Using Eye-tracking to Study Intuitive Reasoning on a Kinematics Graph Tasks

J. Caleb Speirs (he/him), Benjamin Sandlin (he/him)
Department of Physics, University of North Florida, 1 UNF Dr, Jacksonville, FL, 32224

Understanding the ways that unconscious information processing occurs can help educators teach in more efficient ways. In any given science task, there are typically a myriad of relevant and irrelevant features that frame and direct reasoning pathways. To better train students to navigate task features productively, more needs to be understood about the ways that attention-attracting task features interact with reasoning processes. Eye-tracking methodologies allow physics education researchers to record where a student allocates their attention, including the initial focus of attention. In this paper, we discuss results from eye-tracking research into a mechanic graph task that has a strong intuitive reasoning pathway that leads to an incorrect conclusion. We analyze data from students in algebra-based physics courses. We find that analysis of the eye-tracking metrics of first look and dwell time are consistent with predictions made from dual-process theories of reasoning.

2025 PERC Proceedings edited by Pawl, Zwolak, and Leak; Peer-reviewed, doi.org/10.1119/perc.2025.pr.Speirs
Published by the American Association of Physics Teachers under a Creative Commons Attribution 4.0 license.
Further distribution must maintain the cover page and attribution to the article's authors.

I. INTRODUCTION

The myriad benefits that students get from taking physics courses has inspired various majors to require introductory physics courses in order to graduate. Being successful in a physics course does, of course, come with many challenges, and students are typically aware of and discuss these challenges with their peers. It is a common experience for students to feel like they understand the solution to a physics question (or concept) generally but fail to answer the physics question correctly in the moment. In some instances, this failure is attributed to missing a detail about the problem statement or unduly focusing on an irrelevant detail and can result in an incorrect or incomplete answer. Indeed, these phenomena have been well studied in PER literature [1, 2, 3].

Physics education researchers have conducted studies to examine how much of a detriment salient distracting features (SDFs) bring to a student's ability to adequately answer a problem [2, 4, 5, 6, 7]. A salient distracting feature is a surface level feature of a problem that draws attention and cues an incorrect or unproductive model or line of reasoning about the problem. Much of this research involves a screening-target methodology, where screening questions are used to gauge the students' understanding of some principal concepts in physics, sometimes called their 'mindware' in the literature, and the target questions gauge their ability to reason with this mindware when SDF's are present. Some examples of mindware studied in the literature are knowledge of how to find an object's relative speed by observing a position vs. time graph [2, 6, 8], as well as the relations between the friction force and Newton's 2nd Law with respect to a stationary box [5, 6].

Our theoretical framework to examine students' behavior is dual process theories of reasoning (DPToR) [3, 9, 10]. Briefly, dual process theories state that a person working out a problem will quickly and unconsciously create a provisional model to solve that problem and will only seek a new model if they have (1) a disposition or a cognitive cue (such as a decreased feeling of rightness) to scrutinize their provisional model, and (2) the mindware to determine if their initial model will suffice and to create a new model if the provisional model is found unsatisfactory [3].

Generally, the literature highlights that even students who demonstrated adequate mindware in the screening questions displayed difficulty with arriving at the correct answer on questions where salient distracting features were present [2, 5, 6]. One cause of this difficulty could be that the distracting feature introduces a cognitive conflict between formal physics knowledge and the intuitive provisional model.

The Stroop effect describes such a phenomenon wherein tasks with competing dimensions (such as text color and text content, or height and slope) cause longer answering times [8, 11]. In the famous example of the Stroop effect, the name of a color is written in a font-color that may or may not be what the text literally reads (for instance, the word "RED" would be written in blue font or in red font), and the task is to verbally speak the color of the font (for instance, "blue"). When the color of the text is aligned with the word itself, this task takes less time on average than when the color and word are not the same. This is due to the cognitive conflict created by the competing dimensions. Importantly, the effect is erased when the words are not comprehensible (such as when a child who knows their colors but is not yet able to sight-read completes the task [11]). In other words, when there is no mindware, there is no conflict. The Stroop effect will be seen to be relevant to the data presented here.

Most of the research regarding SDF's in physics is done by inference from student answering patterns (though response time has also sometimes been used [8]). However, some studies have investigated reasoning patterns with eye-tracking. For instance, Ref. 12 studied student gaze fixations and scan paths and related those data to performance on kinematics problems.

