

Constructing naive theories of motion on the fly

NANCY J. COOKE

New Mexico State University, Las Cruces, New Mexico

and

SARAH D. BREEDIN

Temple University School of Medicine, Philadelphia, Pennsylvania

People often make erroneous predictions about the trajectories of moving objects. McCloskey (1983a, 1983b) and others have suggested that many of these errors stem from well-developed, but naive, theories of motion. The studies presented here examine the role of naive impetus theory in people's judgments of motion. Subjects with and without formal physics experience were asked to draw or select from alternatives the trajectories of moving objects that were presented in various manners. Results from two experiments indicate that both trajectory judgments and explanations were affected by specific response and display features of the problem. In addition, these data provide little evidence that naive impetus theory plays a significant role in subjects' performance; instead, they suggest that motion judgments and explanations are constructed on the fly from contextual cues and knowledge that is not necessarily naive.

Why do people have difficulty in understanding formal principles of physics? After all, these principles describe the properties of objects with which we interact daily. Yet to master the principles, laws, and formulas of physics is no simple task.

Some researchers have suggested that such difficulties stem from basic misconceptions that people have about physical systems prior to any formal training in physics (Caramazza, McCloskey, & Green, 1981; Clement, 1982; Halloun & Hestenes, 1985; McCloskey, 1983a, 1983b; McCloskey, Caramazza, & Green, 1980). Support for the presence of some of these misconceptions comes from studies in which people are asked to predict the path of a moving object. For instance, many individuals (51%) will predict that a ball that is shot out of a coiled tube (laying flat on a table) will, upon emerging from the tube, continue to follow the same curved path, rather than a

straight-line trajectory (McCloskey et al., 1980). Similarly, some individuals (36%) will predict that a ball dropped from a moving airplane will fall straight down, as opposed to following a parabolic arc (McCloskey, 1983b). The latter error has been attributed to a particular misconception called the "straight-down belief" (McCloskey, Washburn, & Felch, 1983).

The fact that some of the same errors are made by different individuals and on different experimental problems suggests that the misconceptions that underlie them are not random or idiosyncratic. Some investigators have proposed that the errors are instead based on a more general system of beliefs or a "naive theory" (Caramazza et al., 1981; Clement, 1982; McCloskey, 1983a, 1983b). Similar systems of naive beliefs have been proposed for other domains (Kempton, 1986; Vosniadou & Brewer, 1992), although the focus of this paper is on the role of naive theories in motion judgments. Naive theories of motion have been described as systematic, general, coherent, well-developed, and well-articulated conceptions of motion that conflict with the principles of Newtonian mechanics, but that nonetheless are adequate in many situations (McCloskey, 1983b). On the other hand, they may lead to negative transfer when it comes to learning the formal Newtonian principles.

Support for naive theories of motion has been found in individuals' verbal reports on their line of reasoning in trajectory prediction problems. McCloskey (1983b) had subjects think aloud while solving motion problems and later extensively interviewed them about their answers. According to McCloskey (1983b), 11 of the 13 subjects possessed a naive "impetus" theory of object motion, similar to the pre-Newtonian theory of impetus. According to this theory, "the act of setting an object in motion imparts to the object an internal force or 'impetus' that

This work was completed under NASA Contract NAS9-17900 and was carried out at the NASA Johnson Space Center's Human-Computer Interaction Laboratory and Rice University, Houston, TX. The authors would like to thank Douglas Gillan for his valuable contributions to the conceptual underpinnings of these studies and for his helpful input on early drafts of the paper. We would also like to thank Marianne Rudisill, manager of the Human-Computer Interaction Lab at NASA/Johnson Space Center, for her input on the first experiment and for her thoughtful comments on the paper. Thanks also to Laura Yaffee and Laura Magee, who assisted with the preparation of the stimulus materials and subject testing for Experiment 1, and to Amy Sayles and Thea Trötscher, who assisted with subject testing and data analysis for Experiment 2. Finally, we are grateful to Ben Chase, Mary Czerwinski, Lee Gugerty, Tim McKay, Henry Roediger, and Mike Watkins, who carefully read and commented on various drafts of the paper. Correspondence should be addressed to N. J. Cooke, Department of Psychology, New Mexico State University, P. O. Box 30001/3452, Las Cruces, NM 88003. (e-mail: cooke@nmsu.edu).

—Accepted by previous editor, Margaret Jean Intons-Peterson

serves to maintain the motion. Second, the theory assumes that a moving object's impetus gradually dissipates (either spontaneously or as a result of external influences), and as a consequence the object gradually slows down and comes to a stop" (p. 306). Thus, misconceptions that seem to involve an internal force, dissipation of some force, or a curvilinear internal force have been attributed to this naive impetus theory of motion.

Taken together, these errors and naive impetus theory paint a rather bleak picture of people's understanding of motion, in spite of the fact that their everyday interactions with moving objects do not seem to be too impaired (but see McCloskey, 1983b, and McCloskey & Kohl, 1983, for anecdotal evidence to the contrary). Even without formal physics training, individuals do not behave as if naive theories govern their perceptual and motor behavior completely. Intuitively, people seem to know more about motion than experimental results suggest. Our aim in this paper is to examine errors in judgments and explanations of motion further and to evaluate the evidence for naive impetus theory.

Do people who are untrained in formal physics possess a naive impetus theory of motion that influences their trajectory judgments? The existence of naive impetus theory has been based largely on anecdotal evidence and verbal reports from 13 individuals and isolated problems (McCloskey, 1983b). How systematic, consistent, and general is naive impetus theory? How relevant are such naive beliefs to trajectory judgments? A related question is whether untrained individuals possess any correct knowledge of object motion. One possibility is that people have correct knowledge that they fail to apply or access in all situations.

Perhaps the artificiality of the experimental situation makes it difficult for people to see the relation between the problem and a corresponding event about which they have experiential knowledge. Indeed, in many of the studies in this area, researchers have used static presentations of problems that are unfamiliar and abstract (e.g., a ball shot through a coiled tube). In support of this contention, Kaiser, Jonides, and Alexander (1986) found that a subject's familiarity with the stimulus affected predictions of object motion. For instance, fewer errors were made on a version of the tube problem in which the tube was depicted as a coiled garden hose with water running through it as opposed to the standard abstract version of the tube problem. Kaiser et al. (1986) proposed that subjects draw on specific experiences in order to make judgments about object motion and attempt to match the problem to a relevant situation in order to do so. Thus, familiar problems evoke situation-specific knowledge much better than abstract problems and consequently result more often in correct solutions. According to Kaiser et al. (1986), if a problem does not elicit any situation-specific knowledge, individuals must base their responses on their formal knowledge of physics, which may be either correct or naive.

Not only is performance affected by problem familiarity, but it is also affected by the way the problem is dis-

played. Shanon (1976) found that a visual presentation of motion problems resulted in more correct (i.e., Newtonian) responses than did a verbal presentation of the problems. Kaiser, Proffitt, and Anderson (1985) found that a moving problem display resulted in better performance on unfamiliar object motion problems than a static display of the same problem did. These results have been replicated for problems involving collisions and lifting events (Kaiser & Proffitt, 1984). Additionally, in some preliminary work, diSessa and Globerson (reported in diSessa, 1988) indicate that age effects present in the traditional trajectory prediction paradigm disappeared when subjects were required to judge the plausibility of a trajectory presented in a computer simulation.

Although the results from these studies seem to indicate that the errors that people make on these problems may be due to the artificiality of the laboratory situation, other studies suggest that the errors can occur in more realistic situations. For instance, McCloskey and Kohl (1983) found that 25% of their subjects pushed a puck through a curved tube in a manner indicative of curvilinear impetus beliefs. Similarly Kaiser, Proffitt, and McCloskey (1985) found that children show evidence of the straight-down belief in situations in which a real ball is dropped from a moving toy train. To summarize, there is some evidence that people who are untrained in physics have correct knowledge about motion that they fail to apply or access in unfamiliar situations, but there is other evidence that these errors persist even for familiar and concrete tasks.

