Testing alternative explanations for common responses to conceptual questions: An example in the context of center of mass

Paula R. L. Heron

Department of Physics, Box 351560, University of Washington, Seattle Washington, 98195-1560, USA

In physics education research it has been common to interpret student errors on conceptual questions in topic-specific ways, rather than in terms of general perceptual or reasoning difficulties. This paper examines two alternative accounts for responses to questions related to the concept of center of mass. In one account, difficulties are said to be perceptual in nature; in the other, difficulties are said to be tightly linked to the concepts in question. Hypotheses derived from the former perspective are tested in studies conducted among university students in introductory physics courses. The results do not provide strong support for the perceptual hypothesis; in fact, there is evidence that performance on perception tasks may be influenced by subjects' ideas about the physical scenario. While the results do not provide general support for one perspective versus the other, the paper serves as an illustration of the type of investigation needed to develop the kind of rich representation of student thinking that will allow instructional resources to be most effectively targeted.

I. INTRODUCTION

Why do students make errors? Specifically, why do students often answer questions in ways that directly contradict what they have been taught? Traditionalists in physics teaching may attribute errors to flaws in the initial transmission process, insufficient practice, or feeble mathematical and reasoning skills. A key contribution of physics education researchers was to promote an alternative explanation: errors might not reflect lack of knowledge of the relevant concepts and principles; they might instead reflect students' attempts to integrate what they are being told with the complex networks of concepts and intuitions developed prior to their arrival in physics classrooms [1]. Broadly speaking, this "constructivist" theoretical framework continues to dominate thinking about the learning and teaching of physics, as researchers have worked to refine notions of what counts as prior knowledge, how it develops, and how it interacts with formal instruction.

Prior knowledge might be viewed as consisting of robust but erroneous frameworks called up in a wide range of contexts (sometimes referred to as *misconceptions*), or as fragmentary notions (broadly called *resources*) whose activation is highly context dependent and that are neither correct nor incorrect [2,3]. A unifying assumption of both views is that students, presented with a physics question,

search for propositions (e.g., "current is used up" or "closer means stronger") and apply them logically to reach a conclusion. Some propositions are stated explicitly by students, others are inferred from their responses. Some propositions (current is used up) seem to be entangled with physics concepts. Others (closer means stronger) are not. However, even in the latter case, an idea is invoked with a probability that depends on the details of the physical scenario in question. It is only in combination with the context that the ideas have interpretive value. In that sense, explanations built on the activation of resources are topic specific.

A second unifying theme of both views is that implications for instruction can be expressed generally (e.g., "identify and address common conceptual difficulties"), but must be implemented on a topic-by-topic basis. It is not generally assumed that a global strategy will improve conceptual understanding over a broader range of topics than those that are directly addressed.

A growing number of papers have challenged these views, motivated in part by the observation that seemingly superficial changes to a task can sometimes dramatically affect student performance and force us to revise our estimates of their competence [4]. These lines of research typically draw on developments in cognitive science, such as dual process theories [5]. A growing body of physics education research is exploring the implications of such theories for physics learning, suggesting that some answers to conceptual questions are generated swiftly and automatically, with formal reasoning only coming into play if the initial response seems unsatisfactory, or to construct a supporting explanation for the initial response [6,7]. Others are examining the
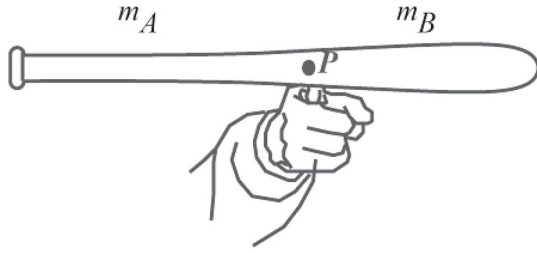
FIG. 1.   Figure used in the bat version of the mass comparison task. Students were told that the bat is balanced on the finger. They are asked (1) if the center of mass of the bat is to the left of, to the right of, or at point P and (2) if the mass of the piece to the left of point P is greater than, less than or equal to the mass to the right of point P. The bat is stylized in order to make it easier for students to gauge the location of the c.m. of each piece by eye.

role of difficulties in spatial or visual reasoning [8]. The unifying theme is the possibility that some errors on seemingly unrelated conceptual questions result from common domain-general reasoning tendencies, which could, in principle, be addressed in a general way.

