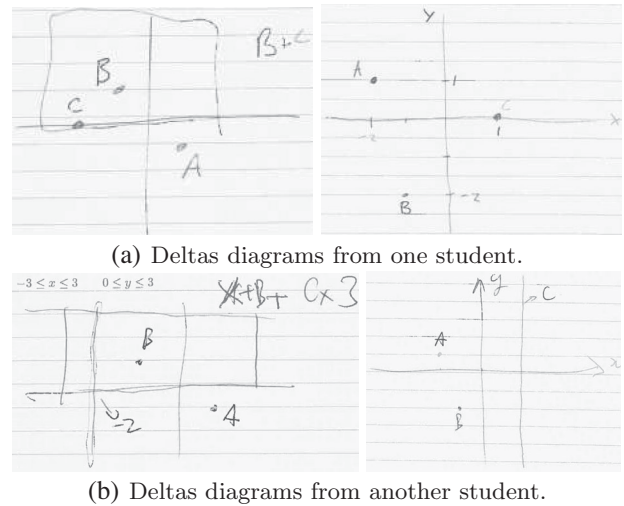(b) Deltas diagrams from another student.

FIG. 7. Examples of unprompted (left) and prompted (right) student diagrams for the maps problem-task pair, with each pair of images being from one student.

other students began to work on the problem but did not select an answer, and the remaining 4 students skipped the problem or were asked to skip because of time concerns.

As 5 of the 6 students who selected an incorrect answer only had a sign error, it is especially noteworthy that 3 of the 5 students with a sign error (who had not indicated direction in their unprompted diagrams) correctly drew the direction of the electric field and included directional labels in the prompted diagram [e.g., the student whose work is shown in Fig. 6(a)]. So while it seems that directional labels are generally necessary to solve this type of problem correctly, many students did not recognize this or chose not to include direction in their diagram without prompting. We believe that this was either an issue of activation or choice, rather than diagramming ability, since these students were largely capable of drawing accurate diagrams when prompted in the paired diagramming task.

6. Deltas

The delta problem-task pair asked students to consider a two-dimensional charge distribution (σ) containing delta functions to represent two point charges (A & B) and a line charge (C). Only 3 of 18 students correctly answered the problem (which asked how much charge is in an enclosed region of space), and only 1 of these 3 students drew an accurate unprompted diagram (in total, 8 students drew diagrams, only 2 of which were accurate). Six students stated or implied that the line charge C was a point, and 4 drew it as such (e.g., Fig. 7). Five students, four of whom indicated C was a point charge, asserted the total charge in the enclosed region was $B + C$, which is especially noteworthy as $B + C$ was not a provided option (though "None of the above" was an option).

One of the students who drew C as a point in the MC problem drew C correctly as a line for the paired task, indicating that they might have simply misread the expression for σ in the problem. Eleven students did not complete the diagramming task, as it was the last item in the interview (though only 1 student was asked by the interviewer to skip this task) and it was the only task asking students to depict a scenario that would not show up in an introductory physics course. However, the high number of students who selected an incorrect answer for the problem and drew a diagram consistent with their incorrect answer suggests that many students who did not complete the diagramming task would not have been able to draw an accurate diagram given more time. The difficulties we observed with this problem-task pair are consistent with research indicating that graphical interpretations of delta functions are challenging even for upper-division physics students [31].

B. Unpaired multiple-choice problems

There were 12 problems for which students may have drawn unprompted diagrams that where were not paired with a diagramming task. We now briefly discuss these problems by grouping them based on content area and patterns in the unprompted diagrams the students drew. A summary of student performance on these problems (as well as the paired MC problems) can be found in Table V.

1. Two cars and E-potential

Two of the problems, the two cars and E-potential, are isomorphic, requiring students to take the ratio of scalar values given in the prompts and then sum these ratios. Despite this mathematical similarity, all students answered the two cars problem correctly but only 12 of 17 students answered the E-potential problem correctly (and 2 students were asked to skip this problem because of time

TABLE V. Student performance on all MC problems, with paired MC problems bolded. Numbers (generally $N = 19$) are number of students. Columns indicate correctness of student answers to problems. Subcolumns categorize *unprompted* diagram content: More than given information (G+); only given information (G); no diagram (ND); or blank (B) if no work was shown. Numbers for the paired MC problems represent a collapsing of the values in Table IV. Students were asked to skip problems for time in the moment of inertia, Snell's law, and E-potential unpaired MC problems, as well as the decay, delta, and E-field paired MC problems.