Regardless of whether or not untrained individuals have correct situation-specific knowledge about motion, results from these studies do not rule out the possibility that a naive impetus theory is responsible for the cases in which errors are made. Proponents of both views have argued that naive impetus theory can account for many of the errors made on abstract or unfamiliar problems. Can errors on abstract problems be attributed to factors other than naive impetus theory? This question was addressed in the two experiments reported in this paper by systematically manipulating various response and display factors and examining their effects on trajectory judgments and explanations. To the extent that these responses are based on coherent, systematic, and general theories (naive or formal), it was hypothesized that problem features that are irrelevant to the theories should have little or no effect on trajectory responses. Moreover, a systematic examination of explanations associated with erroneous responses should reveal the role that naive impetus theory plays in such errors.

One important task characteristic that has varied throughout studies of motion judgments, but that has not been manipulated systematically, is the type of trajectory response. In studies of the effect of moving displays (Kaiser & Proffitt, 1984; Kaiser, Proffitt, & Anderson, 1985; McCloskey & Kohl, 1983), subjects had selected their responses from two or more complete trajectories, whereas in most other studies subjects had drawn trajectories (e.g., McCloskey, 1983b; McCloskey et al., 1980) or predicted the path of a concrete object (Kaiser, Proffitt, & McCloskey, 1985; McCloskey & Kohl, 1983). The presentation of

alternatives in the selection paradigm provides additional cues in the form of alternatives that may serve as a more sensitive test of knowledge, although it has not been directly compared with the drawing or production paradigm. Thus, in the present experiments response type (production vs. selection)¹ was systematically manipulated. If naive impetus theory is responsible for erroneous trajectory judgments, it should lead subjects to either draw or select the trajectory most congruent with that theory. On the other hand, responses in the two paradigms may differ to the extent that the type of response influences judgments.

Contextual features of the problem itself may also affect subjects' problem representations and subsequent solutions, as has been found for a number of other tasks such as choice behavior (Tversky & Kahneman, 1981), logical reasoning (Johnson-Laird, Legrenzi, & Legrenzi, 1972), and question answering (Feldman & Lynch, 1988; Kreitler & Kreitler, 1981). In fact, context effects have been found in solutions to physics problems similar to those discussed here (Anzai & Yokoyama, 1984; diSessa, 1988; Hojnacki, 1988; McAfee & Proffitt, 1991). One type of contextual feature that has been previously manipulated is whether the display is static or moving. This contextual feature may enable subjects to rely more on situation-specific knowledge, as opposed to naive or formal theoretical knowledge. Thus, in order to investigate the effects of familiarity further, the display type (motion or no motion) was manipulated in Experiment 1.

In addition, there are various problems that embody the same physics principles, but that are associated with different diagrams and verbal descriptions. For instance, the trajectory of a ball rolling off of a cliff involves a combination of constant horizontal velocity and vertical acceleration due to gravity. The same combination occurs when the string is cut at the bottom of a pendulum's swing. A theory that is general should apply to all problems isomorphic with respect to that theory. Consequently, in each experiment three or more problems of object motion were presented to each subject, some of which were isomorphic with each other.

Even within the same problem, subtle features of the diagram can be manipulated that may affect trajectory judgments. For instance, subjects may focus on the dominant feature of the coiled tube—the curve—and may respond with a curvilinear trajectory, simply because it creates a good gestalt. In contrast, the curve may not be as dominant in a picture of a tube with a straight segment appended to the initial opening (see Appendix A, first row, middle diagram). Similarly, the coiled tube problem is typically drawn so that the opening from which the ball emerges is on the lower or right pointing "upward" (see Appendix A, first row, left diagram). Although the tube is supposedly lying flat, this feature may encourage curvilinear trajectories over a tube in which the opening is on the lower left or pointing "downward" (see Appendix A, first row, right diagram). Thus, some of the common errors that subjects make may be the result of demands of specific features of the problem, rather than a consistent naive theory of motion. Therefore, in

these studies various versions of the same problem were presented to each subject. Specifically, two versions, labeled *perceptual set* and *orientation*, were compared with the standard versions of each problem. *Perceptual set* refers to specific perceptual features of the problem drawing (e.g., a completely coiled tube) that provide a context for subsequent interpretations of the drawing (Leeper, 1935), and in this case, for subsequent trajectory predictions. *Orientation* refers to the spatial arrangement of the stimulus or component of the stimulus (e.g., tube exit pointing up or down).

In summary, this research addresses the extent to which "naive errors" pervade motion judgments across various response and display problem features. More importantly, the explanations given for these judgments should indicate whether a naive impetus theory of motion is indeed responsible for the errors that are made and may suggest an alternative account. We propose that the errors that are made on motion problems can be explained without relying on naive impetus theory.

EXPERIMENT 1

In Experiment 1, effects of response type, display type, problem, and version on judgments of object motion were investigated. In addition, formal physics experience was measured and taken into account. The extent to which task characteristics influence trajectory judgments should provide an indication of the generality, consistency, and systematicity of subjects' beliefs. In addition, a systematic analysis of explanations should clarify the role of naive impetus theory in these judgments.

Method

Design. Response type (production or selection), display type (motion or no motion), physics experience (experienced or inexperienced), problem (tube, cliff, pendulum, and rocket), and version (standard, perceptual set, and orientation) were manipulated in a $2 \times 2 \times 2 \times 4 \times 3$ factorial design. Problem and version were within-subjects factors. All other factors varied between subjects.

Subjects. Sixty individuals participated in this experiment. Forty were students enrolled in undergraduate psychology classes at Rice University who voluntarily participated in order to partially fulfill course credit. An additional 20 employees of Lockheed-ESC and the NASA/Johnson Space Center in Houston, TX participated voluntarily. The subjects were randomly assigned to one of four conditions: (1) production-motion, (2) production-no motion, (3) selection-motion, and (4) selection-no motion, with the constraint that there be an equal number of subjects in each condition.

On a postexperiment questionnaire, the subjects were asked how much formal physics training they had had. The questionnaire was administered after the experiment so that it would not prompt any subjects to respond to the experimental problems by using principles of formal physics. The subjects were classified as experienced if they reported having at least one college physics course that included a lab; otherwise, they were classified as inexperienced. According to this criterion, 30 subjects were classified as experienced (14 from Lockheed or NASA and 16 from Rice) and 30 as inexperienced (6 from Lockheed or NASA and 24 from Rice). This postexperiment classification resulted in 7 experienced subjects in the production-motion condition, 8 in the production-no motion condition, 10 in the selection-motion condition, and 5 in the selection-no motion condition.

Materials. The stimuli consisted of 12 problems that described moving objects and required a judgment about the trajectory that the object would follow in a particular situation. The problems, including diagrams and accompanying verbal descriptions, were based on four problems that have been used in previous studies (Clement, 1983; McCloskey, 1983b): (1) a ball moving through a coiled tube, (2) a ball rolling off of a cliff, (3) a ball falling after being cut from a swinging pendulum, and (4) a rocket moving through space before and after its engines have been fired. In addition to the standard versions of these problems, there were two modified versions of each. One modification involved a change in a perceptual feature of the problem display. For example, the starting end of the tube was extended in a straight direction. The other modification involved changes in the orientation of the object. In the tube problem, the tube was rotated clockwise about 135° so that the end of the tube from which the ball emerged was pointing in a downward direction (although the verbal description of the problem clearly stated that the tube was lying flat on a table) instead of upward and toward the right. The pictorial representations of the 12 problems and their corresponding verbal descriptions are presented in Appendix A. Note that with exception of the pendulum-orientation problem, the cliff, pendulum, and rocket problems are isomorphic with respect to the physics principles that apply.