The implications for research and instruction strategies differ sufficiently that it is essential that the relative roles of domain-specific and non-domain-specific reasoning in physics learning be understood. Developing this understanding will require a body of empirical work to test specific alternative interpretations in a range of different physics topics and instructional settings. This paper represents one such study. The context is the concept of center of mass.

II. BACKGROUND

In 2005, two colleagues and I published a paper that reported on a widespread tendency to attribute balancing of a rigid body to the presence of equal mass (or weight or force) to either side of the fulcrum [9]. The data included responses to a set of related questions administered in introductory physics courses. The *mass comparison* task provides the clearest example.

In the baseball bat version of the mass comparison task, students are shown a diagram (see Fig. 1) and told that the bat is balanced on a person's finger. Students are asked (1) if the center of mass (c.m.) of the bat is to the right, left or at point P (directly above the finger); and (2) if the mass of the piece to the left of point P is greater than, less than, or equal to the mass of the piece to the right. A correct response to part (1) requires recognizing that the c.m. of the bat must be at point P (this is essentially the operational definition of c.m.). More than one approach to part (2) is possible; all require recognizing that the c.m.'s of each piece are approximately at their respective midpoints. A high degree of precision is not required. Thus, the bat can be modeled as two point masses located at different distances from the fulcrum. The left piece, having a c.m. farther from the

fulcrum, must have a smaller mass. A correct answer to part (1) is not necessary to arrive at a correct answer to part (2) if the principle of torque equilibrium is applied (i.e., that the torque produced by each piece with respect xto the fulcrum must be of equal magnitude).

When this task was administered in calculus-based physics courses at the University of Washington (UW) in the late 1990s and early 2000s, 95% of the students answered part (1) correctly but only 20% answered part (2) correctly ($N = 674$). The majority of students claimed that the two pieces must have the same mass. The tasks were given either on course exams or as paper-and-pencil "pretests" prior to instruction in a tutorial [10]. In the latter case, students were seated in a lecture hall and conditions were roughly similar to exam conditions (limited time, no access to course notes or textbooks, no talking permitted) except that the stakes were much lower: students had only to demonstrate a good-faith effort on the pretest to obtain credit. In general, the pretests were administered after instruction on the concept of center of mass in the context of momentum, as well as instruction on torques and the dynamics of rigid bodies. In some cases equilibrium of rigid bodies had also been covered but no student had attended a tutorial on the topic. Neither the test-taking circumstances nor the prior instruction affected the percentage of correct answers, therefore the results were pooled.

As described in Ortiz, Heron, and Shaffer [9], we administered many other tasks, conducted individual interviews and developed a research-based tutorial on static equilibrium. Our analysis led us to conclude that students were having difficulty distinguishing between the concepts of mass (or weight or force) and torque (or moment). A central part of the argument was that students were very successful at mass comparison tasks when the distribution consisted of two pointlike masses at rest on a massless plank, which was in turn balanced on a fulcrum. In such cases, they almost universally took both mass and its location relative to the fulcrum into account. However, on tasks involving rigid bodies, mass was the only variable considered. Reliably using the proper variables when they are highly salient, but neglecting one or the other otherwise is a hallmark of confusion between related quantities. This tendency has been identified in a wide variety of topic areas. For example, mass, volume, and the more ambiguous "amount" are often used interchangeably. However, it is *specific concepts* that are interchanged. Thus the interpretation offered in Ortiz, Heron, and Shaffer is tightly linked to the concepts in questions.