Our study also seeks to use eye-tracking to investigate student reasoning difficulties but focuses on understanding the allocation of attention when SDFs are present. The larger goal of our research is to better understand students' behavior while answering physics questions with SDFs as a means of understanding how best to teach students to navigate situations where their provisional models are incompatible with accepted physics models and to train them to investigate their mindware with greater scrutiny. These are our general and long-term goals.

In this paper, we explore whether a salient distracting feature on a graph task draws significant attention as measured by eye-gaze data. We pose the following research questions.

- Do students who have relevant knowledge look first at the salient distracting feature (SDF) on the graph task?
- Do students (even those who answer correctly) spend significant dwell time on the graph task SDF?

The answers to these questions have implications for the applicability of DPToR to understanding student in-the-moment reasoning on questions with salient distracting features. DPToR would predict that both questions would be answered in the affirmative.

II. METHODOLOGY

All participants in this study were students enrolled in algebra-based physics courses at a primarily undergraduate institution in the southeast. Our sample consists of 63 such students. The university' gender and racial demographics are similar to those of U.S. college-bound first-year students, with a student body that is approximately 60% female and 40% male. Racial and ethnic demographics are majority

White (55%), with notable representation from Hispanic/Latino (16%), Black or African American (12%), and Asian (6%) students, alongside smaller proportions identifying as multiracial, American Indian/Alaskan Native, Native Hawaiian/Pacific Islander, or of unknown race.

The students booked time once a semester to participate as a homework assignment for their course (they could opt out of using their data for research and/or request an alternate assignment unrelated to eyetracking). Students completed the screening and target questions presented in this manuscript alongside one other set of screening/target questions in the context of friction and another kinematics activity unrelated to this research. Generally, these appointments took around 20-30 minutes. On each activity, their eye-movements recorded using a GazePoint eyetracker.

The Gazepoint Eyetracker had a 150 Hz sampling rate, 0.5–1.0 degree visual angle accuracy with 9-point calibration. We used a 24” monitor with the stimuli filling the screen. The participant was typically around 15” away, and there were no movement restrictions, but movement was monitored by interviewer during the tasks and students were prompted to realign when needed. Each student performed the tasks alone using a think-aloud protocol. However, due to the nature of eye-tracking research, the interviewer typically only prompted the student to “think aloud” when confirming their final answer to each question.

Students completed two screening questions and one target question related to finding the velocity from a position vs. time graph. The tasks were drawn from previous research [6] and are shown in Figure 1 and 2.

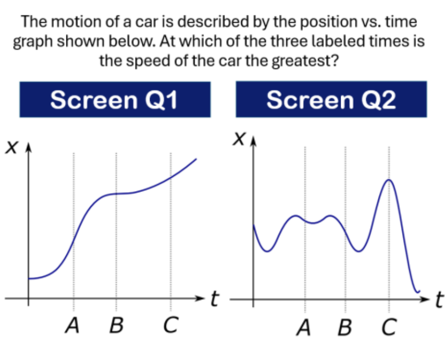

FIG 1. The screening questions used to assess whether students understood how to arrive at the speed of an object when present with a position vs. time graph. The correct answer to screening question 1 is “Time A” and the correct answer to screening question 2 is “Time B”.

We defined areas of interest (AOIs) related to the SDF and relevant features (RF). The RF AOI defined as the point on the graph where the lines are most parallel, and the SDF was defined where the lines intersect. The target question and their defined AOI’s are shown in Figure 2. From each AOI, we derived dwell time measurements for each participant as well as which AOI received the first fixation.

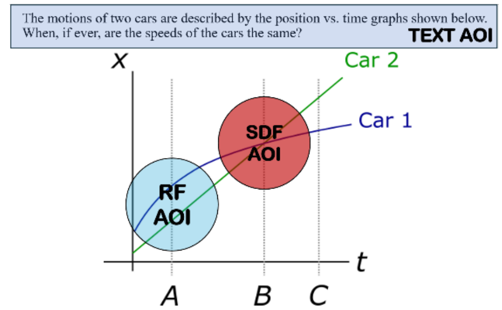

FIG 2. Target graph task used in this study, along with the AOIs defined. The correct answer to is “Time A”, where the slopes are the same (RF for *relevant feature*). The common incorrect answer is “Time B”, where the intersection of the two graphs occurs (SDF).