Three distractors for each problem were generated to serve as the incorrect alternatives in the selection condition. Care was taken to ensure that the distractors were representative of common erroneous responses found in the literature (Clement, 1983; McCloskey, 1983b) and in a pilot study conducted at Rice University. In addition, in order to ensure that the selection test was well constructed (i.e., one alternative should be nearly correct and one completely wrong), the distractors were selected with the constraint that they differ in degree of correctness. Other research has demonstrated that specific trajectory errors are made to different degrees by experienced subjects (Clement, 1983). This criterion was verified by giving the final set of distractors, along with the correct solutions, to four judges who had extensive formal physics training. The judges independently ranked the four stimuli for each problem according to degree of correctness. Final ranks were based on the rank given by a majority or all of the four judges. Thus, three or four of the four judges agreed on the rank of each distractor, although all four judges agreed on the correct answer for each problem. The correct trajectory and the three distractors for each problem (ordered from left/correct to right/incorrect) according to degree of correctness are presented in Appendix B.

In the production conditions, the verbal description of the problem was followed by the associated drawing (Appendix A). In the selection conditions, the verbal description was followed by drawings of the correct trajectory and three distractor trajectories in a two by two arrangement. Note that in all cases, the drawings in the selection condition were identical to those in the production condition, except that the trajectories (one correct and three incorrect) were available in the selection condition. The positions of the four selection stimuli were randomly determined for each problem.

The problem drawings were created with Videoworks II (MacroMind, Inc.) animation software. In the motion conditions, the animation capabilities of Videoworks II were used to add movement to the drawing. Movement in each of the four standard problems consisted of (1) the ball moving through the tube, (2) the ball rolling across the cliff, (3) the ball swinging back and forth with the string as a pendulum, and (4) the rocket moving from point A to point B and then firing its engine. In the selection-motion condition, the animation not only depicted the problem situation, but also the complete trajectory for all four of the alternatives. In the production-motion condition, the animation only depicted the problem situation (e.g., the ball was shown moving through the tube, but it stopped at the end of the tube). The animation was continually replayed for as long as the subject viewed the display. The stimuli were presented and some of the responses were collected on a Macin-

tosh computer using HyperCard. In addition, some responses were collected on paper.

Procedure. Each subject was seated in front of a Macintosh computer. The experimenter instructed the subject about the task requirements and went over two example problems. The examples were similar in style to the test problems, but their solutions did not reveal any principles about the trajectories of moving objects (i.e., one involved weights and a pulley; the other, weights balanced on a board and fulcrum). The subjects were given feedback about their performance on the example problems, but they were not given feedback about the test problems. After the examples had been completed, the experimenter answered questions and the experimental trials began.

The order of the 12 problems was randomized for each subject. For each trial, a verbal description of the problem was presented first. The subjects had as long as they wished to read the description. When they were ready to continue, they used the mouse to select a computer button labeled READY, at which point a pictorial representation of the problem was displayed.

In the production conditions, a single drawing (animated in the motion condition, static in the no-motion condition) of the object was displayed. The subjects viewed this drawing as long as they wished and responded by drawing the path corresponding to the correct trajectory on an answer sheet that displayed a drawing of the object. Next, they wrote a brief verbal description of their drawing. The subjects were told that the description would help judges interpret their drawing if there was any question or ambiguity. They then wrote an explanation for the path that they had drawn. When they were finished responding, they used the mouse to select a CONTINUE button. At this point they were given the chance to go back and view the problem description and the drawing again if they wished; if not, they selected a second CONTINUE button.

In the selection conditions, the four drawings (correct trajectory and three distractors) appeared simultaneously on the screen (animated in the motion condition, static in the no-motion condition). Each of the four drawings was identical in size to the single drawing presented in the production condition. The drawings were numbered 1, 2, 3, and 4 from left to right and top to bottom. The subjects viewed the display as long as they wished and were told to indicate the correct alternative by selecting it with the mouse. These subjects also wrote a verbal description of what they had selected and a reason for their selection on a paper answer sheet. Next, a message was displayed on the screen that indicated the number of the selected alternative. At this point, the subjects were allowed to go back and view the problem description and the drawings and to select a different answer if they wished. If they selected a different answer, they then changed their answer sheets accordingly. If the subjects did not wish to view the problem again, they used the mouse to select a CONTINUE button.

In all conditions, the subjects were asked to rate the confidence of their responses (drawing or selection) before moving on to the next problem. A four-point scale appeared on the screen. The subjects used the mouse to select one of the four points along the scale labeled from left to right: *not confident*, *somewhat confident*, *moderately confident*, and *very confident*. After the confidence rating was entered, the next trial began and a new problem description appeared on the screen. The subjects were not allowed to go back to previous problems once a new trial had been initiated. After having completed all 12 trials, the subjects filled out a questionnaire pertaining to their physics experience.

Results

Trajectory judgments. The trajectory responses were scored for correctness. Because it was assumed that a subject's particular trajectory response could be the result of a number of factors, no assumptions about naive theories or beliefs were made on the basis of trajectory responses.

Whereas there is an objective criterion for a correct response in the selection condition (i.e., the selection of the Newtonian alternative), the criterion for a correct drawing in the production condition is more subjective. Judges could decide which drawings depicted correct trajectories; however, this strategy does not allow for a direct comparison of the two response types. For example, a slightly deviant trajectory judgment might be associated with a drawing judged to be "correct" in the production condition. However, the same judgment might lead to the selection of an incorrect alternative in the selection condition.

To equate the criteria for scoring the production and selection responses, the drawings in the production conditions were classified into one of four categories corresponding to the four selection alternatives. The classification was made by two judges and was based purely on the physical similarity of the drawings to those in each of the selection alternatives. This strategy ensures that responses are subject to the same four possibilities in both conditions. Thus in both conditions responses fell into one of four categories; in the selection condition, however, the category was selected by the subject, and in the production condition, it was selected by judges on the physical basis of the subject's drawing. This scoring strategy does not require that the four responses be equally probable, merely that they be the same for the two conditions. In addition, the strategy does not require the assumption that subjects' behavior in terms of responding in one of four ways be equivalent across the conditions. Precisely these differences are of interest.

The two judges agreed on the classification of 338 (94%) of the 360 (30 subjects \times 12 problems) production drawings. Disagreements were discussed by the judges until an agreement on the closest match was reached. Most (16) of these disagreements involved a conflict between two categories, whereas only 6 of the drawings could not easily be classified into any of the four categories. The small proportion of unclassified drawings supports the representativeness of the response categories determined by pilot testing.

Trajectory scores of 1, 2, 3, or 4 were assigned to each response on the basis of the rank of the response category previously determined by the four expert judges. A score of 4 indicates that the trajectory is correct, and scores of 1, 2, and 3 indicate incorrect responses, but in increasing order of plausibility. A total of 453 (63%) of the judgments were correct (i.e., scores of 4). A total of

270 (75%) of the judgments by subjects with formal physics training were correct and 183 (51%) of the judgments by inexperienced subjects were correct. These proportions are comparable to the proportion of correct trajectories in previous experiments (e.g., 47%–68% in McCloskey et al., 1980).

To verify the ordinal nature of the trajectory scores, a median split was done on the number of correct responses (4s) for each subject. Of the 30 *four-scoring* subjects, 20 were experienced and 10 were inexperienced. If the scores are truly ordinal, there should be a greater proportion of 3s scored by four-scoring than 2s and a greater proportion of 2s scored by four-scoring than 1s. In fact, 38% of the 3s, 14% of the 2s, and 6% of the 1s were scored by four-scoring. This pattern, coupled with the interjudge agreement on ranks, justifies the use of parametric statistical analyses to test for differences among mean trajectory scores.²

Trajectory scores were subjected to a $2 \times 2 \times 2 \times 4 \times 3$ (response type \times display type \times experience \times problem \times version) analysis of variance (ANOVA). Neither display type (motion, no motion) nor version (standard, perceptual set, orientation) factors were significant, nor did either enter into significant interactions with any other factors. Mean trajectory scores collapsed across display type and version are presented in Table 1. The number of correct problems in each cell is also noted.