This view was challenged in a recent paper by Sattizahn *et al.* [11] that demonstrates that locating the center of mass of an extended, asymmetric object is perceptually difficult. They conducted studies among university students who were not enrolled in a physics
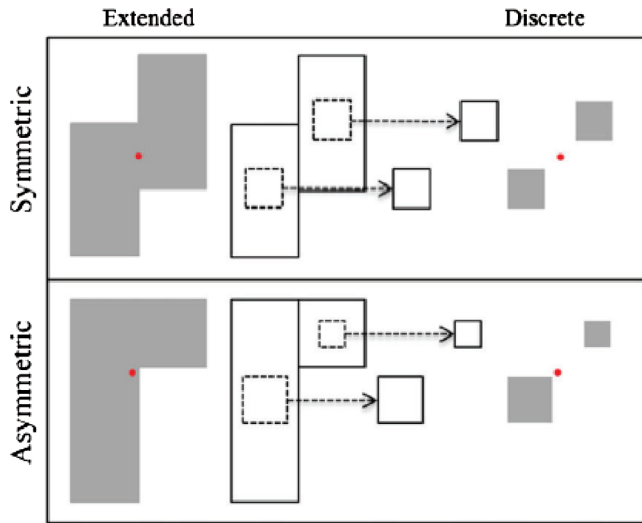
FIG. 2. A figure (Fig. 3) from Sattizahn *et al.* [11]. Subjects were shown a series of such shapes on a touch-screen device and asked to locate the center of gravity with a stylus. The original figure caption states that the red circle indicates the correct answer while the green circle is the average of all responses.

course, although some had previously taken physics at the university level. Subjects were shown a series of flat objects depicted on a touch-screen device (see Fig. 2 for examples) and asked to indicate the location of the center of gravity[1] (COG) of each one with a stylus. The subjects were first told that the COG is the "point of balance or equilibrium" of an object. Results were reported in terms of accuracy and how long it took subjects to answer. Performance was strongest on "discrete, symmetrical" arrangements and weakest on "extended, asymmetric" objects. Specifically, errors (defined as the number of pixels between the predicted and actual locations) were reported as being roughly 4 times the stylus width for extended, asymmetric arrangements, while for other arrangements, errors were "around or less than 2 times the stylus width." Subjects were briefly interviewed after completing a full set of 192 tasks (48 different configurations in four different orientations) and asked to rank the types of configuration by degree of difficulty. Some discussed the strategies they used. The most common strategies included simply marking the center point on symmetric arrangements, mentally balancing the shapes, and separating the extended shapes into discrete parts.

On the basis of these results, they argued that some of the errors we had reported may not be conceptual in nature, but result from distinctly different cognitive processes:

We show how student difficulty in applying COG [center of gravity] to an object such as a baseball bat can be accounted for, at least in part, by general principles of perception (i.e., not exclusively physics-based) that make perceiving the COG of some objects more difficult than others.

The "perceptual hypothesis" expressed in this passage suggests that some fraction of the students in our study were simply unable to perceive the location of the center of mass of the bat, or one of its pieces, and thus unable to answer our questions correctly. The studies described below were intended to test this hypothesis.

III. PREDICTIONS BASED ON THE PERCEPTUAL HYPOTHESIS

A set of testable predictions that follow from the perceptual hypothesis are described here. Results are presented in Sec. IV.

A. Student ability to estimate the locations of the c.m. of each part of the bat should be poor

If students apply a proper procedure to the mass comparison task but are unable to locate the approximate c.m. of one or both of the two parts, then their incorrect answers may not, in fact, reflect a lack of understanding of the concept. If this is the case, we would expect them to perform poorly if we ask them to locate those c.m.'s.

To test this prediction, we developed the *c.m. location* task. Students were shown two pieces of a bat but not told anything about the c.m. or balance point of the bat as a whole (see Fig. 3). They are asked where the c.m. of each piece is relative to some marked points.[2] A high degree of accuracy is not needed for answering the mass comparison task correctly; in fact, students need only to recognize that the c.m. of the left piece is farther from the balance point than that of the right. Therefore, on the c.m. location task, it is not necessary to be very accurate, only to identify a general region.