	Correct				Incorrect				No answer			
	G+	G	ND	B	G+	G	ND	B	G+	G	ND	B
Two cars	1	1	17	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯	⋯
Maps	12	4	⋯	⋯	2	⋯	⋯	⋯	1	⋯	⋯	⋯
Projectile	⋯	1	14	⋯	⋯	⋯	2	⋯	⋯	⋯	1	1
Blocks	9	4	⋯	⋯	1	2	⋯	⋯	1	2	⋯	⋯
Rolling disk	2	5	2	1	2	2	3	⋯	1	⋯	1	⋯
Stage ramps	2	4	1	2	1	4	1	3	1	⋯	⋯	⋯
Center of mass	1	7	⋯	⋯	5	3	⋯	⋯	⋯	3	⋯	⋯
Collision	2	6	6	⋯	⋯	3	2	⋯	⋯	⋯	⋯	⋯
Circular track	⋯	8	8	⋯	⋯	⋯	⋯	⋯	⋯	3	⋯	⋯
Torque	12	2	⋯	⋯	2	⋯	⋯	⋯	2	1	⋯	⋯
M. of inertia	5	1	6	⋯	⋯	⋯	⋯	⋯	2	⋯	3	1
Decay	2	5	3	3	⋯	1	1	1	⋯	⋯	⋯	1
Mirrors	6	2	⋯	⋯	4	4	⋯	⋯	3	⋯	⋯	⋯
Thin lens	2	4	2	⋯	4	2	⋯	1	⋯	1	⋯	3
Snell's law	11	⋯	1	⋯	3	⋯	2	⋯	1	⋯	⋯	⋯
Deltas	⋯	2	1	⋯	⋯	5	4	2	⋯	1	1	2
E-potential	3	2	6	1	5	⋯	⋯	⋯	⋯	⋯	⋯	⋯
E-field	5	1	⋯	⋯	3	2	1	⋯	2	⋯	1	⋯

constraints). Only 2 students drew diagrams for the two cars problem, while 10 students did so for the E-potential problem, including the 5 who selected an incorrect answer, all of whom selected "not enough information."

From this answer selection of not enough information, their unprompted diagrams, and some student comments, we believe these 5 students thought the positions of the charges (not just their distance from the point of interest) impacted the value of the electric potential. This dependency on position would be true for the value of the electric field at the point of interest, so it is possible students were conflating electric potential and electric field.

There are multiple possible explanations for these E-potential diagrams: students who did not know how to solve the problem may have started by drawing the situation; students who thought the position of the charges mattered began to draw a diagram to organize this information; and/or students started to draw a diagram, but as the problem prompt did not list the positions of the charges, trying and being unable to draw the diagram may have lead students to an incorrect conclusion.

2. Projectiles

Fifteen of 19 students answered the projectiles problem correctly. Only 1 student drew an unprompted diagram—an upside-down parabola with no axes, labels, or annotations—and this student selected the correct answer. Three of the 4 students who did not select the correct

answer set $y(x) = 0$ rather than $y'(x) = 0$ when solving for the x position where the projectile begins to fall. As with the decay problem discussed earlier, this issue seems to be an instance in which, while the answer could be obtained directly from a carefully drawn diagram (a graph of the given function), drawing a diagram was not necessary for students to answer the problem correctly.

3. Rolling disk and moment of inertia

For the rolling disk and moment of inertia problems, the majority of students applied the correct physical principles, but a sizable fraction of students struggled with proper execution. For the rolling disk problem, 4 of the 17 students who tried to solve the problem using conservation of energy did not take into account all three forms of energy: linear kinetic, rotational kinetic, and potential energy. For the moment of inertia problem, 4 of the 15 students who tried to use moment of inertia did not set up an appropriate integral. In addition, for both problems, a number of students misremembered formulas (e.g., $\omega = vr$) and/or made algebraic mistakes while solving the problems.

For the rolling disk problem, 7 of 12 students who drew a diagram selected the correct answer, while 5 of 9 who did not draw a diagram selected the correct answer. For the moment of inertia problem, 5 of 7 students who drew a diagram selected the correct answer, while 6 of 10 who did not draw a diagram selected the correct answer. For these problems, it seems like students who drew diagrams were

slightly more likely to answer the question correctly, but this does not mean that the diagram helped the students answer correctly. As may be the case with the delta problem (Sec. III A 6), it is possible that students who had a better grasp of these situations were simply more able to draw them, and it did not seem as though drawing diagrams for these problems helped students catch mistakes that they had made, either in setting up their equations or in their algebra.

4. Center of mass and torque

The center of mass and torque problems both required students to take into account multiple "real world" objects (not just "masses" or "charges"). Fifteen students used the center of mass equation, but only 7 of those got the correct answer. Six students did not take into account the mass of the tray on which the items sat for the center of mass problem, either because they forgot it or because they intentionally—and erroneously—excluded it because its center of mass was at the origin. Two other students solved for just the x position of the center of mass and ignored both the tray and the plate, possibly because they lie on the y axis. One student got the correct answer by, we gather from verbal utterances, sense making about the center of mass being closer to heavier objects.

Eleven students correctly applied the principals of torque to answer the torque question correctly, while one student correctly used proportional reasoning and superposition. Two students made mistakes while trying to use torque, and 4 students only used or calculated forces while trying to solve this problem.

Every student drew an unprompted diagram for both of these problems. As with the mirrors problem, the diagrams drawn by students who did not select an answer contained the largest number of details. Along with other analyses (discussed in Sec. IV), this suggests that in these instances, students tried to use diagrams to help reach an answer that they were unable to obtain through other means (e.g., through algebra). Additionally, many students who did not select the correct answer were able to generate accurate unprompted diagrams, which again parallels the mirrors problem.

5. Stage ramps

Although friction was not discussed in the stage ramps problem statement, students needed to consider friction in order to answer the Stage Ramps question correctly (which 9 students did). Twelve students drew diagrams for the stage ramps problem, including 6 of the 9 who considered friction. Only one student included a sketch that depicted a friction force. As we saw in the mirrors, center of mass, and torque problems, the most detailed diagrams were drawn by students who did not select an answer.