The main effects of response type [$F(1,52) = 5.66$, $MS_e = 2.35$, $p = .02$] and experience [$F(1,52) = 22.4$, $MS_e = 2.35$, $p < .0001$] were significant, as was the interaction between these two factors [$F(1,52) = 6.45$, $MS_e = 2.35$, $p = .01$]. In general, scores were higher in the selection condition than in the production condition and were higher for experienced subjects than for inexperienced subjects. However, the advantage of the selection over the production response held for inexperienced subjects [$F(1,26) = 8.51$, $MS_e = 3.34$, $p = .007$], but not for experienced subjects [$F(1,26) < 1$]. Furthermore, experienced subjects were more accurate than inexperienced subjects in the production condition [$F(1,26) = 26.25$, $MS_e = 2.51$, $p = .0001$], but this difference was not significant in the selection condition [$F(1,26) = 2.44$, $MS_e = 2.19$, $p = .13$]. Note that performance of the experienced subjects, although very good, was not perfect (75% correct, mean score = 3.6), and therefore this interaction cannot be completely attributed to a ceiling effect.

Table 1
Mean Trajectory Scores (1 = Incorrect, 4 = Correct) for Experiment 1,
With the Number of Correct Problems (Score = 4) for Each Cell ($n = 45$)

Problem Context	Experienced				Inexperienced				Total	
	Selection		Production		Selection		Production			
	<i>M</i>	Correct	<i>M</i>	Correct	<i>M</i>	Correct	<i>M</i>	Correct	<i>M</i>	Correct
Tube	3.93	43	3.93	44	3.82	39	3.76	39	3.86	165
Cliff	3.80	38	3.93	43	3.47	31	2.87	18	3.52	130
Pendulum	3.58	33	3.47	28	3.56	30	2.64	11	3.31	102
Rocket	3.16	20	3.27	21	2.56	10	1.91	5	2.72	56
Overall	3.62	134	3.65	136	3.35	110	2.79	73	3.35	453

In addition, the main effect of problem [$F(3,156) = 38.41, MS_e = 1.01, p < .0001$] and the interaction of problem and experience [$F(3,156) = 5.67, MS_e = 1.01, p = .001$] were significant. In general, the subjects' trajectories were most accurate for the tube problem, followed by the cliff, pendulum, and rocket problems. Also, the effects of experience were greatest for the cliff and rocket problems. All other interactions with problem failed to reach significance.

Although the main effect of version was not significant, some subtle effects of version within individual problems may have been masked, given that this factor was manipulated differently within each problem. For instance, the removal of the frame in the pendulum perceptual set problem may not have been as salient as the rounding of the cliff in the cliff perceptual set. Thus, the effect of version was investigated separately for each problem. Results indicated that version was significant for the tube problem only with subjects scoring slightly, but significantly [$F(2,104) = 3.01, MS_e = .188, p = .05$] better on the perceptual set and orientation versions (mean score = 3.9 for each) than on the standard version (mean score = 3.7).

Confidence ratings. An ANOVA of confidence ratings resulted in small, but significant effects of experience [$F(1,52) = 6.23, MS_e = 3.03, p = .016$] and problem [$F(3,156) = 8.26, MS_e = .463, p < .0001$]. No other effects were significant. As expected, experienced subjects were more confident than inexperienced subjects (mean confidence ratings = 3.6 and 3.3, respectively, where 4 is *very confident*). Also, subjects were more confident on the cliff (3.6), tube (3.5), and pendulum (3.5) problems than on the rocket (3.2) problem. Thus, confidence rating patterns are consistent with the trajectory scores, except that inexperienced subjects were equally confident (3.3) for selection and production responses, providing additional evidence that the selection alternatives were good distractors.

Number of explanation errors. Each explanation was independently classified as correct or incorrect by two judges. The criteria for correct explanations are presented in Table 2. Incorrect explanations were also classified according to type of error, as described in the next section. The judges agreed on the classification of 635 (88%) of the 720 explanations. Disagreements were discussed by both judges until agreement was reached.

There were a total of 514 (71%) incorrect explanations. Data indicating whether explanations were correct or incorrect were subjected to a $2 \times 2 \times 2 \times 4 \times 3$ (response type \times display type \times experience \times problem \times version) ANOVA. The results indicated that the inexperienced subjects made more of the 514 explanation errors (299) than the experienced subjects did (215), and that this difference was significant [$F(1,52) = 12.89, MS_e = .79, p = .0007$]. In addition, subjects made fewer explanation errors on the tube problems (72) than on the pendulum (144), cliff (147), and rocket (151) problems [$F(3,156) = 36.52, MS_e = .20, p < .0001$]. Note that significant problem differences exist even for the isomorphic cliff and pendulum problems (ignoring orientation) [$F(1,52) =$

Table 2
Criteria Used to Score Explanations in Experiments 1 and 2

Problem	Correct Explanation
Tube	Because the tube no longer exerts force on the ball, it travels in a straight line after emerging from the tube.
Cliff	Two components, a constant horizontal velocity and vertical acceleration of gravity, combine to produce a parabolic path from the cliff to the ground.
Pendulum* (standard & perceptual set)	Two components, a constant horizontal velocity and vertical acceleration of gravity, combine to produce a parabolic path to the ground.
Pendulum* (orientation)	The ball falls straight down due to gravity.
Rocket	Two components, a constant horizontal velocity and vertical acceleration from the engine's thrust, combine to produce a parabolic path from B to D.

Note—Underscored segments constitute the minimum requirements for correctness, although exact wording was not necessary. *Experiment 1 only.

4.17, $MS_e = .06, p = .046$]. In addition, problem interacted significantly with display type [$F(3,156) = 3.49, MS_e = .20, p = .02$], response type [$F(3,156) = 8.38, MS_e = .20, p < .0001$], and version [$F(6,312) = 4.30, MS_e = .20, p = .0003$]. In general, these results indicate that specific display and response features of the problems affected the accuracy of explanations. Specifically, there were more explanation errors in the no-motion condition for the cliff (53%) and rocket (52%) problems, but more explanation errors in the motion condition for the tube problems (60%). Also, with the exception of the pendulum problems (46% errors in selection), more explanation errors were generally made in the selection condition (55%) than in the production condition (45%). Finally, the perceptual set problems were associated with fewer errors (32%) than other types of problems were, except for the pendulum problems, for which the orientation variation resulted in the fewest errors (29%). All other differences failed to reach significance.

For 263 (37%) of the problems, both trajectories and explanations were incorrect, and for 202 (28%) of the problems, both were correct. Furthermore, when the trajectory was incorrect, the associated explanations were almost always also incorrect. In only four cases (less than 1% of the problems) did subjects give a correct explanation for an incorrect trajectory. In these cases, although the trajectory was incorrect (e.g., 1 subject predicted that the ball in the pendulum orientation problem would fall downward and to the left in a parabolic arc), the explanation was correct, given the conditions that the subject described (e.g., the ball was not at the top of its swing, but in motion to the left). More interesting are the 251 (35%) cases in which subjects gave incorrect explanations for correct trajectories. In other words subjects gave incorrect explanations for over half (55%) of the correct trajectories. To examine the pattern of explanation errors of this type, the percentage of correct trajectories associated with incorrect explanations across all 12 problems

was computed for each subject and subjected to a $2 \times 2 \times 2$ (experience \times response type \times display type) ANOVA.

The effect of experience on explanation performance was significant [$F(1,52) = 5.37$, $MS_e = .09$, $p = .02$]. On the average, the experienced subjects gave incorrect explanations for 50% of their correct trajectories; for the inexperienced subjects, the average was 68%. Percent incorrect explanations for correct trajectories was also greater for selection responses (mean = 68%) than for production responses (mean = 49%) [$F(1,52) = 7.02$, $MS_e = .09$, $p = .01$]. The main effect of display, as well as all interactions, failed to reach significance.