While the perceptual hypothesis does not generate a quantitative prediction, if the success rate on the c.m. location task is similar to that on the mass comparison task, i.e., a clear majority of students gives incorrect responses or claims there is insufficient information, then there would be grounds for supporting the hypothesis.

---

[1]Though not strictly speaking equivalent, the concepts of center of mass and center of gravity are treated interchangeably here. Also, no equivalences of mass, weight, or force is implied, but students seemed to treat these as interchangeable.

[2]It can be argued that there is not enough information to locate the center of mass of the right piece, however, students have not tended to give that answer, nor have most experts (graduate students, faculty, etc.,) who have seen this question. In any case, while "at or very near point *X*" is referred to as the "correct" answer, its correctness is not the critical factor in the analysis that follows—consistency between the different questions is.
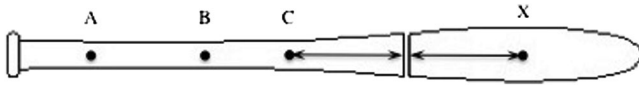
FIG. 3. Figure used in the c.m. location task. Students were told that a bat has been cut into two pieces. They were asked (1) if the center of mass of the left piece is closest to point A, B, or C, and (2) if the center of mass of the right piece is to the left of, to the right of, or at point X. They were told that points C and X are the same distance from the break. In both cases, students were given the option "there is not enough information."

B. Student responses to the mass comparison task should be consistent with their estimates of the locations of the c.m.'s of the two pieces

If students apply a proper procedure to the mass comparison task, then their answers to that task and the c.m. location task should be mutually consistent. For example, students who claim that the c.m. of each piece is the same distance from the break should also claim that the two pieces have the same mass. To check this prediction, we gave some students the mass comparison task first, followed by the c.m. location task; others had the questions in the reverse order. In both cases, students had to submit answers for the first task before they saw the second task and no backtracking was permitted. If answers to the c.m. location task are predictive of answers to the mass comparison task, it would lend support to the perceptual hypothesis.

C. Student performance should be superior on tasks for objects that are not as perceptually challenging as the bat

If students are applying a proper procedure for the mass comparison task but are tripped up by the fact that the bat is an irregular shape, then they should be better able to answer an analogous task with an object composed of flat rectangles.

To test this prediction we can use the *shaded bar* version of the mass comparison task, which was also used in the 2005 study. Students are shown a rectangular bar made of two pieces of different length and made of different material; one is shaded, the other is not (see Fig. 4). The bar is supported on a frictionless pivot at the junction between the two parts. Students are asked if the mass of the unshaded (left) piece is greater than, less than, or equal to the mass of the shaded (right) piece. They can use the same reasoning as in the bat version of this task to conclude that the mass of the left piece is less than the mass of the right piece.

We can also examine results from the *T-shaped object* task, which most closely resembles the Sattizahn *et al.* tasks as balancing is not involved. Students are shown a flat *T*-shaped object made of two pieces (see Fig. 5.) The dimensions of the two pieces and their relative densities are provided so that students can conclude that the two pieces
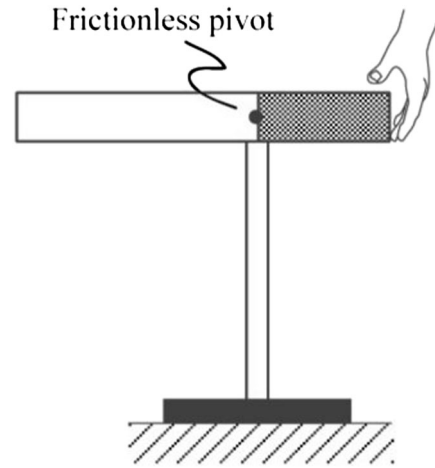

FIG. 4. Figure used in the shaded bar version of the mass comparison task. Students were told that the bar is made of two pieces of different material and that it is balanced on a frictionless pivot and will remain at rest when released. They were asked if the mass of the unshaded (left) piece is greater than, less than, or equal to the mass of the shaded (right) piece.

have the same mass. They are asked whether the c.m. of the object is to the right of, to the left of, or at the junction. They can reason that the c.m. of the object must be midway between the c.m.'s of the two pieces and thus it should be to the right of the junction.