A *fixation* occurs when eye-gaze rests on a single object for more than 50 ms and is the primary means that visual information is acquired [13]. Dwell time is defined here as the total amount of time that an eye-gaze rests inside the AOI throughout the task, in other words, the sum of the durations of the various fixations on the AOI throughout the task.

III. RESULTS

The students’ overall performance was rather poor. As seen in Figure 3, 15 (24%) students answered the target question correctly. Of the 25 (40%) who answered both screening questions correctly, 11 answered the target question incorrectly, reproducing previous findings [2, 6]. The 27 students who answered incorrectly on the screening questions universally answered the target question incorrectly.

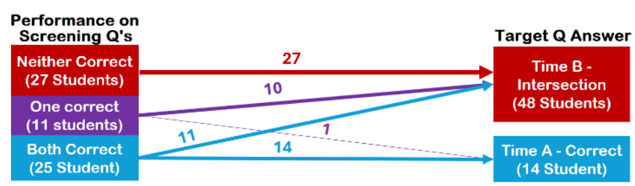

FIG 3. Performance breakdown on the target question. The left-hand side of this diagram shows the number of students with each score on the two screening questions, and the right-hand side shows the number of students who answered the target question correctly (“A”) or incorrectly (“B”).

In every case, when the target question was presented, students first read the question prompt and then examined the graph. Upon viewing the graph for the first time, 56 (89%) students looked at the SDF first, 3 (5%) students looked at the RF first, and another 4 (6%) looked at some other feature (such as the letter labels on the time-axis). Additionally, every student who ultimately answered the target question correctly looked at the intersection point (the SDF) first.

The data also showcased a difference in behavior among students that answered the target graph task correctly compared to those who answered incorrectly (see Fig. 4).

- The total time on task for the target question was longer for those who demonstrated mindware on the screening questions compared to those who did not ($t(52) = 3.33, p = .001$).
- The mean dwell time for the SDF was observably longer than that for the RF for the groups of students who answered the target question incorrectly, regardless of whether they demonstrated the requisite mindware on the screening questions. (See Figure 4.b and 4.c.)
- Those who demonstrated mindware and answered the target question incorrectly spent more time looking at the SDF compared to those who did not demonstrate mindware who answered incorrectly according to a two-sample equal variance t-test ($t(38) = 2.4, p = 0.01$).
- The behavior of those that got all 3 questions correct differs greatly from that of the first two groups. This group spent about an equal amount of time looking at both features with them dwelling on the RF a bit more than the SDF, but not by a statistically significant amount. (See Figure 4.a.)

IV. DISCUSSION

The purpose of this research was to use eye-tracking methodologies to create a more detailed picture of the interactions between mindware and salient distracting features on physics problems. Our research reproduces earlier research that demonstrated that even when adequate mindware is demonstrated, students may fail to employ correct reasoning on physics questions with SDFs. Much of that previous research relied on answering patterns. The data gathered from the eye-tracking methodology used here gave us more to observe about the students' behavior beyond how they answer.

Previous research had posited that performance on the target kinematics graph task was controlled by the tendency to unconsciously produce provisional models based on task features – that is, that students notice the eye-catching feature of the intersection before they notice anything else about the graph, and thus their subsequent reasoning is built off of the intersection point unless that model is deemed unsatisfactory. The eye-tracking methodology presented here allowed us to confirm this suspicion. We saw that the SDF was overwhelmingly what students darted their eyes to first – even among those who ultimately answered the question correctly, and that the SDF drew at least equal but often more attention compared to the relevant feature.

The mean dwell time of the SDF for those who answered all three questions correctly was not statistically different than the dwell time on the RF, even though the SDF was universally the first attention of this group (see Figure 4.a). One might expect that a correct answer would derive from more attention on the relevant feature, and that irrelevant features might be glanced at but not fully considered. However, from a dual-process perspective, a first glance at the intersection point would cue a provisional model related to that point that the student would have to abandon in favor of a conceptually correct model. Therefore, we would expect to see considerable attention given to the SDF as they struggle with whether their provisional related to the SDF is satisfactory or must be abandoned.