The number of subjects ($n = 60$) giving incorrect explanations to correct trajectories for each of the 12 problems was also computed. Explanation performance on the tube problems (33% incorrect) and pendulum orientation (27% incorrect) problems was quite good in comparison with performance on the other pendulum problems (43% incorrect) and the cliff problems (54% incorrect), probably owing to differences in the amount of detail required (see Table 2). On the other hand, performance was quite good on the rocket problems (15% incorrect), which are the most difficult. The latter result suggests that if subjects responded correctly to these difficult problems, they also knew enough to give the correct explanation.

Types of explanation errors. The types and frequencies of explanation errors for both correct and incorrect trajectories are presented in Table 3, along with an example of each type of error. Thirteen miscellaneous errors that are not included in the table ranged from statements

such as "it looked good" to vague references to physics principles. The most frequently occurring errors were (1) errors of omission—omission of one or more critical factors (i.e., gravity, velocity, thrust, acceleration); (2) descriptive errors—the subject simply repeated the problem situation; (3) impetus errors—the object possessed an internal force such as curvilinear momentum or one force overcame another; and (4) surface errors—inappropriate focus on surface features of the problem, such as the curve of the cliff or the shape of the tube. Note that impetus errors were defined very broadly and included any explanations that mentioned an object's internal impetus (including curvilinear impetus), the dissipation of impetus, and the overcoming of one force by another. In addition, explanations containing terms such as *energy*, *momentum*, and *force* were included in this category when they were mentioned in an impetus context. Although this grouping admittedly blurs many interesting distinctions among varieties of impetus explanations, it at the same time maximizes the chances of identifying consistent impetus beliefs.

Interestingly, 48% of the 720 responses contained explanations that omitted critical factors, were vague, or restated the problem situation. As a consequence, these explanations did not reveal any particular naive theory or belief. It is likely that more extensive probing along the lines of the 2-h interviews carried out by McCloskey (1983b) would suggest one or more correct or incorrect beliefs, but this procedure was not practical for such a large number of subjects. Indeed many of the explanations associated with omissions may have been conceptually

Table 3
Types of Explanation Errors and Frequencies of Occurrence in Experiment 1

Error Type (Frequency)	Description	Example
Omission (275)	omits horizontal component, gravity, thrust, or acceleration	<i>Problem:</i> rocket standard. Subject describes the rocket as traveling forward/downward. <i>Explanation:</i> "The firing of the engine will cause the rocket to move forward, and since there are no other forces acting on the rocket, it will continue in this direction to points C & D."
Descriptive (73)	vague or repeats problem statement	<i>Problem:</i> coiled tube-standard. Subject states that the ball shoots out straight. <i>Explanation:</i> "Since there is not any spin on the ball, it shoots out straight."
Impetus (83)	ball has momentum, impetus, energy, curvilinear momentum, or one force overcomes another	<i>Problem:</i> pendulum orientation. Subject says that the ball will drop at a slight curve. <i>Explanation:</i> "The ball had impetus and momentum on it. Because of these forces, the ball will curve ever so slightly."
Surface (43)	focus on physical features in problem description or picture	<i>Problem:</i> cliff perceptual. Subject explains that the ball follows the cliff in a curved trajectory. <i>Explanation:</i> "Because of the curve of the cliff, the ball will stay along the cliff when rolling."
Adds velocity (19)	adds velocity component in the pendulum orientation problem	<i>Problem:</i> pendulum orientation. Subject draws a curved path from ball to the ground right of the ball. <i>Explanation:</i> "The ball continues in the direction it was moving when the string was cut and is gradually turned downward by gravity. Cutting the string such that the ball is released at 45 degree angle will result in the farthest distance the ball will travel."
Replaces gravity with weight (5)	weight replaces gravity	<i>Problem:</i> cliff orientation. Subject correctly draws a parabolic arc from the cliff to the ground. <i>Explanation:</i> "The ball doesn't continue on its straight path off the cliff all the way to the ground because the momentum of the ball combines with the weight of the ball to curve its path in a downward motion to the ground."
Incorrect combination of components (3)	all components are indicated but are combined incorrectly	<i>Problem:</i> pendulum perceptual set. Subject draws a diagonal path from the ball to slightly left of the ball on the ground. <i>Explanation:</i> "The ball will travel the path which is a combination of the swing speed (max at this point) and gravity acceleration constant."

correct, although in some cases these errors seemed related to associated trajectory errors (e.g., omission of the horizontal velocity of the rocket, accompanied by the belief that the rocket goes straight up). In order to avoid making arbitrary decisions about whether or not these vague explanations were correct, the previously reported scoring of explanations as correct was based on strict conceptual adherence to the criteria in Table 2. Explanations did not have to correspond word for word, but they had to include the major components of the explanation. Therefore, because the explanation error data presented in the previous section encompass errors that range from vague, but potentially correct explanations to those containing impetus-like statements, the ANOVA done on overall errors was repeated with error type included as an additional within-subjects factor. Error type was coded in terms of the presence or absence of errors of three different types (*impetus*, *surface*, or *vague*). Vague errors included both descriptive errors and errors of omission. Because the analysis of overall error rate was presented in the previous section, the results presented in this section will be focused exclusively on the new factor, error type.

The number of impetus and surface explanations ranged from 0 to 12 across all 12 problems ($n = 60$) in contrast to a range from 8 to 45 for vague explanations. Only the tube problems were associated with vague answers at the same low frequency as that for impetus and surface explanations. Results of the ANOVA indicate that the main effect of error type was significant [$F(2,104) = 85.49$, $MS_e = .43$, $p < .0001$], as well as the interactions between error type and problem [$F(6,312) = 15.66$, $MS_e = .25$, $p < .0001$] and error type and version [$F(4,208) = 5.72$, $MS_e = .06$, $p = .0002$] and the three-way interaction between error type, problem, and version [$F(12,624) = 9.47$, $MS_e = .08$, $p < .0001$]. In addition, there were more impetus errors than surface errors [$F(1,52) = 6.51$, $MS_e = .17$, $p = .014$], although this was not true for the tube problems and cliff perceptual set problem. This interaction between problem and error type (impetus vs. surface only) was also significant [$F(3,156) = 2.79$, $MS_e = .17$, $p = .04$]. In addition, the four-way interaction between error type, problem, version, and response type was also significant [$F(12,624) = 1.81$, $MS_e = .08$, $p = .04$]. This interaction can be attributed mostly to the general finding that more impetus and surface explanation errors are made in the selection conditions than in the production conditions, but the opposite pattern holds for vague errors in the pendulum standard problem. All other interactions with error type failed to reach significance. In general, the analysis of error type indicates that the types of explanation errors that were made in this study were dependent on particular features of the problems.

If one assumes that a naive theory is an abstract entity that can be applied to a variety of similar situations, one should expect subjects to make the same type of explanation error across problems. A subject was classified as having a naive theory if he/she made the same error on all three versions within a particular problem context and if this error occurred across two or more of the four prob-

lems. According to this criterion, and ignoring errors of omission and descriptive errors, only 3 out of 60 subjects had naive theories. These 3 all made consistent impetus errors, but they made other types of errors as well. Even with a lenient criterion (4 or more errors of the same type across the 12 problems) only 12 subjects were identified as having naive theories (10 made consistent impetus errors, and 2 made consistent surface errors). In 9 of these 12 cases, both impetus and surface errors were made by the same subject. These results raise questions about the role, if any, of a naive impetus theory of motion.

The degree of these inconsistencies can be assessed by comparing the relative consistency of explanations associated with naive impetus theory and those associated with Newtonian theory. That is, how consistently (or inconsistently) do subjects apply naive impetus theory, as opposed to Newtonian theory? Given the strict criterion described above (i.e., correct/Newtonian explanations for all three problem versions across at least two of the four problem contexts), 11 of the 60 subjects produced consistent Newtonian explanations. Note that these differences (5% impetus vs. 18% Newtonian subjects) are conservative, given that explanations were scored strictly, erring on the side of finding impetus explanations and against finding Newtonian explanations.