Strong performance on these tasks would suggest that the shape of the bat is a major impediment to success on the bat version, thus lending support to the perceptual hypothesis.

D. Data collection

All of the questions were administered in introductory physics courses at UW as part of a pretest for the tutorial
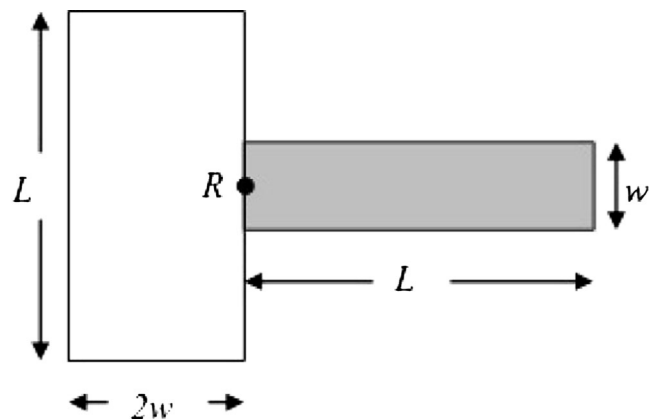

FIG. 5. Figure used in the *T*-shaped object task. Dimensions are given on the diagram. Students were told that the density of the shaded part is twice that of the unshaded part. They were asked if the center of mass of the object is located to the left of, to the right of, or at point R.

Equilibrium of Rigid Bodies. As noted, the pretests may occur before or after lecture coverage of that topic, depending on the course schedule. In the academic quarters in which the data here were collected, previous lectures had covered c.m. (in the context of momentum conservation), rotational kinematics, rotational energy, moment of inertia, and torque. Continuous mass distributions and equilibrium of rigid bodies had not yet been covered.

Pretests are currently administered online. Students have 15 minutes to take a pretest, which they can do at any time during a roughly 24 h period that precedes the tutorial. Several conceptual questions are posed. Students are asked to select answers from a menu of choices and to type brief explanations in a text box. They receive a small amount of credit, whether or not their answers are correct, but they are warned that "Credit will not be given for a pretest that does not show a serious attempt at providing an explanation of reasoning." Not all pretest responses are examined during the quarter in which they are collected, but some fraction are examined to help ensure that that students take them seriously.

Several versions of the pretest were used. The two versions that included both the mass comparison task and the c.m. location task were given in two different course sections that followed the same schedule, and had the same homework and exams, but that met at different times of day and had lectures prepared and delivered by different faculty members ($N = 262$). Two other versions contained either the mass comparison task ($N = 179$) or the c.m. locationtask ($N = 105$). All four included an identical set of additional questions. Results from these four versions allowed us to assess whether the *presence* of the c.m. location task affected responses to the mass comparison task, perhaps by suggesting a productive strategy.

A fifth pretest featured the *shaded bar* version of the mass comparison task followed by the same set of additional questions as the other versions. This version was given to 78 students. The *T*-shaped object task was administered several years ago, to 76 students.

It is important to note several differences between the Sattizahn *et al.* [11] study and ours. In theirs, subjects were volunteers whereas in ours they are enrolled in a physics course and engaged in regular class activities for which they will obtain course credit. Also, our tasks involved only selecting from a short menu of answer choices and (in most cases) furnishing an explanation. Finally, none of the tasks in the Sattizahn *et al.* paper involve balancing and, therefore, they may not invoke the same reasoning strategies. None of these differences are important for interpreting the results, however, as no comparison is made between the performance of their subjects and those in this study.