6. Collision

All students used conservation of momentum in the collision problem, but 5 students had a sign error. Two of these students set up the equations with an incorrect sign, and 3 students had the correct equations but made a sign error while solving for the answer. The other 14 students all got the correct answer.

Eleven students drew unprompted diagrams for this problem, including 8 who got the correct answer. Three of 5 students with a sign error drew an (accurate) unprompted diagram. One student who drew a diagram, and two students who did not set up their conservation of momentum equation with a sign error: the other two students who selected the wrong answer made an algebraic mistake. Overall, 10 of 11 students who drew a diagram set up their equations correctly, so it is possible (as with the E-field problem) that drawing a diagram that depicts direction can help students avoid sign errors.

7. Circular track

Ten students drew diagrams for the circular track problem (one of whom showed no other work). Eighteen students used rotational kinematics to solve the problem, with 16 students getting the correct answer and three students, including a student who only showed a sketch, not selecting an answer. Seven of the 16 students who selected the correct answer drew a diagram. As with the mirrors, center of mass, torque, and stage ramps problems, the most detailed diagrams were drawn by students who did not select an answer.

8. Thin lens and Snell's law

To find the magnification of an image from a thin lens, 13 students drew diagrams: 6 got the right answer (2 of whom got the answer from just the diagram), 6 got the wrong answer, and 1 did not select an answer. Two students got the correct answer without drawing any diagrams. The diagrams drawn for this problem are similar to those drawn for the decay problem in that many of the diagrams lacked the features that would have been necessary to aid in solving the problem from the diagram.

One student was asked to skip the Snell's law problem and four students only provided a sketch (one of whom selected the correct answer). All 14 other students used Snell's law, with 11 getting the answer correct and 3 making minor mistakes leading to an incorrect answer. Of students who drew a diagram, 6 of 11 who got the correct answer never drew the refracted rays of light, as was the case with 1 of the 3 students with a diagram who selected an incorrect answer.

It seems that many of these diagrams were drawn before the students switched to a completely algebraic mode of problem solving that did not depend on the diagram, similar to the decay problem.

IV. QUANTITATIVE RESULTS AND INTERPRETATIONS

As discussed in Sec. II, we qualitatively coded every detail that appeared on student diagrams. An illustrative example of this coding is shown in Fig. 1. We would also like to remind the reader that all of the students who participated in our study were physics majors who had completed at least a full year of physics, and thus it is likely that some of our findings may not generalize to physics majors in introductory courses or nonphysics majors in physics courses.

In total, across the 19 students and 18 problems (with no diagram prompts), students worked through 331 problems (with 11 instances where we asked a student to skip a problem due to time constraints). Students drew unprompted diagrams for 66.2% of these problems ($N = 219$), and Fig. 8 shows what fraction of students drew a diagram for each problem. We identified a total of 1771 diagram details, including 1051 markings (e.g., arrows, objects, etc.), 644 (nonaxis) labels (141, or 21.9%, of which had units), and 76 axis details (drawing or labeling axes). Since student use of axes was not a primary focus of this study, the coding of axes details was not as granular as the coding of other diagram details: for example, drawing both an x and y axis would be coded as a single axis detail, which is part of why the number of axis details appear to represent such a small portion of the overall details.

Also shown in Fig. 8 is that there were very few instances—just 18 of 219—in which students drew a diagram but wrote something else down first. In these few instances, students typically did some form of orienting (writing down general formulas, rewriting key information from the problem statement, identifying relevant physical laws, etc.), and the diagram quickly followed. There were only two instances in which it appears a student tried to solve a problem without a diagram, then after being unsuccessful, drew a diagram as part of a second attempt. With 91.8% of diagrams being the first thing students drew, and with diagrams coming very early in the problem-solving process for the other 18 instances, we believe that most, if not all, unprompted diagrams drawn in this study likely helped students orient themselves to the problem they were trying to solve. While this set of interviews did not include time to discuss the purpose of the diagrams with interviewees, a subsequent set of interviews, using a modified subset of these problems and diagramming tasks, were designed and conducted with the goal of having extended conversations with participants about their diagrams [40]. In these interviews, all 10 participants discussed using diagrams to orient to the situation described in the problem and/or to the question that was being asked of them, with several participants stating that they often start generating a diagram to help them orient to the problem even before they even finish reading the problem statement.
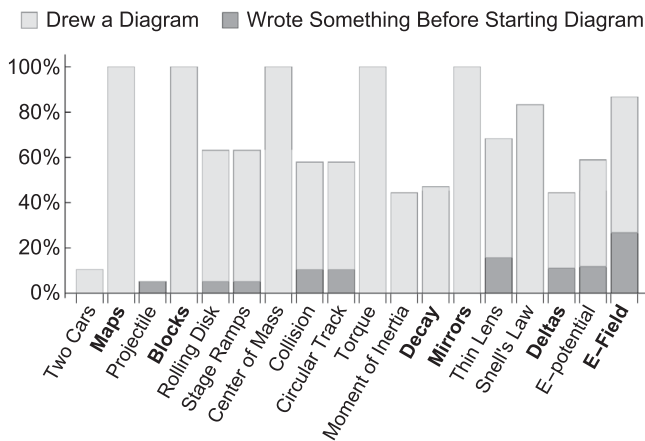

FIG. 8. The percentage of students who drew an unprompted diagram for each MC problem. The darker shading indicates instances in which a diagram was drawn but not the first thing the student put onto the paper.