Not only were impetus explanations inconsistently applied, but they were unrelated to subjects' performance on individual problems. Of the 83 impetus explanations, only 49% were associated with an incorrect trajectory. The remainder were associated with correct trajectories. This dissociation between impetus explanations and trajectory performance raises concerns about the pragmatic value of identifying and targeting "impetus theorists" for remediation.

Discussion

The results of Experiment 1 replicate previous findings that indicate that people often make errors when predicting the trajectories of moving objects and that in general, subjects with formal physics training make fewer errors than those with no training. Additionally, these results extend previous results indicating that trajectory judgments depend on problem features to include the context of the problem and type of response required. For instance, subjects responded with the correct trajectory for 72% of the cliff problems, but only 31% of the isomorphic rocket problems. Also, inexperienced subjects were more accurate at selecting the response from a set of alternatives than they were at drawing it.³ Furthermore, an analysis of explanations resulted in very little evidence that a consistent set of beliefs held by an individual leads to trajectory judgments. A number of the explanation errors were impetus-related, replicating previous findings; however, other types of errors were represented across and within individuals. Explanations also depended on various problem characteristics.

The failure to find a significant effect of motion on trajectory judgment like that found by Kaiser, Proffitt,

and Anderson (1985) is most likely due to the way that motion was depicted. In Experiment 1, the animated motion displays failed to preserve the kinematics of veridical motion, and consequently, they appeared artificial. Another possibility is that responses in the production motion condition were made on static diagrams, potentially negating any effects of motion.⁴ It should be noted that McCloskey and Kohl (1983) also failed to find an effect of motion under conditions similar to these. On the other hand, Kaiser, Proffitt, and Anderson (1985) obtained the effect with a selection response and videotapes of moving objects. Although the presence or absence of motion did not affect trajectory accuracy, it did affect explanation accuracy for some problems. Explanations were more accurate in the motion conditions for the cliff and rocket problems. Perhaps even though the motion display did not preserve the dynamics necessary to cue correct trajectory judgments, in some situations it provided cues that could remind subjects of the important problem components.

The benefit of a selection response over a production response not only adds another factor to the list of those relevant to trajectory judgments, but also has implications concerning the knowledge of those without formal physics training. This knowledge does not appear to be limited to familiar problems (Kaiser et al., 1986), but can be applied to more abstract problems if the trajectory can be selected from a set of alternatives. The existing knowledge enables individuals to "know" the correct trajectory when they see it, either because the selection situation provides sufficient retrieval cues for typically inaccessible knowledge or because it provides additional information that can be used in conjunction with partial knowledge. Thus, the error rates reported in previous studies may exaggerate the extent to which subjects are truly naive about object motion. Instead, errors in a production task may reflect either a complete absence of knowledge or the existence of knowledge that is insufficient or inaccessible in that particular situation.

This hypothesis is supported by the finding that for correct trajectories, experienced subjects gave more accurate explanations than inexperienced subjects did, and that subjects in the production condition gave more accurate explanations than those in the selection condition did. These results can be explained easily if it is assumed that trajectory explanations require at least as much knowledge as trajectory predictions, if not more. Furthermore, because the production condition is impoverished in comparison with the selection condition, it requires more knowledge than selection does. Therefore, given that the trajectory is correct, those who have produced it are more likely to have sufficient knowledge to explain it than are those who have selected it. Similarly, most of the correct trajectory predictions made by inexperienced subjects were selected, whereas those made by experienced subjects were produced and selected equally often. In other words, a trajectory can be selected by using knowledge that is insufficient to produce or explain it or that is inaccessible.

Results from Experiment 1 and previous studies (e.g., Hojnacki, 1988; Kaiser, Proffitt, & Anderson, 1985; Ran-

ney, 1987) have demonstrated inconsistency in subjects' trajectory judgments. For instance, Kaiser, Proffitt, and Anderson (1985) found that when 105 subjects were presented with pairs of six alternative trajectories for a C-shaped tube problem, 17 (16%) of these subjects demonstrated no consistent preference for an alternative as measured by four or more circular triads. On the other hand, these inconsistencies do not rule out the possibility that naive impetus theory underlies some errors of motion judgments. That is, it could be argued that naive theories are applied consistently across problems that are isomorphic with respect to the naive theory. Thus, subjects may respond differently to the cliff and pendulum problems because the former involves active motion and the latter passive motion—a meaningful distinction, according to a naive impetus theory. However, this type of explanation cannot explain all of the inconsistencies that have been demonstrated.

From a pragmatic point of view, it seems less important to determine the existence of naive impetus theory than the ultimate utility of the construct. Are they applied so inconsistently as to be of little practical value in education? Are there other alternatives that are equally or more useful from this pragmatic perspective? The results in Experiment 1 suggest that problem context plays a role at least equal to that of naive impetus theory. Not only did subjects' responses vary as a function of context, but in many of their explanations, their focus on specific contextual features was apparent.

Before further exploring an alternative account of errors in motion judgments, we must address two potential limitations of Experiment 1. First, subjects were asked to provide explanations in a written form. It could be argued that these written reports were impoverished in comparison with the oral reports elicited in 2-h interviews by McCloskey (1983b). As a result, many errors of omission and description were obtained in this study that may have been eliminated by more extensive probing. Some of these errors may have been impetus, surface, or other errors. Furthermore, more extensive probing would be required to separate truly naive misconceptions from linguistic slips. For example, some individuals who say that an object "has momentum" may incorrectly mean that the object has an internal force, whereas others may correctly mean that the object has velocity.

Alternatively, written reports allow for a more extensive analysis of explanations given by the same individual across different problems. Approximately half of the incorrect explanations revealed some initial misconception. At the least, one can conclude from these results that impetus misconceptions are not immediately obvious in all cases. Also, extensive probing in an interview is likely to have several undesirable characteristics, including the subjective nature of the probe questions, the uncertainty of knowing when to stop, and the possibility that some probe questions may lead the subject to respond in a particular way (Donley & Ashcraft, 1992). In addition, explanations could be impoverished because that is the nature of people's explanations of phenomena that they have not

intensely pondered (i.e., some things just happen and do not require explanation). DiSessa (1988) points out that people tend to explain only that which is problematic or puzzling, rather than extract fundamental principles to explain every situation. Therefore, as a compromise between these two positions, in Experiment 2 subjects were explicitly instructed to write very detailed explanations and were given example explanations that were quite lengthy.

A second potential limitation of Experiment 1 is that there is an alternative explanation for the effect of response type (i.e., selection vs. production). It could be argued that individuals in the selection condition used test-taking strategies (e.g., select the alternative that is most different from the others; select one alternative from the two that are the most similar) to make sophisticated guesses on problems for which they possessed no knowledge. There are several reasons to suspect that test-taking strategies cannot account for the advantage of selection over production. First, an attempt was made to construct a good selection test by choosing distractors that represented common mistakes and that varied in degree of plausibility. The fact that inexperienced subjects were no more confident on the selection test than on the production test provides some evidence of success in this attempt. Also, the experienced subjects did no better on the selection test than on the production test, even though they had room to improve.

Second, a pilot study was conducted in which 34 introductory psychology students at Rice University were given static drawings of 9 of the 12 selection pictures used in Experiment 1 (for unrelated reasons, the pendulum problem was not included). The 9 problems (four alternatives in each) were presented without a verbal description or question. The subjects were told that the experiment had to do with test-taking strategies and that they were simply to select the alternative from the four that they thought was correct. After the subjects had selected an answer for each of the 9 problems, they were given a questionnaire concerning any strategies that they had used and they were also asked to guess at what question was being tested. Not too surprising was the finding that 25 of the 34 subjects guessed the correct question (i.e., which is the correct trajectory?). These 25 correctly answered 63% of the problems (mean of 5.7 out of 9 correct per subject). The 9 who did not know the question answered 38% of the problems correctly (mean of 3.4 out of 9 correct per subject). Thus, although the 9 naive subjects did better than chance (25%), they did not do as well as the subjects who knew the question, providing additional support that the selection condition was superior to production for inexperienced subjects for reasons other than test-taking strategies.