IV. RESULTS

The results of the mass comparison task were essentially the same for all versions (chi-square test, $p > 0.9$), as were

the results of the c.m. location task (chi-square test, $p > 0.7$). Therefore, they are pooled in the analysis below. Moreover, results on the additional questions (not discussed here) are consistent across all sections. Therefore, direct comparisons of results from different classes can be made.

A. Student accuracy in locating the c.m. of the bat pieces

Students were mostly successful at locating the c.m. of each piece of the bat. The correct answers were given by about 73% of students for the left piece and 83% for the right piece ($N = 367$). A total of 62% answered both questions correctly. In contrast, only 35% answered the mass comparison task correctly ($N = 441$).[3]

B. Consistency between the c.m. location task and the mass comparison task

Results from the c.m. location task can also be classified according to what each student's pair of responses suggests about the relative distances of the c.m.s of the left and right pieces to the break point. For instance, the responses of a student who claimed that the c.m. of the left piece is at point B, but the c.m. of the right piece is to the left of point X could be categorized as consistent with "c.m. of the left piece farther from the break than that of the right piece" (abbreviated as $x_{c.m.L} > x_{c.m.R}$). It should be noted that some pairs of responses do not have an unambiguous interpretation. For instance, if a student indicates that the c.m. of the left piece is at point B while that of the right piece is to the right of point X, it is not possible to say how they compare with certainty.

Using this scheme, 69% of students gave answers implying $x_{c.m.L} > x_{c.m.R}$ while 16% gave answers implying $x_{c.m.L} = x_{c.m.R}$ ($N = 367$). The corresponding answers to the mass c.m.parison task ($m_L < m_R$ and $m_L = m_R$) were given by 35% and 58%, respectively ($N = 441$). Thus almost three times as many students gave the "equal mass to both sides" answer as the responses to the *c.m. location* task would predict.

The responses of individual students who answered both the c.m. location task and mass comparison task provide a more stringent test of consistency. (It is important to note that there is no evidence that the order of the questions affected the results: the frequency of answers in each class

---

[3]It is noteworthy that performance on the mass comparison task is consistently higher than was reported in Ortiz, Heron, and Shaffer [9]. There are several possible reasons, including the fact that the course design has changed significantly, now including clicker questions, online homework with instant feedback, etc. Also, the tasks in Ortiz, Heron, and Shaffer [9] were administered under more controlled circumstances (paper and pencil in a lecture hall under supervision), as opposed to being completed online, and out of class as is now the case. It is not known whether these or other factors are responsible, but in any case no comparisons are being made with the original data.

TABLE I. Results from pretests in which students were asked both the c.m. location and mass comparison tasks (with percentages of the total).

$N = 262$		Mass comparison task (bat)			
		$m_L = m_R$	$m_L < m_R$	$m_L > m_R$	other
c.m. location task (bat)	$x_{\text{c.m.}L} = x_{\text{c.m.}R}$	19 (7%)	18 (7%)	1 (<1%)	4 (2%)
	$x_{\text{c.m.}L} > x_{\text{c.m.}R}$	98 (37%)	67 (26%)	1 (<1%)	16 (6%)
	$x_{\text{c.m.}L} < x_{\text{c.m.}R}$	2 (1%)	0	0	0
	other	25 (10%)	7 (3%)	0	4 (2%)

are consistent on these two tasks as well as on the additional tasks that are not discussed here.) The results are summarized in Table I.

Only 7% of the students gave a set of responses consistent with the "equal mass to both sides" conclusion ($N = 262$). A total of 33% gave a mutually consistent set of responses, correct or incorrect. (This fraction rises to 40% if all ambiguous responses are eliminated). The equal mass to both sides conclusion was the most probable answer to the mass comparison task for students in every category on the c.m. location task: selected by 45% and 54% of those in the $x_{\text{c.m.}L} = x_{\text{c.m.}R}$ and $x_{\text{c.m.}L} > x_{\text{c.m.}R}$ categories, respectively. Considering only the two most frequent responses to both tasks (the four cells in the upper left corner of the table, which represent 77% of all students), the split between "$m_L = m_R$" and "$m_L < m_R$" is comparable (Fisher exact test, two-tailed, $p > 0.4$). Thus responses to the c.m. location task are not strongly predictive of responses to the mass comparison task.

