

Naive misconceptions of Cooke and Breedin's research: Response to Ranney

NANCY J. COOKE

New Mexico State University, Las Cruces, New Mexico

and

SARAH D. BREEDIN

Temple University School of Medicine, Philadelphia, Pennsylvania

In this paper, we respond to Ranney's (1994) comment on our paper on naive physics (Cooke & Breedin, 1994). In his comment, Ranney criticizes our methods and analyses, as well as our claims of theoretical inconsistency. In response to Ranney, we show that most of his criticisms can be traced to several misconceptions. Some of these misconceptions seem to stem from his drawing of inappropriate similarities between his own research and ours. Specifically, Ranney seems to hold four misconceptions about our research: (1) the belief that trajectory responses in our study were relevant to our claims about theoretical consistency, (2) the belief that aggregation of theoretical variations weakens claims of inconsistency, (3) the belief that the method of written reports is inferior to the method of structured interviews, and (4) the belief that our proposal of on-the-fly theorizing is at odds with temporal consistency and nonimpetus beliefs.

In an earlier paper (Cooke & Breedin, 1994), we describe two experiments on naive theories of motion that support two main conclusions: (1) performance on motion problems is context sensitive and (2) naive impetus theory is expressed infrequently and inconsistently by subjects and is unrelated to performance. In regard to the first conclusion, Ranney (1994) states that our research adds to the growing literature supporting the anti-theory or post-McCloskeyan view of performance on physics tasks. In general, this is the view that performance on physics tasks is greatly affected by the context in which the problem is presented, calling into question the extent to which theory-like knowledge guides performance. Ranney happens to agree with this position.

Ranney's critical comments are primarily directed toward our second conclusion and specifically toward our claims that subjects apply naive impetus theory inconsistently. Ranney raises several methodological and analytical issues that he thinks are problematic for this claim. In the present paper, we address these issues and show how many of them seem to be based on four basic misconceptions that Ranney has about the research presented in Cooke and Breedin (1994). In general, Ranney seems to overassociate our work with his research, thereby drawing parallels between the two that do not exist. In particular, Ranney's four major misconceptions are (1) the be-

lief that trajectory responses in our study were relevant to our claims about theoretical consistency, (2) the belief that aggregation of theoretical variations weakens claims of inconsistency, (3) the belief that the method of written reports is inferior to the method of structured interviews, and (4) the belief that our proposal of on-the-fly theorizing is at odds with temporal consistency and nonimpetus beliefs. These are discussed in turn below.

MISCONCEPTION 1: Trajectory Responses Are Relevant to Theoretical Inconsistency Claim

One of Ranney's central criticisms revolves around the fact that we "pigeonholed" trajectory responses of our subjects into four categories (Experiment 1) or two categories (Experiment 2). Ranney believes that such data aggregation is problematic for our claims of theoretical inconsistency. Specifically, he argues that the response categories are so coarsely aggregated that chances of finding consistency seem questionable. For instance, he notes that we "combine a curvilinear-impetus response for the tube problem with a 'straight-down response' for the cliff problem"—two different varieties of impetus. Ranney proceeds to point to his own research (see his Appendix), which employs a cataloging procedure that is purportedly more sensitive to subjects' theories. In general, this criticism seems to stem from Ranney's misconception about the relevance of our categories to our claims of theoretical inconsistency. Before addressing Ranney's specific position, we review our reasons for aggregating trajectory responses in the way that we did.

Although we did categorize trajectory responses, we did not use, nor did we intend to use, these response cat-

We would like to thank Anna Rowe and Laura Thompson for their feedback on earlier versions of this manuscript. Correspondence should be addressed to Nancy J. Cooke, Department of Psychology, New Mexico State University, P.O. Box 30001/3452, Las Cruces, NM 88003.

—Accepted by previous editor, Margaret Jean Intons-Peterson

egories to make inferences about subjects' theories. As we explain in the paper (Cooke & Breedin, 1994), such categorization was done so that the production (i.e., draw-a-trajectory) and selection (i.e., select-a-trajectory) conditions could be equated in terms of alternative responses available to the subjects. We did not attempt to infer subjects' theories from this categorization scheme. Instead, we used the categories to generate accuracy scores. Because of this goal, it was important that the categories be representative of common responses and that they vary in degree of accuracy (see McCloskey & Kohl, 1983, for a similar approach). In this way, we could account for as many drawings as possible and, at the same time, create a good selection test. In fact, using our categories, we were able to account for all but 2% of the drawings in Experiment 1 and all but 3% in Experiment 2. As a result of this *empirical* rather than *theoretical* orientation to category generation, we undoubtedly omitted categories that were theoretically distinct but infrequently produced. In addition, as Ranney points out, our foils were not maximally divergent and our response typologies were not isomorphic across problems. Although this is not the place to discuss test construction, it is our belief that good multiple-choice tests contain distractors similar to one another and/or the target and that the best distractors are based on common errors. At any rate, although other response categories are surely possible, we only wish to argue that our response categories accounted for the majority of subjects' drawings, varied in degree of accuracy, and were equivalent across conditions. Thus, in our analysis, straight-down cliff responses and curvilinear tube responses were given equivalent accuracy scores. This does not, however, imply that the two types of responses are theoretically equivalent.