Indeed, this same "struggle with the intersection" could be inferred from the dwell time on the SDF for those who demonstrated mindware but answered the intersection answer incorrectly (see Figure 4.b), compared to the mean dwell time on the SDF for those who did not demonstrate mindware (Figure 4.c). If one does not have adequate mindware to fully investigate the correctness provisional model (more likely for those students who do not answer either screening question correctly), then there would be no need to linger on the SDF because there would be no conflict with that answer. Thus, from a dual-process perspective, the

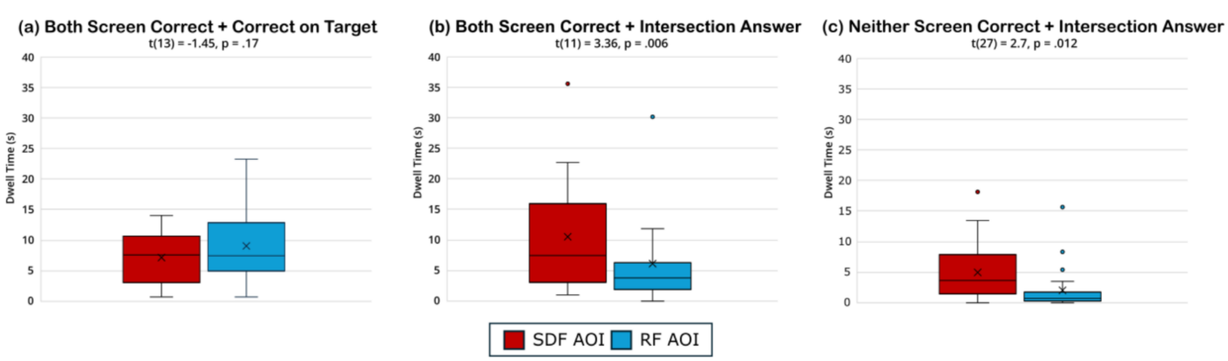

FIG 4: Dwell time data for both AOIs under study for the target graph task broken down into three categories of students, (a) those who answered all three questions correctly, (b) those who answered both screening questions correctly but answered the target question incorrectly, and (c) those who answered all three questions incorrectly. A paired two sample t-test on the means was conducted for each chart, with the test statistic and p-value reported above the chart.

comparison of Figure 4.b and Figure 4.c shows exactly the pattern one would predict for the dwell times.

Our data could be viewed as a Stroop effect in this regard. Recall that the Stroop effect describes a situation where competing dimensions (in this case, “slope” and “height”) cause a delay in answering times, even when the participant is readily able to answer correctly. The intersection point creates a competition between that feature and the relevant slopes, but it would only create this competition for those who understood the slope as relevant to the task. We thus observed that, for students that seemed to understand the relevance of the slope (*i.e.,* those who answered both screening questions correctly), the task took longer, in accordance with a Stroop effect.

Additionally, there is a wider variance for those who answered both screening questions correctly in the RF for those who answered the target question correctly compared to those who answered the target question incorrectly, and the effect is reversed for the SDF (*i.e.*, those who answered target incorrectly had a wider variance compared to those who answered target correctly). This may be evidence of a complex interplay between process 1 reasoning and process 2 reasoning for those who have the relevant mindware to answer the question. This suggests that those features are operating differently in each group, but the exact cause is hard to pin down with our data alone.

V. CONCLUSIONS AND FUTURE WORK

The results of this study demonstrate that the salient feature of the intersection point does indeed draw first and significant attention from students completing the target graph task. Given this as a starting point, dual-process theories of reasoning would predict that student’s reasoning and answering patterns would tend towards the answer associated with the intersection. The behavior of the students in our sample are consistent with this prediction. Reaffirming previous research, many of the students in this study who demonstrate the ability to relate the velocity of an object to the slope of a position vs. time graph fail to do so on a question with a salient distracting feature. Secondly, the mean dwell times on the SDF are consistent with the view that student internal reasoning is tending towards the answer associated with the intersection point, even for those who demonstrated that they know what the relevant feature associated with velocity is for a position vs. time graph. This work, therefore, adds support for the use of dual-process theories of reasoning as a way to study the mechanistic causes of student reasoning and answering patterns.