As noted earlier, it is not necessary to give a correct response to part (1) of the mass comparison task (in which the student is asked where the c.m. of the bat as a whole is relative to the fulcrum) in order to answer part (2) correctly, but incorrect answers on the former may reflect attempts to "eyeball" the c.m. location of the bat rather than to rely on the fact that the bat is balanced at that point. Therefore, it is possible that the perceptual difficulty of estimating the location of an object as irregularly shaped as the bat is the main issue, not the comparatively easier task of estimating the c.m. location of the two pieces. To account for this possibility, we can eliminate all those who gave an incorrect answer to part (1), 23% of all students. Of the remaining 338 students, 84 (or 25%) answered the mass comparison task correctly, which is slightly lower than the overall success rate of 35%.

C. Student accuracy with simpler objects

The results for the shaded bar version of the mass comparison task are similar to those for the bat version: 37% gave the correct answer for the shaded bar ($N = 78$) and 35% for the bat (Fisher exact test, two-tailed, $p > 0.4$). Although no student was asked both questions, the samples appear to be comparable based on responses to other, matched questions.

Results for the related T-shaped *object* task are also not directly comparable, but they tell a similar story: 25% of the students responded correctly, while 67% claimed that the c.m. is at the junction between the two equal-mass pieces ($N = 76$).

V. DISCUSSION AND CONCLUSION

These results do not provide strong support for the perceptual hypothesis, at least in explaining results of the mass comparison task (bat). If locating the c.m. of each piece of the bat were part of their strategy, the majority of students would be able to do so accurately enough to be able to answer the mass comparison task. The fact that most answers to the two tasks are not mutually consistent also suggests that locating the c.m. of each piece does not play a role in the reasoning of the majority of students. Even when the object has a more regular shape, similar to those deemed "extended, symmetric" in Sattizahn *et al.*, most students claim that there is the same amount of mass to both sides of the balance point. Finally, when balancing is not an issue, and students must reason about the location of the c.m. from a known mass distribution, rather than the reverse, the most frequent answers still suggest that the c.m. effectively divides an object into pieces of equal mass.

It is noteworthy that in the one example given in Sattizahn *et al.* the incorrect answers cluster at the border between two identical, rectangular portions of the object (see Fig. 2). It may be that what appear to be perceptual difficulties in fact reflect conceptual difficulties.

An account consistent with dual-process theories is that students' initial, "intuitive" response relies on a balancing heuristic—not necessarily *literal* balancing, simply a heuristic in which the at-rest condition is taken as evidence of equal influences counteracting one another. In other words, the fact that the bat is at rest indicates that *something* must be the same to both sides of the fulcrum. (This heuristic could be an instance of a *phenomenological primitive* [12].) In situations in which mass seems unsatisfactory—as in the case of two discrete objects at obviously different distances from the fulcrum—incorporating distance into the quantity assumed to be equal to both sides provides a satisfactory resolution. When distance does not obviously matter, then mass seems satisfactory on its own, and no further analysis occurs. In this account, students simply experience *no need*

to consider modeling the extended object as a collection of point objects. It may be that they are not even aware that this is a legitimate strategy. It is also possible that they are not fully satisfied with their responses but see no alternative. That is, they have no means of including distance into their thinking. The results presented here do not allow us to distinguish between these possibilities.