Ranney seems to believe that we used trajectory responses as an index of theoretical consistency. This misconception is demonstrated by his comment that "cutting the alternatives to two (C&B's Experiment 2) may lower the odds of finding consistency, as more subtly different paths are theoretically miscategorized/pigeonholed." Beside the fact that Ranney seems to be confused about our reasons for reducing the number of alternatives (explained in Cooke & Breedin, 1994), their number, like the alternatives themselves, is irrelevant to the issue of theoretical consistency. The small drop in explanation consistency between Experiments 1 and 2 (1% vs. 5%) that Ranney notes has nothing to do with the reduced number of trajectory alternatives. Instead, this difference is most likely due to the more lenient criterion for explanation consistency used in Experiment 1 (two out of four problems given similar written explanations) than in Experiment 2 (two out of three). Although we implied that reducing the number of response categories should increase response consistency, we were not referring to theoretical consistency, but rather to consistency in response accuracy across varying contexts. This position is a prime example of Ranney's drawing inappropriate parallels between his research in which trajectory responses were used to make inferences about theoretical consistency and our re-

search in which such responses were used to investigate accuracy and context sensitivity.

This misconception is again indicated where Ranney contrasts our categorization of responses that we used to measure response accuracy to qualitative measures such as interviews and verbal protocols, which have been used to measure theoretical consistency. Ranney also proposes that each of our response categories should be mapped onto a specific theory or misconception. As mentioned above, this practice would be counter to our goals of constructing a good test of trajectory accuracy. Thus, from the perspective of our response categories, Ranney seems to confuse our goal of measuring response accuracy with his goal of measuring theoretical consistency.

Interestingly, Ranney himself acknowledges that "as the sortings are based on physical—not conceptual—similarity, the paths are seemingly categorized largely on accuracy." Indeed, the paths were categorized *completely* on the basis of accuracy. Although Ranney seems to recognize the nature of our categories, he apparently misunderstands that the reason for such categorization was to address the question of the sensitivity of accuracy to context, not theoretical consistency.

Misconceptions aside, the fact that we did not choose to use trajectory responses to make claims about theoretical consistency can also be questioned. As Ranney himself points out, *trajectory drawings* are ambiguous in terms of theory. For example, he states that one might draw a diagonal trajectory for several reasons. A single trajectory drawing may be interpreted as stemming from a number of theories or beliefs. For this reason, as well as because only half of the subjects had the freedom to draw a trajectory (the rest selected one from a set of alternatives), we chose not to make inferences about theories from trajectory responses. Ranney claims to be able to disambiguate subjects' trajectory drawings via their verbal reports. However, because verbal reports are often ambiguous (see below), we do not view this as a satisfactory solution to the theoretical interpretation of subjects' drawings.

MISCONCEPTION 2:

The Aggregation of Theoretical Variations Weakens Claims of Inconsistency

Most of Ranney's complaints of "inappropriate data aggregation" implicate our trajectory response data. As discussed above, these data were not relevant to our claims of theoretical inconsistency. In contrast, subjects' explanations of their trajectory responses *were* relevant to these claims. Unlike the categorization of trajectory responses, explanations were categorized in a post hoc fashion, with a specific model in mind. Specifically, we were interested in identifying any and all instances of naive impetus theory. Ranney is concerned that, by analyzing the data in this way, we may have overlooked other theoretical variations and, in doing so, increased our chances of finding inconsistency. This concern seems to reveal another misconception held by Ranney, on the basis of drawing erroneous associations between his research goals and ours.

That is, Ranney seems to have the erroneous belief that we set out to investigate general theoretical consistency, as opposed to the consistency of a single theory—namely, impetus.

Because we were interested in investigating the prevalence and consistency of impetus theory, we categorized explanations with this model in mind. In particular, we set a very low criterion for categorizing an explanation as one of the impetus variety. To do this, we grouped a variety of types of terms and statements into the impetus category. These terms and statements were based on descriptions and examples of impetus theory found in the published literature.¹ If we did not make the impetus category as broad as possible, any claims of inconsistency could have been attributed to an overly narrow definition of impetus. Indeed, on the basis of some early comments from reviewers along these lines, we broadened our impetus category to include any mention of “forces overcoming one another” and “momentum.”