We also wish to note that the use of the diagram to orient to a problem is not the only type of orientation that may occur. There may be a larger process of internalized orienting at work, of which orienting with a diagram is only one component. At the same time, as discussed below, we believe many of these diagrams served additional functions (later in the problem-solving process) beyond orienting.

The overall results of time-stamping the details in unprompted diagrams is displayed in Figs. 9 and 10, which show how long after starting the problem the students made various types of markings (both in absolute time—seconds—and as a fraction of the total time the student spent on the problem). Both in absolute and relative terms, the vast majority of diagramming occurs near the start of problem-solving process. As a whole, median time of the last diagram detail was 55 sec or 45% of the way through solving the problem.

With 19 students completing a total of 331 problems, drawing 219 unprompted diagrams and a total of 1771 markings, we are able to look at patterns in when and how the students generated and used unprompted diagrams. When comparing two categorical variables—where our categorical variables include Student, Problem, Cohort, Answer, or if a Diagram was drawn—we used a Fisher exact (FE) test to determine statistical significance, where p values were simulated via Monte Carlo methods. When comparing a categorical value with a numerical value (duration, number of details), we used a Kruskal-Wallis (KW) test, or in cases where we had multiple independent variables, a multiple linear regression (MLR). These statistical tests were run in the programming language R.

In the sections below, we first evaluate hypotheses that were generated during the development of these interviews and during our initial analyses of the problem-task pairs: and we refer to these analyses as confirmatory analyses. We then conduct exploratory analysis to identify any
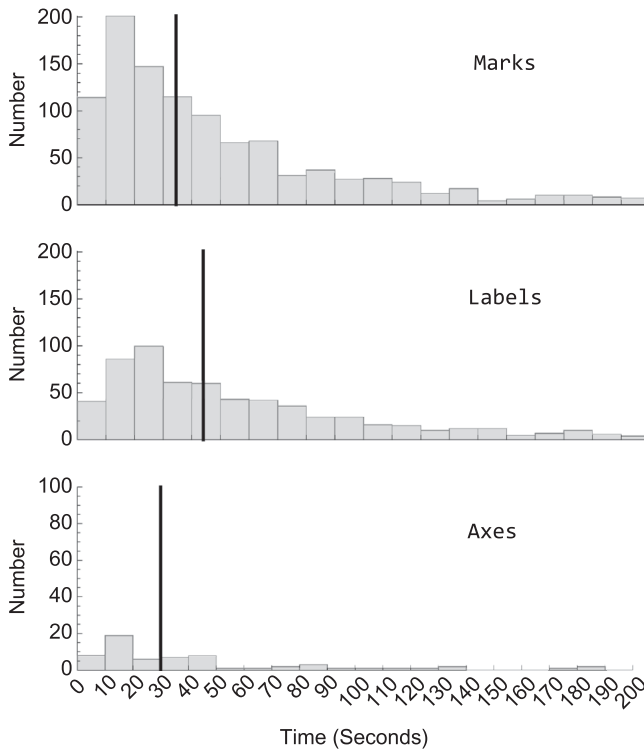
FIG. 9. Histogram of student diagram details (marks, labels, and axes) by 10 sec intervals for 219 unprompted diagrams. For example, between 10 and 20 sec after starting a problem, students put a total of 201 markings (e.g., arrows, objects, etc.), 86 nonaxis labels, and 19 axis details (axes or axis labels) on diagrams. These graphs include 97% ($n = 1705$) of diagram details, with the remaining 3% ($n = 46$) occurring between 200 and 598 sec. Vertical bars indicate median times. Note: The vertical axis for the axes plot is different because of the low number of axes details. The times for 20 marks were not recorded, generally because the students' hand obscured their work.
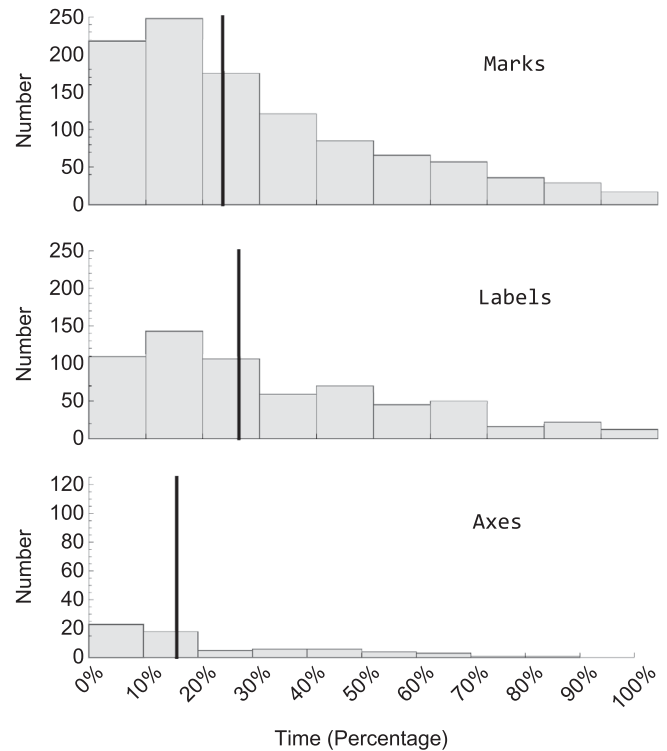
FIG. 10. Histogram of timing of student diagram details (marks, labels, and axes) as a fraction of total time spent on each problem. For example, a total of 218 markings, 109 nonaxis labels, and 23 axis details were drawn by students in the first 10% of time spent problem solving. Vertical bars indicate median times. Note: The vertical axis for the axes plot is different because of the low number of axes details.