In addition, the strategies reported by the subjects would not have proved helpful most of the time, given the way in which the alternatives were designed. For instance, the reported strategies of picking one of the two most similar pictures or of not choosing the most different alternative would have resulted in many incorrect responses. Despite these arguments against the test-taking strategy account,

the strongest evidence would be a replication in a situation in which this possibility was eliminated. This was done in Experiment 2 by limiting the response to one of two alternatives.

EXPERIMENT 2

The purpose of Experiment 2 was to attempt to replicate the selection advantage found for inexperienced subjects in Experiment 1 by using a two-alternative selection task. Decreasing the number of alternatives eliminates any advantages in the selection condition that are due to the use of test-taking strategies that rely on the physical similarity of alternatives. Subjects will be unable to make judgments that are based on the prototypical or odd instance in a group. In addition, the subjects in Experiment 2 were given explicit instructions and examples to encourage careful and detailed written explanations.

Method

Design. Response type (production or selection), physics experience (experienced or inexperienced), problem (tube, cliff, and rocket), and version (standard, perceptual set, and orientation) were manipulated in a $2 \times 2 \times 3 \times 3$ factorial design. Problem and version were within-subjects factors. All other factors varied between subjects.

Subjects. A total of 148 individuals voluntarily participated in this experiment. They included 56 students enrolled in undergraduate psychology classes at Rice University and 60 students enrolled in undergraduate psychology classes at the University of Houston. Additional experienced subjects were obtained from several other sources, including faculty from the University of Houston (2), undergraduate science majors at Rice University (21), physics graduate students at MIT (5), and engineers at Lockheed Engineering and Sciences Corporation in Houston, TX (4). The subjects were randomly assigned to either the production or the selection condition.

The subjects were categorized as experienced or inexperienced on the basis of information provided on a postexperiment questionnaire. The questionnaire was identical to the one used in Experiment 1 with the exception of 10 additional test questions taken out of a physics text book (Tipler, 1976), which are listed in Appendix C (the first 5 questions and Numbers 1, 3, 5, 8, and 11 of the true/false questions). These were included in order to have a performance-based measure of experience. One point was assigned to each correct answer, including correct answers to the three parts of the first question, for a total possible score of 12. The questionnaire also included seven true/false questions (Numbers 2, 4, 6, 7, 9, 10, 12) that were relevant to impetus theory, but not included in the test score. The wording of these impetus questions was taken from previous accounts of impetus theory (McCloskey, 1983b).

Subjects were classified as experienced if they had taken at least one college physics course and scored 67% or higher on the physics test. On the basis of these criteria, 50 subjects were classified as experienced and 98 as inexperienced. The experienced subjects had taken a mean of 4 college physics courses and had a mean test score of 10.2; the inexperienced subjects had a mean test score of 5.6. Of the experienced subjects, 21 were in the selection condition and 29 in the production condition. There were 50 inexperienced subjects in the selection condition and 48 in the production condition.

Materials. The stimulus materials consisted of the same two example problems and 9 of the 12 test problems used in the first experiment. The pendulum problem was eliminated because the point

of breakage seemed to be confusing to some subjects in the first experiment. In addition, the three versions of the rocket problem were altered slightly (see Appendix A). Apart from these changes, the problems and their verbal descriptions were identical to those in Experiment 1.

A single distractor was chosen from the three used for each problem in Experiment 1. The choice was based on the most frequent error found in Experiment 1 and in the literature. The distractors for the tube problem represented the typical "curvilinear impetus belief," although not in its extreme form (see Appendix B, almost correct alternatives). The cliff and rocket distractors represented the typical "straight-down belief" (see Appendix B, slightly wrong alternatives for the rocket and cliff perceptual set; wrong and almost correct options for the cliff standard and cliff orientation, respectively).

Two example problems were presented in a test booklet, followed by the nine test problems. The order of the test problems was randomized across subjects. In both conditions, the verbal description of the problem was followed by the associated drawing (Appendix A). In the production condition, the drawing was duplicated to serve as the response diagram. In the selection condition, the verbal description and drawing were followed by drawings of the correct trajectory and the distractor trajectory in a side-by-side arrangement. The position of the correct and incorrect drawings in each selection problem was randomly determined for each subject. The selection alternatives were identical to the problem drawing (Appendix A), except that the trajectories were included in the drawings.

Procedure. The subjects were seated at a desk, and the experimenter instructed them about the task requirements and went over the two example problems used in Experiment 1. The instructions stressed the importance of writing detailed explanations: "It is important that you give us as much relevant detail as you can for the explanation. In other words, assume that you are explaining your answer to a person who knows nothing about the problem." In addition, detailed explanations were provided for both of the example problems. Apart from these changes, the instructions were the same as those in Experiment 1.

In the production conditions, the subjects drew their predicted trajectory on the response diagram, and in the selection condition, the subjects placed a mark underneath the alternative that they thought was correct. Next, the subjects in both conditions wrote brief descriptions of their drawings or selections, rated their confidence in their answers, and wrote explanations for their answers as in Experiment 1. After all nine problems had been completed, the subjects filled out the questionnaire/test pertaining to their physics experience.

Results

Trajectory judgments. As in Experiment 1, in order to equate the two response conditions, the drawings in the production conditions were classified by two judges into one of two categories corresponding to the two selection alternatives for each problem. Again, this strategy ensures that responses are subject to the same possibilities in both conditions. As before, the classification was based on physical similarity. This distinction was not perfectly correlated with a correct/incorrect distinction. For instance, diagonal trajectories (Appendix B, rocket, standard cliff, and perceptual set cliff—almost correct alternatives), although incorrect, were judged to be more physically similar to the correct parabolic trajectory (Appendix B, rocket and cliff correct alternatives) than to the "straight-down" distractor. Thus, responses that would have been classified in one of the two omitted response categories of Experiment 1 were mapped onto the remaining two categories according to physical similarity. Two judges agreed on the classification of 97% of the 693 production drawings. Disagreements were resolved by a third judge. The disagreements all involved drawings that could not easily be classified into either of the categories. The small proportion of unclassified drawings supports the representativeness of the response categories.

Trajectory scores of 1 (correct) or 0 (incorrect) were assigned to each response on the basis of the selected or judged response category. A total of 1,017 (76%) of the judgments were correct. As anticipated, the percent correct was slightly greater in this experiment than in Experiment 1, probably owing to the more lenient scores that result from a binary classification of responses. The percentage of correct responses in each condition is presented in Table 4.

Trajectory scores were subjected to a $2 \times 2 \times 3 \times 3$ (response type \times experience \times problem \times version) ANOVA. As in Experiment 1, the main effects of response type [$F(1,144) = 7.48, MS_e = .43, p = .007$] and experience [$F(1,144) = 55.0, MS_e = .43, p < .0001$] were significant. In general, more correct trajectory responses were made in the selection condition than in the production con-

Table 4
Percent Subjects in Each Condition Who Gave Correct Trajectory Responses for Each Problem in Experiment 2

Problem	Experienced		Inexperienced		Overall <i>n</i> = 148
	Selection <i>n</i> = 21	Production <i>n</i> = 29	Selection <i>n</i> = 50	Production <i>n</i> = 48	
Tube					
Standard	100	97	80	65	81
Perceptual set	100	97	84	73	85
Orientation	100	97	76	88	87
Cliff					
Standard	100	97	86	69	85
Perceptual set	95	97	68	65	76
Orientation	100	100	86	85	91
Rocket					
Standard	95	86	72	25	63
Perceptual set	95	83	68	27	62
Orientation	91	86	56	29	58
Overall	97	93	75	58	76

dition (82% vs. 72% correct) and by experienced subjects than by inexperienced subjects (95% vs. 67% correct). Although the response type \times experience interaction was only marginally significant [$F(1,144) = 2.66$, $MS_e = .43$, $p = .11$], as in Experiment 1, the experienced subjects tended to benefit less from selection (93% vs. 97% correct) than the inexperienced subjects did (58% vs. 75%). However, the results of the experienced subjects are difficult to interpret because their performance approached ceiling levels in this experiment.