One could nonetheless argue, as Ranney does, that some of our nonimpetus errors (e.g., our “adds velocity” and surface errors) may actually be covert manifestations of impetus. (Interestingly, in this case Ranney is arguing that our impetus category is *not broad enough*.) However, we see no reason to suspect that our categorization of responses was systematically biased against finding impetus. In fact, the intention was to bias the categorization in the opposite direction, making it more, not less likely, that we would find impetus.

Analyzing explanations into different and finer categories as Ranney suggests, while interesting, would only provide even stronger support for our position. That is, if we found few impetus theorists when impetus was defined broadly, we would find even fewer theorists holding a consistent curvilinear impetus view. We do not, as Ranney suggests, reject the notion that there may be a number of varieties of impetus theory (or other naive theories for that matter); however, if we cannot find prevalence or consistency when grouping these varieties, we will certainly find less prevalence and consistency when varieties are regarded separately. (We are unconvinced by Ranney’s interesting conjecture that fewer categories lead to greater inconsistency.) This concern of Ranney’s seems to stem from a confusion between the goals of our research and his research. We set out to investigate naive impetus theory; he set out to investigate general theoretical consistency.

MISCONCEPTION 3: Written Reports Are Inferior to Structured Interviews

Ranney criticizes our method of collecting written explanations (see also Champagne, Klopfer, & Anderson, 1980; McCloskey & Kohl, 1983) compared with his and others’ use of structured interviews. We believe that because different methods trade off with one another in terms of costs and benefits, good science should approach a problem using a variety of methods that hopefully con-

verge on the same conclusion. Ranney seems to subscribe to the unilateral superiority of the structured interview technique and attributes any findings at odds with his beliefs to the use of a different technique. For instance, Ranney attributes the relatively large number of explanation omissions and descriptives that we found to shortcomings of written reports, rather than the actual state of subjects’ knowledge. In his singular acceptance of the structured interview technique, Ranney seems to have overlooked many of the positive features of written reports and negative features of structured interviews.

Ranney criticizes our method of obtaining subjects’ explanations through written reports on several grounds. One of his biggest concerns is that subjects will tire of writing and will therefore provide impoverished explanations, compared with those obtained with a technique that uses oral reports. First, our written reports were not necessarily brief. Many subjects wrote lengthy explanations. As a consequence, our protocol data, like Ranney’s, were time consuming to analyze. However, much of what was written by subjects simply repeated the problem statement or described the problem diagram. Therefore, subjects did not seem to be getting tired of writing. They simply did not know what to write. At the same time, we would argue that concerns about subjects’ fatigue are relevant to structured interviews as well. Surely, subjects and even well-trained experimenters would tire and lose their vigilance during the 3-h interviews that Ranney conducted.

Ranney praises the structured interview for its ability to disambiguate subjects’ responses. We do not question the proposal that a lengthy structured interview, complete with experimenter prompts, will disambiguate subjects’ responses. We do question whether such a method will disambiguate it in an unbiased way. Unfortunately, this question cannot be answered without access to more detail about the structured interviews that were carried out (i.e., interviewer prompts, questions, transcripts, coding schemes). This point is also elaborated below.

Although we recognize that there are potential costs to written reports, we also are aware of benefits. For instance, using our method of written reports, we were able to test three to five times as many subjects as Ranney tested using interviews. In fact, we are unaware of any other published paper on naive physics that reports analyses of the explanations of so many subjects.

Finally, although there may be cases in which we could have learned more from our subjects by careful prompting, we have no reason to believe that impetus errors were uncovered by written reports less so than other errors. In fact, we tried to err in the other direction by setting a low criterion for an impetus explanation. In summary, both written reports and lengthy structured interviews have the potential of wearing down even the best subjects and experimenters. Also, contrary to Ranney’s claim, oral reports are often retrospective. Furthermore, there is a tradeoff between the disambiguation that results from the additional information gleaned from structured interviews and the unbiased collection of large data sets possible with written reports.

It is worth noting that, in reports of naive physics research in which structured interviews have been used (Clement, 1982; Halloun & Hestenes, 1985; Kaiser, Proffitt, & McCloskey, 1986; McCloskey, 1983), the method is never spelled out in sufficient detail to replicate. Ranney's vague reporting of the methodology for his "fairly nonintrusive structured verbal-protocol session" typifies this state of affairs. Typically, there is no indication of what went on during the interviews. What questions were subjects asked? How were they prompted to give more information? Also, how were these data analyzed? At best, reports of the structured interviews used in naive physics research contain little more than selected excerpts from the interview. It seems futile to compare the method of written reports with one that is not described in sufficient detail to replicate.