unanticipated but interesting trends in the data. When considering a large number of statistical tests, as we did in this study, it is important to account for the possibility of false positives. Since a p value of 0.05 signifies a 5% chance of the observed result being seen if the null hypothesis is true, running a large number of tests increases the likelihood that some of the results will be false positives. To avoid this pitfall, we adjusted p values for our exploratory analyses using the Holm-Bonferroni method. We do not use the adjusted p values for our confirmatory analyses, since we have independent reasons (explained in each analysis below) to suspect these results might be significant.

A. Confirmatory analyses

In this section, we discuss findings that, from our interview development and previous work [23], we expected to be statistically significant. The first analyses in this section test hypotheses formed during interview development that did not fit within the scope of our previous paper, either because they looked at differences

between student groups or to problems that were not analyzed in that study. Analyses of hypotheses from our previous work (in the final two paragraphs of this section) were run only data from the 12 MC problems that were not a part of the previous study.

We hypothesized that graduate students would answer the most questions correct for a number of reasons: they have had the most formal physics education, they elected to pursue more advanced physics degrees, and many graduate students teach in introductory physics courses. Indeed, we found that graduate students were the most likely group to select a correct answer (78.7% of the time), compared to seniors (55.1%), juniors (55.8%), and lower-division students (61.2%). That graduate students answered more questions correctly than did any other cohort is statistically significant (FE $p < 0.001$), but the differences between the other groups were not (FE $p > 0.5$).

In choosing the order of questions when designing the survey, we intentionally placed the E-potential problem before the E-field problem. While there is enough information in the E-potential problem to solve the problem, there is not enough information to draw a complete diagram since the coordinates of the charges are not know. As the

E-fields problem requires students to consider relative positions and not just distances, we did not want to cue students to think about coordinates in the E-potential problem by having it come after the E-fields problem. Such cuing, we argue, could bias students towards both generating diagrams that they otherwise might not have drawn and then concluding that there was not enough information to solve the problem. Furthermore, we hypothesized that students who generated unprompted diagrams for the E-potential problem would be less likely to select a correct answer than those who did not. This ended up being the case (FE $p < 0.05$), where 100% of students (7 of 7) without a diagram selected the correct answer, compared to 50% (5 of 10) of students who drew a diagram. As discussed briefly in Sec. III B 1, it could be that students who did not know how to solve the problem were more likely to generate a diagram for it, or that the process of generating a diagram hindered students ability to solve the problem.

From our initial study of the 6 problem-task pairs, we noted that graduate students were the most likely group to draw unprompted diagrams and that lower-division students were the least likely. We hypothesized that this would also be true for the other 12 MC problems (that were not paired with a diagramming task). We found that lower-division students drew fewer diagrams than did graduate students (FE $p < 0.05$), but neither of these groups were statistically different from juniors and seniors (FE $p > 0.1$ for both graduate and lower-division students). Overall, across the 18 MC problems, graduate students drew diagrams 74.2% of the time, compared to 65.2% for seniors, 68.6% for juniors, and 53.7% for lower-division students. We wish to note that it is possible that a larger dataset might have the statistical power to distinguish juniors and seniors from the other groups.

Finally, in our initial study of the problem-task pairs, we had also found that lower-division students, while they drew the fewest unprompted diagrams, tended to include the highest number of details in those diagrams. Again, we tested this hypothesis using the 12 unpaired MC problems (this time using a multiple regression since, as described in the next section, there is a strong correlation between the problem and the number of markings on students' diagrams, and the regression takes into account for which problems students were drawing diagrams), and we found the difference in number of details between lower-division students and other students is not statistically significant (MLR $p > 0.5$).

B. Exploratory analyses

As can be seen in Fig. 8, between 5.3% and 100% of students drew diagrams on any given problem, and the correlation between the specific problem being solved and the fraction of students who drew unprompted diagrams was statistically significant (adjusted FE $p < 0.001$). This result is somewhat surprising, given that most of these interview questions were designed to evoke unprompted diagrams in students. Furthermore, and as mentioned in the previous section, the specific problem being diagrammed impacted the number of details included in unprompted diagrams (adjusted KW $p < 0.01$). This test took into account only the 219 problems for which students drew a diagram.

Overall, the presence of a diagram was not a significant predictor of whether a student got the answer correct (adjusted FE $p = 1$) when looking across all 18 MC problems (though, as mentioned above, it was significant in the case of the E-potential problem). This result makes sense with our previous findings that, for some problems, diagrams do not seem to be necessary in order for students to select the correct answer, and for other problems, a large number of students who drew diagrams were still unable to select the correct answer.

In addition to certain problems evoking student diagrams more or less frequently, different students were more or less likely to draw unprompted diagrams. The percentage of problems for which individual students drew a diagram range from 33.3% to 94.7%, a correlation that was statistically significant (adjusted FE $p < 0.001$). Considering this result in light of our findings that, as cohorts, only lower-division and graduate students were statistically distinct, we believe this shows that the variation within each cohort of students is greater than the variation between the cohorts.