There are limitations to this study, however. More tasks with clear salient distracting features need to be studied with eye-tracking methodologies, and the type of salience can be varied. For instance, some task features are visually appealing, such as the intersection point in this study. Other tasks have salient features that are more conceptual in nature (such as including coefficients of friction that cue reasoning about the equation $f = \mu N$) and the allocation of attention may follow different patterns on tasks such as these.

By studying the patterns created by student reasoning mechanisms, we hope to eventually shed more light on how curriculum can be devised that helps support the development of robust reasoning skills. Not all salient distracting features are the same. Some are visual, such as the intersection point in the graph task here. Some are verbal, as when specific but irrelevant values are given for the coefficient of friction. Additionally, the impact of the SDF may in some cases cue an incorrect answer independent of physics reasoning, or it may cue specific physics models that then lead to an answer. Once these differences are better understood, one could imagine developing a curricular activities that have students navigate their intuition in productive ways. For instance, one could pose a screening question to students, have them elaborate on the knowledge they have regarding the problem, then pose them a question with an SDF and elicit an intuitive but incorrect response, tell them that it is incorrect, and then prompt them to consider what the formal reasoning they used on the screening question would conclude on the target question. Reviewing multiple questions across the semester in this way may improve student navigation of distracting features.

However, instructors may consider removing SDF’s from questions when trying to assess content knowledge alone. There is still much work needed to understand the patterns of reasoning and answers when SDF’s are present. Eye-tracking methodologies seem to be a productive way to study these mechanisms.

ACKNOWLEDGMENTS

We would like to acknowledge helpful contributions from MacKenzie Stetzer, and Beth Lindsey, Mila Kryjevskaia, Vincent Doan, Grant Hawthorne, and AJ Hutchinson, W. Brian Lane, Andrew Heckler, and members of the PER@UNF group.

This material is based upon work supported by the National Science Foundation under Grant Nos. DUE-2142416, DUE-2141975, DUE-2142276, and DUE-2142436.

[1] D. Hammer, More than misconceptions: Multiple perspectives on student knowledge and reasoning, and an appropriate role for education research, *Am. J. Phys*. **64**, 1316 (1996).

[2] A. F. Heckler, The role of automatic, bottom-up processes: In the ubiquitous patterns of incorrect answers to science questions, in *Psychology of Learning and Motivation* (Academic Press, Cambridge, MA, 2011), pp. 227–267.

[3] M. Kryjevskaia, P. R. L. Heron, A. F. Heckler; Intuitive or rational? Students and experts need to be both. *Physics Today* 1 August 2021; **74** (8): 28–34.

[4] P. R. Heron, Testing alternative explanations for common responses to conceptual questions: An example in the context of center of mass, *Phys. Rev. Phys. Educ. Res.* **13**, 010131 (2017).

[5] M. Kryjevskaia, M. R. Stetzer, and T. K. Lê, Failure to engage: Examining the impact of metacognitive interventions on persistent intuitive reasoning approaches, in *Proceedings of the 2014 Physics Education Research Conference*, Minneapolis, MN (AIP, New York, 2015).

[6] J. C. Speirs, M. R. Stetzer, B. A. Lindsey, and M. Kryjevskaia, Exploring and supporting student reasoning in physics by leveraging dual-process theories of reasoning and decision making. *Phys. Rev. Phys. Educ. Res.* **17**(2) 2021. https://doi.org/10.1103/physrevphyseducres.17.020137

[7] J. C. Speirs, R. Leuteritz, T. K. Lê, R. Deng, and Shawn W. Ell, Investigating the efficacy of attending to reflexive cognitive processes in the context of Newton's second law. *Phys. Rev. Phys. Educ. Res.* **19**, 2023.

[8] A. F. Heckler and T. M. Scaife, Patterns of response times and response choices to science questions: The influence of relative processing time, *Cogn. Sci.* **39**, 496 (2014).

[9] J. S. B. T. Evans and K. E. Stanovich, Dual-process theories of higher cognition: Advancing the Debate, *Perspect. Psychol. Sci.* **8**, 223 (2013).

[10] D. Kahneman, *Thinking, Fast and Slow* (Macmillan, New York, 2013).

[11] AR Jensen, W D Rohwer. The Stroop color-word test: a review. *Acta Psychologica*. **25** (1): 36–93. 1966.

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

[13] K. Holmqvist et. al., "Eye Tracking: A comprehensive guide to methods and measures" (Oxford University Press, Oxford, 2015).