Along similar lines of reasoning, Ranney criticizes the true/false questions that we used in Experiment 2. He seems to think that we regard them as a reasonable method for eliciting theories. We do not. It is for this reason that we do not conclude that the $-.48$ correlation between trajectory accuracy and impetus true/false responses indicates that impetus is related to accuracy. On the other hand, Ranney seems to associate the true/false questions with structured interviews when he concludes that the relative paucity of written impetus explanations, relative to true/false questions answered in an impetus manner, indicates that written explanations are inferior to a structured interview.

Our point in using these true/false questions was to demonstrate that we too could elicit more impetus theories with the use of carefully constructed, albeit misleading and often ambiguous, questions. Thus, we agree with Ranney that these questions are poorly worded and ambiguous. This is exactly our point. If those who conduct structured interviews were to ask these types of questions, more impetus responses would be elicited, even from experienced subjects (i.e., MIT graduate students in physics). Even the most experienced subjects can be tricked or misled by the specific wording of questions. Again, this potential feature of structured interviews makes it imperative that researchers report the specific questions and prompts that they use.

MISCONCEPTION 4: On-the-Fly Theorizing Is at Odds With Temporal Consistency and Nonimpetus Beliefs

Although Ranney agrees with a context-specific responding or a "loose reasoning" perspective, he is less apt to accept our proposal of on-line reasoning. Ironically, we feel that context-specific responding necessitates some form of on-line reasoning. Also, Ranney seems to misunderstand not only the basis for our proposal but also its implications. Specifically, Ranney believes that our proposal is at odds with the existence of beliefs or theories and temporal consistency.

Although theoretical inconsistency is compatible with on-the-fly theorizing, these findings do not provide the

strongest evidence for this view, as Ranney suggests. That is, people could be inconsistent because they are inconsistently applying several well-formed theories, perhaps theories that are independent of the experimenter's theories. However, the fact that people are sensitive to subtle contextual features in the environment suggests that a more dynamic interplay between knowledge and the environment is taking place. In fact, we find it difficult to account for context-sensitivity results without postulating some form of on-line reasoning.

In addition, the suggestion that subjects can create theories on the fly does not rule out the existence of well-formed beliefs or theories. We found little evidence for impetus theories, but we have no problem with Ranney's suggestion that subjects may be consistently applying organizing principles other than impetus theory. Interestingly, he at the same time points out that, thus far, no model—including naive impetus—has been proposed that can account for the data. We claim that subjects use existing knowledge and beliefs, coupled with context, to construct responses.

Ranney also apparently believes that temporal consistency is at odds with on-the-fly theorizing. He discusses an unpublished master's thesis (Hojnacki, 1988) in which subjects were found to be remarkably consistent in their responses to physics problems after a delay of nearly a month. Ranney attributes this to "somewhat stable personal 'theories,'" not to on-line theorizing. Unfortunately, the finding of temporal consistency cannot discriminate between these two positions. Subjects should also be expected to be consistent over time if on-line theorizing is occurring, given that they are presented with the same problem and contextual features. To the extent that the context and knowledge is preserved over time, responses should be equivalent.

Finally, although the issue of theoretical consistency is interesting, it is probably not the most important issue from a pragmatic perspective. Even if, as Ranney suggests, some subjects are found to have stable personal theories, it may still be the case that subjects fail to apply or consistently apply these theories to problems.² That is, subjects may be so swayed by contextual features in many instances that their beliefs are swept under the table. In these cases, knowing that an individual holds a particular theory has little pragmatic utility, and intervention to correct the naive theory would have little effect. Thus, the importance of uncovering consistent beliefs or theories seems to hinge on the relative strength of context effects, relative to those beliefs.

CONCLUSION

In summary, Ranney raises several issues that he thinks are relevant to our claims of theoretical inconsistency. Unfortunately, Ranney seems to have attributed many of his research goals and methods to our research, thereby raising a number of irrelevant issues. In his study, Ranney used subjects' trajectory responses to investigate theoretical consistency. We used trajectory responses to investigate

context sensitivity. Ranney investigated general theoretical consistency. We focused on theoretical consistency in regard to naive impetus theory. Besides confusing our goals with his, Ranney dismisses our methodology (written reports) in favor of his (structured interviews). In general, we understand that there are tradeoffs involved in selecting a method, and, for these reasons, we see merit in pursuing a question using a variety of methods. Finally, Ranney seems to agree with our context-sensitivity findings that demonstrate that people attend to environmental cues when responding to naive physics problems, yet he rejects our on-line theorizing view and misunderstands its implications.