It is also noteworthy that Sattizahn *et al.* [11] found that performance on their tasks was not strongly influenced by the physics background of their subjects. They interpret this observation as evidence that the tasks draw heavily on non-domain-specific reasoning. We also found that prior physics experience, even in the immediate past, is not a significant factor [9]. This phenomenon is not unique to this context—it is well known in physics that certain incorrect reasoning patterns are observed both before and after formal instruction with roughly equal frequencies [13]. In contrast to Sattizahn *et al.*, we view this as evidence that the underlying topic-specific intuitive ideas do not interact strongly with formal instruction.

If the results presented here do not strongly support the perceptual hypothesis, neither do they rule it out. Additional investigation is needed to determine whether this is a significant issue, even if it affects a relatively small number of students. If so, the implications for instruction are not obvious. One can simplify tasks to reduce the perceptual burden, or provide general support in accurately drawing inferences from sketches, figures, or apparatus. In the recent past, it was important for STEM students to learn to read the intercept of a hand-drawn graph or the level of the meniscus in a fluid-filled beaker accurately. Such skills may be less and less frequently employed in STEM, but lack of ability may present a barrier to students in unexpected ways.

The studies in this article were not intended to determine which perspective on the attribution of errors, domain specific or domain general, is right in a general sense. It is highly probably that both are needed for developing a rich understanding of student thinking. This paper illustrates how this process can play out. The results indicate that domain-general theories of perception and cognition can shed light on responses to domain-specific tasks, and vice versa.

ACKNOWLEDGMENTS

[1] R. Driver, H. Asoko, J. Leach, E. Mortimer, P. Scott, R. Driver, H. Asoko, and J. Leach, Constructing Scientific Knowledge in the Classroom, Educ. Res. **23,** 5 (1994).

[2] D. Hammer, A. Elby, R. E. Scherr, E. F. Redish, D. Hammer, A. Elby, R. E. Scherr, and E. F. Redish, Resources, framing, and transfer, in *Transfer of Learning from a Modern Multidisciplinary Perspective*, edited by J. Mestre (IAP, Greenwich CT, 2004), pp. 89–119.

[3] D. E. Brown and D. Hammer, Conceptual change in physics, in *International Handbook of Research on Conceptual Change*, edited by S. Vosniadou (Routledge, New York, 2008), pp. 127–154.

[4] B. W. Frank, S. E. Kanim, and L. S. Gomez, Accounting for variability in student responses to motion questions, Phys. Rev. ST Phys. Educ. Res. **4,** 020102 (2008).

[5] J. S. B. T. Evans, The heuristic-analytic theory of reasoning: extension and evaluation, Psychon. Bull. Rev. **13,** 378 (2006).

[6] M. Kryjevskaia, M. R. Stetzer, and N. Grosz, Answer first: Applying the heuristic-analytic theory of reasoning to examine student intuitive thinking in the context of physics, Phys. Rev. ST Phys. Educ. Res. **10,** 020109 (2014).

[7] A. F. Heckler, The ubiquitous patterns of incorrect answers to science questions: The role of automatic, bottom-up processes, Psychol. Learn. Motiv. **55,** 227 (2011).

[8] M. Kozhevnikov, M. Motes, and M. Hegarty, Spatial visualization in physics problem solving, Cogn. Sci. **31,** 549 (2007).

[9] L. G. Ortiz, P. R. L. Heron, and P. S. Shaffer, Student understanding of static equilibrium: Predicting and accounting for balancing, Am. J. Phys. **73,** 545 (2005).

[10] L. C. McDermott and P. S. Shaffer, and The PEG at UW, *Tutorials in Introductory Physics* (Pearson, Upper Saddle River, NJ, 2002).

[11] J. R. Sattizahn, D. J. Lyons, C. Kontra, S. M. Fischer, and S. L. Beilock, Physics Education, Perception Matters, Mind, Brain, Educ. **9,** 164 (2015).

[12] A. DiSessa, Toward an epistemology of physics, Cognit. Instr. **10,** 105 (1993).

[13] P. R. L. Heron, Effect of lecture instruction on student performance on qualitative questions, Phys. Rev. ST Phys. Educ. Res. **11,** 010102 (2015).