For 122 of the problems in which a student drew an unprompted diagram (55.7% of the time), students added details to their diagram that were not given in the problem statement. These details were significantly correlated with which problem a student was solving (adjusted FE $p < 0.001$), but were independent of which student (or cohort) was solving the problem or if students selected the correct answer or not (adjusted FE $p = 1$ for these 3 tests). From this we conclude that the extent to which a student used their diagram to organize information depends on the problem but does not predict success.

In 102 instances (46.6% of times when a diagram was drawn), a student revisited the diagrams to add to it (or created additional diagrams) after doing something else (generally performing algebra). This was associated with a higher likely-hood of not getting the answer right—selecting an incorrect answer or no answer—with 72% of non-revisited diagrams and 52% of revisited diagrams being drawn on diagrams with the correct answer selected (adjusted FE $p < 0.05$). Revisiting a problem is highly correlated with which problem the student was trying to solve (adjusted FE $p < 0.001$), but was not correlated with the student (adjusted FE $p = 1$) or group of students (adjusted FE $p = 1$). From this, we do not conclude that revisiting a diagram is detrimental to students, rather that it seems most likely that students would revisit a diagram while attempting to solve a challenging problem.
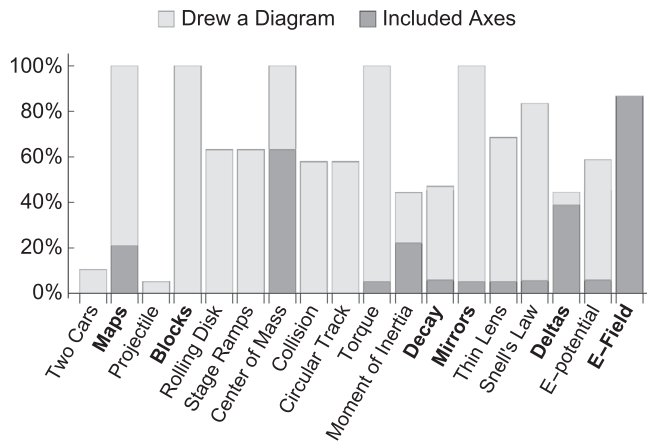
FIG. 11. The percent of diagrams drawn (for each problem prompt) that included axes. Axes were most common for the problems that directly (center of mass, E-potential) or indirectly (delta) provided coordinate information in the problem prompt.

In our initial design of these interview problems, we wanted to present students with problems that were possible to solve entirely from a diagram, and we believe 6 of the problems could be reasonably solved in a purely geometric way. This happened rarely. There were only 9 times students obtained an answer directly from a diagram, 5 times for the delta problem (1 correct answer, 2 incorrect answers, and 2 students who did not select an answer because the answer they obtained was not listed as an answer option.) and 3 times for the thin lens problem (2 correct answers, 1 incorrect answer). As we posited previously, this finding suggests that there may be room to improve or increase instruction around geometric solutions in physics [23].

When looking at how long it took students to solve problems, we found the largest difference was that lower-division students spent on average 9% more time solving problems than did students in the other cohorts, though this finding was not statistically significant (adjusted LR $p = 1$).

Finally, while looking at student use of axes was not a primary focus of this work, we did find that students drew axes on only 21.0% of unprompted diagrams (46 of 219), with no individual student drawing axes for more than 6 problems. Figure 11 shows the percentage of student-generated diagrams for each problem that included axes. As the problems for which students drew axes the most were problems that gave students information (directly or indirectly) in the form of coordinates, these findings are not surprising. These findings do, however, allow us to nuance our previous claim that students are much about twice as likely to add axes to prompted diagrams than to unprompted diagrams: this discrepancy—which is not necessarily bad and may reflect reasonable decisions on the part of students—may be even larger for problems with noncoordinate information.

V. SYNTHESIS AND CONCLUSION

In this study, we presented 19 physics majors with 18 multiple-choice physics problems and 6 diagramming tasks, where each task was of a similar physical situation as 1 of the multiple-choice problems. The 18 multiple-choice problems gave us the opportunity to study diagrams that students generate without prompting during the problem-solving process. The paired diagramming tasks let us compare prompted and unprompted diagrams across 6 different physical contexts.

Previous work addressing **RQ1** (How do spontaneously generated student diagrams used in problem solving compare with similar, prompted student diagrams?) has been published previously [23] and expanded on here. We found that prompted diagrams were generally larger, objects in the diagrams were more to scale, and these diagrams were more likely to include axes and units and were unprompted diagrams. These findings suggest that these features are more valued in diagrams created for the purpose of communicating information than they are in diagrams generated to aid the student in problem solving.

This previous work also found specific problems in which students often selected the correct answer without drawing a diagram, and other problems where students often selected an incorrect answer after drawing a diagram. This finding was confirmed in the current study when looking at a separate subset of the multiple-choice problems. For this reason, we reiterate our previous suggestion that instructors consider when and how to teach and assess diagrams in problem solving, as problem context seems to influence how helpful diagrams can be. This is not to say that diagramming should not be taught in these contexts, merely that there should be alignment between the goals of the course and assessment when it comes to diagramming.