Ranney's research, and the other related research that he presents, is interesting and not at odds with our conclusions. Ranney generally supports the loose reasoning perspective and acknowledges findings of context sensitivity. His work on theoretical consistency takes a broad look at subjects' theories of motion, going beyond naive impetus theory and asking whether there is any theoretical consistency of any form.³ As discussed above, our proposal that theorizing is done on line is not at odds with the existence of knowledge in the form of theories and beliefs. What we find interesting is the relative importance of theories and beliefs versus context in these types of judgments.

REFERENCES

- CHAMPAGNE, A. B., KLOPFER, L. E., & ANDERSON, J. H. (1980). Factors influencing the learning of classical mechanics. *American Journal of Physics*, **48**, 1074-1079.
- CLEMENT, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, **50**, 66-71.
- COOKE, N. J., & BREEDIN, S. D. (1994). Constructing naive theories of motion on the fly. *Memory & Cognition*, **22**, 474-493.
- FISCHOFF, B. (1982). For those condemned to study the past: Heuristics and biases in hindsight. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 335-351). Cambridge: Cambridge University Press.
- HALLOUN, I. A., & HESTENES, D. (1985). Common sense concepts about motion. *American Journal of Physics*, **53**, 1056-1065.
- HOJNACKI, S. K. (1988). *Consistency in naive physical reasoning*. Unpublished master's thesis, University of Pittsburgh, Learning Research and Development Center.
- KAISER, M. K., PROFFITT, D. R., & McCLOSKEY, M. (1986). Development of intuitive theories of motion: Curvilinear motion in the absence of external forces. *Developmental Psychology*, **22**, 67-71.
- McCLOSKEY, M. (1983). Naive theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 299-324). Hillsdale, NJ: Erlbaum.
- McCLOSKEY, M., & KOHL, D. (1983). Naive physics: The curvilinear impetus principle and its role in interactions with moving objects. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 146-156.
- RANNEY, M. (1994). Relative consistency and subjects' "theories" in domains such as naive physics: Common research difficulties illustrated by Cooke and Breedin. *Memory & Cognition*, **22**, 494-502.

NOTES

1. In both of our experiments, we based our materials on those that have been traditionally used and accepted in the published literature on naive physics. Undoubtedly, different materials (i.e., definitions of impetus, problem statements, problem diagrams) could be examined. Also, it may well be that some of the traditional materials are flawed. However, our goal was not to address this aspect of the literature. Instead, we were interested in context specificity and impetus prevalence/consistency as demonstrated using traditional materials. If we had attempted to address these issues using definitions or problems different from those traditionally used, proponents of impetus would have probably attributed our results to such differences.

2. Given that Ranney acknowledges dissociations between actions and conceptions, we find it surprising that he attributes the dissociations between trajectory explanations and responses that we report to methodological flaws. In addition, the odds that Ranney reports for being wrong given an impetus explanation versus being wrong given no impetus explanation are misleading. First of all, the probability of being wrong given an impetus explanation is nearly equal to the probability of being right given such an explanation (49% vs. 51% in Experiment 1 and 44% vs. 56% in Experiment 2). To us, this result indicates that the identification of an impetus explanation is not diagnostic of performance on the trajectory problems. Also, in Experiment 2, there is a 44% chance of being wrong given an impetus explanation and a 22% chance of being wrong given no impetus explanation. However, the latter probability is deflated due to the fact that correct explanations, which in our analyses preclude incorrect responses, are included. We do not doubt that the presence of correct versus incorrect explanations is predictive of performance. The question is whether understanding the nature of subjects' explanation errors is predictive of performance. Thus, in Experiment 2, the probability of being wrong given incorrect explanations other than impetus is 34% (vs. 44% wrong given impetus), and in Experiment 1 this probability is 52% (vs. 49% wrong given impetus). In sum, there is not much predictive power in impetus explanation errors, relative to other explanation errors.

3. Although we agree with Ranney that subjects in naive physics studies are theoretically inconsistent, we do not agree that such inconsistency is to be expected. To us, this prediction of Ranney's exemplifies the hindsight bias or the "I knew it all along" phenomenon (Fischhoff, 1982). We see no reason to presume inconsistency. Because many aspects of the environment are not chaotic but predictable, there are many situations in which people can and do behave in a consistent fashion. In fact, human-factors engineers take advantage of such consistency in their design of machines that behave in predictable ways (e.g., turn knobs clockwise to increase volume).