The ways in which unprompted diagrams aid students was the focus of **RQ2** (When and how do students use diagrams during problem solving, such as to orient to the problem, organize information, or directly obtain an answer?). In this study, we found that the vast majority, if not all, of unprompted diagrams generated by students were used (at least initially) to help students orient themselves to the problem. A subsequent study, in which we present students and faculty with a modified subset of these interview questions and then explicitly ask them about the role of diagrams in problem solving, supports and elaborates upon this finding [40]. Future work could use eye-tracking software to help us more fully understand when and how participants refer back to their unprompted diagrams [19].

Looking at differences among groups of students was the focus of **RQ3** (How does generation and use of diagrams differ between lower-division undergraduate, upper-division undergraduate, and graduate physics majors, if at all?). We found only minor differences in diagram frequency between some groups, and our data suggest that, when it

comes to generating unprompted diagrams, even if differences between cohorts were to be identified, differences within a cohort of students are likely larger than the differences between cohorts.

Finally, having coded 1771 diagram details for 219 multiple-choice problems, we are confident in asserting that problem-solving diagrams are messy. We encourage instructors to consider how they might value messy diagrams in assessment, making sure that assessment "should always be relative to the task" [30]. In addition, we believe it is important for instructors to model messy diagrams for students. While not the primary focus of our follow-up study, the instructor-generated diagrams created during interviews were also messy and included mistakes, corrections, and second attempts, and the differences between these messy diagrams and carefully constructed figures in textbooks and lectures are stark. However, if students are never exposed to examples of messy work (including messy diagrams) done by experts, and if students' messy work is not valued in assessment, then students may come to believe that the messy, productive work they are doing is not good physics. More direct investigation into student perceptions of the validity of their messy work, perhaps using the epistemic frames framework, may be of interest to the research community.

Our study is limited in that all students interviewed were physics majors at the University of Colorado Boulder, which has a strong undergraduate physics program and highly selective graduate physics program. Furthermore, the student demographics of this department are overwhelmingly white and male. Our findings may not generalize to other populations, and so further study of student diagramming, especially with nonphysics majors or physics majors in introductory courses, may be necessary.

We believe there are many affordances to our study design: by having students solve problems without specific prompts (diagramming or otherwise), and then by giving students explicit prompts, this study design can be used to explore other areas of student problem solving and how tools used in problem solving manifest across different epistemic frames. One change our study design may have benefited from would have been setting aside time at the end of the interview to explicitly discuss diagramming with students, and this modification is the primary focus of ongoing work. Together with this current study, this work can help inform further work into student-generated diagrams, as well as research and instructional approaches regarding student interactions with expert-generated diagrams.

ACKNOWLEDGMENTS

[1] H. H. Tairab and A. K. K. Al-Naqbi, How do secondary school science students interpret and construct scientific graphs?, J. Biol. Educ. **38**, 127 (2004).

[2] D. E. Meltzer, Student learning in upper-level Tthermal physics: Comparisons and contrasts with students in introductory courses, AIP Conf. Proc. **790**, 31 (2005).

[3] D.-H. Nguyen, N. S. Rebello, M. Sabella, C. Henderson, and C. Singh, Students difficulties in transfer of problem solving across representations, AIP Conf. Proc. **1179**, 221 (2009).

[4] A. Savinainen, A. Mkynen, P. Nieminen, and J. Viiri, Does using a visual-representation tool foster students ability to identify forces and construct free-body diagrams?, Phys. Rev. ST Phys. Educ. Res. **9**, 010104 (2013).

[5] R. J. Dufresne, W. J. Gerace, and W. J. Leonard, Solving physics problems with multiple representations, Phys. Teach. **35**, 270 (1997).

[6] D. Rosengrant, A. Van Heuvelen, and E. Etkina, Case study: Students use of multiple representations in problem solving, AIP Conf. Proc. **818**, 49 (2006).

[7] A. F. Heckler, Some consequences of prompting novice physics students to construct force diagrams, Int. J. Sci. Educ. **32**, 1829 (2010).

[8] P. B. Kohl, N. D. Finkelstein, L. Hsu, C. Henderson, and L. McCullough, Expert and novice use of multiple representations during physics problem solving, AIP Conf. Proc. **951**, 132 (2007).

[9] D.-H. Nguyen, E. Gire, N. S. Rebello, C. Singh, M. Sabella, and S. Rebello, Facilitating Students Problem Solving across Multiple Representations in Introductory Mechanics, AIP Conf. Proc. **1289**, 45 (2010).

[10] P. B. Kohl and N. D. Finkelstein, Student representational competence and self-assessment when solving physics problems, Phys. Rev. ST Phys. Educ. Res. **1**, 010104 (2005).

[11] D. McPadden and E. Brewe, Network analysis of students' representation use in problem solving, in *Proceedings of the 2015 Physics Education Research Conference, College Park, MD* (AIP, New York, 2015), pp. 219–222.

[12] E. Gire and E. Price, Structural features of algebraic quantum notations, Phys. Rev. ST Phys. Educ. Res. **11**, 020109 (2015).

[13] R. R. Bajracharya, P. J. Emigh, and C. A. Manogue, Students' strategies for solving a multirepresentational partial derivative problem in thermodynamics, Phys. Rev. Phys. Educ. Res. **15**, 020124 (2019).

[14] B. Hand and A. Choi, Examining the impact of student use of multiple modal representations in constructing arguments in organic chemistry laboratory classes, Res. Sci. Educ. **40**, 29 (2010).

[15] O. Parnafes and A. Disessa, Relations Between types of Reasoning and Computational Representations, Int. J. Comput. Math. Learning **9**, 251 (2004).

[16] S. H. Schwartz, Modes of representation and problem solving: Well evolved is half solved, J. Exp. Psychol. **91**, 347 (1971).

[17] R. Cox and P. Brna, Supporting the use of external representations in problem solving: The need for flexible learning environments, J. Artificial Intelligence Educ. **6**, 239 (1995), https://psycnet.apa.org/record/1996-29469-001.

[18] S. P. Schmidgall, A. Eitel, and K. Scheiter, Why do learners who draw perform well? Investigating the role of visualization, generation, and externalization in learner-generated drawing, Learn. Instr. **60**, 138 (2019).

[19] A. Susac, A. Bubic, M. Planinic, M. Movre, and M. Palmovic, Role of diagrams in problem solving: An evaluation of eye-tracking parameters as a measure of visual attention, Phys. Rev. Phys. Educ. Res. **15**, 013101 (2019).

[20] E. Gire, A. Wangberg, and R. Wangberg, Multiple tools for visualizing equipotential surfaces: Optimizing for instructional goals, in *Proceedings of the 2017 Physics Education Research Conference*, Cincinnati, OH, edited by L. Ding, A. Traxler, and Y. Cao (AIP, New York, 2017).

[21] A. A. diSessa, The many faces of a computational medium: Teaching the mathematics of motion, in *Computers and Exploratory Learning*, NATO ASI Series, edited by A. A. diSessa, C. Hoyles, R. Noss, and L. D. Edwards (Springer, Berlin, Heidelberg, 1995) pp. 337–359.

[22] B. L. Sherin, How students invent representations of motion: A genetic account, J. Math. Behav. **19**, 399 (2000).

[23] M. Vignal and B. R. Wilcox, Comparing unprompted and prompted student-generated diagrams, in *Proceedings of the 2020 Physics Education Research Conference*, virtual conference, edited by S. Wolf, M. B. Bennett, and B. W. Frank (AIP, New York, 2020), pp. 551–556.

[24] R. Cox, Representation construction, externalised cognition and individual differences, Learning Instr. **9**, 343 (1999).

[25] M. Scaife and Y. Rogers, External cognition: How do graphical representations work?, Int. J. Human-Comput. Stud. **45**, 185 (1996).

[26] T. M. Foster, Implications of distributed cognition for PER, in *Proceedings of the 2020 Physics Education Research Conference*, Boise, ID, edited by S. Franklin, K. Cummings, and J. Marx (AIP, New York, 2002).

[27] D. Kirsh, Thinking with external representations, AI & SOCIETY **25**, 441 (2010).

[28] A. A. diSessa and P. Cobb, Ontological innovation and the role of theory in design experiments, J. Learn. Sci. **13**, 77 (2004).

[29] A. Gupta and A. Elby, Beyond epistemological deficits: Dynamic explanations of engineering students difficulties with mathematical sense-making, Int. J. Sci. Educ. **33**, 2463 (2011).

[30] A. A. diSessa, Metarepresentation: Native competence and targets for instruction, Cognit. Instr. **22**, 293 (2004).

[31] B. R. Wilcox and S. J. Pollock, Student difficulties with the Dirac delta function, in *Proceedings of the 2014 Physics Education Research Conference* (AIP, New York, 2015), pp. 271–274.

[32] R. D. Knight, B. Jones, and S. Field, *College Physics* (Pearson Education, London, 2015).

[33] D. C. Giancoli, *Physics for Scientists and Engineers with Modern Physics* (Pearson Education, London, 2008).

[34] D. Hammer, A. Elby, R. E. Scherr, and E. F. Redish, Resources, framing, and transfer, in *Transfer of Learning from a Modern Multidisciplinary Perspective* 89 (Information Age Publishing, Greenwich, CT, 2005).

[35] A. Maries, S.-Y. Lin, and C. Singh, The impact of students' epistemological framing on a task requiring representational consistency, in *Proceedings of the 2016 Physics Education Research Conference, Sacramento, CA* (AIP, New York, 2016), pp. 212–215.

[36] Roy D. Pea, Practices of distributed intelligence and designs for education, in Distributed cognitions: Psychological and educational considerations, in *Learning in Doing*, 1st ed., edited by G. Salomon (Cambridge University Press, Cambridge; New York, 1997), pp. 47–87.

[37] V. K. Otero, Conceptual development and context: How do they relate?, in *Proceedings of the 2016 Physics Education Research Conference, Boise, ID* (AIP, New York, 2002).

[38] S. E. Stemler, Content analysis, in *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource* (Wiley, 2015), p. 1, 10.1002/9781118900772.etrds0053.

[39] J. W. Cresswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 4th ed. (SAGE Publications, Inc., Thousand Oaks, CA, 2014), pp. 183–213.

[40] M. Vignal and B. R. Wilcox, *Physics Problem-Solvers on Why They Generate Unprompted Diagrams* (AAPT, College Park, MD, 2021), pp. 420–425, https://www.compadre.org/portal/items/detail.cfm?ID=15791.