Also, the main effect of problem was significant [$F(2,288) = 21.34$, $MS_e = .55$, $p < .0001$], in addition to the interactions with response type [$F(2,288) = 5.10$, $MS_e = .55$, $p = .007$] and experience [$F(2,288) = 6.57$, $MS_e = .55$, $p = .002$]. In general, subjects' trajectories were more accurate for the tube and cliff problems (85% and 84% correct) than for the rocket problem (61% correct). This result is interesting, given that the cliff and rocket problems are isomorphic with respect to the physics principles that underlie them, as well as in the naive sense that both involve active motion. Also, the effects of experience and response type were greater for the more difficult rocket problems. In fact, the marginally significant three-way interaction of problem, response type, and experience [$F(2,288) = 2.31$, $MS_e = .55$, $p = .10$] can be attributed to a significant experience \times response type interaction for the more difficult rocket problems [$F(1,144) = 4.94$, $MS_e = .43$, $p = .03$], although the ceiling effects mentioned above may have contributed to this interaction. The pattern of results found in Experiment 1 (i.e., that selection is easier than production, especially for inexperienced subjects) was also found in Experiment 2, in which selection was limited to two choices instead of four. This general pattern was most pronounced for the more difficult rocket problems.

Responses were also affected by the interaction of problem with version [$F(4,576) = 4.03$, $MS_e = .06$, $p = .003$]. The subjects were least accurate on different versions within each problem (i.e., the tube standard problem, the cliff perceptual set problem, and the rocket orientation problem). Version also interacted with response type [$F(2,288) = 5.92$, $MS_e = .06$, $p = .003$], and the three-way interaction of version, response type, and experience was also significant [$F(2,288) = 3.47$, $MS_e = .06$, $p = .03$]. In general, the difference between selection and production was greater for the standard and perceptual set problems than for the orientation problems, and this was especially true for the inexperienced subjects. Much of this effect can be explained by a slight, but nonsignificant, reversal of the response type effect for the tube orientation problem (i.e., for this problem, production resulted in more correct responses [88%] than selection [76%]).

Because the tube orientation problem depicted the tube opening at the bottom of the drawing, this reversal might be due to subjects in the production condition forgetting that the tube is on a table in a horizontal position and consequently drawing a straight line trajectory for the wrong reasons (i.e., gravity). In the selection condition, the pres-

ence of the curved trajectory may have reminded subjects that gravity was not a factor, so that they ruled out the straight trajectory and concluded that the curved alternative must be correct. Alternatively, in Experiment 1, when all four alternatives were present, three of which were gravity irrelevant, it was clear that there was more to the question than remembering that gravity is not a factor and, consequently, further analysis occurred. Therefore, it seems possible that sophisticated test-taking strategies may have allowed subjects to make better judgments for some of the problems in Experiment 1, although reducing the number of alternatives served to eliminate such strategies. The results for the difficult problems make it clear that the benefits of the selection response extend beyond test-taking strategies and persist in cases in which information from multiple distractors has been eliminated.

Confidence ratings. As in Experiment 1, an ANOVA of confidence ratings resulted in small, but significant effects of experience [$F(1,144) = 43.96$, $MS_e = 2.71$, $p < .0001$] and problem [$F(2,288) = 9.94$, $MS_e = .513$, $p < .0001$]. As in Experiment 1 and in consistency with trajectory effects, experienced subjects were more confident than inexperienced subjects (mean confidence ratings = 3.8 and 3.2, respectively, where 4 is *very confident*) and subjects were more confident about the easier tube (3.4) and cliff (3.5) problems than they were about the rocket problems (3.3). In addition, there was a significant main effect of version [$F(2,288) = 5.53$, $MS_e = .197$, $p = .004$], with subjects being more confident on the standard problems (3.5) than on the orientation (3.4) or perceptual set (3.4) versions.

Number of explanation errors. Each explanation was independently classified as correct or incorrect by two judges, on the basis of the criteria used in Experiment 1 (see Table 2). Incorrect explanations were also classified according to type of error. Data from 15 subjects were randomly selected and used to train the judges on the scoring criteria. The two judges agreed on the classification of 97% of the remaining 133 subjects (1,197 problems). Disagreements were settled by a third judge.

There were 881 (66%) incorrect explanations. Data indicating whether explanations were correct or incorrect were subjected to a $2 \times 2 \times 3 \times 3$ (response type \times experience \times problem \times version) ANOVA. Results indicated that inexperienced subjects made significantly more explanation errors (82%) than experienced subjects did (35%) [$F(1,144) = 104.96$, $MS_e = .62$, $p < .0001$]. In addition, subjects made fewer explanation errors on the tube problems (46%) than on the cliff (73%) and rocket (79%) problems [$F(2,288) = 44.45$, $MS_e = .26$, $p < .0001$]. More errors were made on the rocket problems than on the isomorphic cliff problems [$F(1,144) = 6.97$, $MS_e = .13$, $p = .009$]. In addition, problem interacted significantly with response type [$F(2,288) = 3.93$, $MS_e = .26$, $p = .02$]. As in Experiment 1, there were more explanation errors made in the selection condition than in the production condition (68% vs. 65%, respectively), but this difference can primarily be attributed to the cliff problems, in which 80% errors were made in the selection

condition and 67% in the production condition. The rocket and tube problems resulted in 80% and 42% selection errors and 78% and 49% production errors, respectively.

For 313 (23%) of the problems, both trajectories and explanations were incorrect, and for 449 (34%) of the problems, both were correct. As before, in very few cases (.2%) was the trajectory incorrect but the associated explanation correct. In the two cases in which this pattern did occur, subjects' drawings (e.g., a curved tube trajectory) did not correspond to their explanations (e.g., the ball takes a straight path). For 568 (43%) of the problems, subjects gave incorrect explanations to correct trajectories. As in the first experiment, the percentage of correct trajectories associated with incorrect explanations produced for the nine problems was computed for each subject and subjected to a 2×2 (experience \times response type) ANOVA. The effect of experience on explanation performance was significant [$F(1,143) = 64.07$, $MS_e = .09$, $p < .0001$]. On the average, the experienced subjects gave incorrect explanations for 33% of the correct trajectories; this figure was 75% for the inexperienced subjects. Also, as in Experiment 1, the percentage of incorrect explanations for correct trajectories was also greater for selection responses (64%) than for production responses (58%), although this difference, as well as the interaction between experience and response type, failed to reach significance.

The percentage of subjects giving incorrect explanations to correct trajectories for each of the nine problems was also computed. Explanation performance on the tube problems was quite good (mean errors, 36%) in comparison with performance on the cliff (68%) and rocket (66%) problems.

Types of explanation errors. As in Experiment 1, the most frequently occurring errors were: (1) 383 errors of omission (43% of all errors), (2) 206 descriptive errors (23%), (3) 128 impetus errors (15%), and (4) 117 surface errors (13%). In addition, 28 of the errors involved the addition of an unnecessary component to an otherwise correct explanation (e.g., adding gravity in the tube or rocket problems), 11 of the errors were cases in which no explanation was given, 4 involved an incorrect combination of components, 2 involved replacing the concept of gravity with weight, and 2 involved anthropomorphizing (e.g., "the ball wants to leave the cliff"). Again, the criteria for impetus explanations were intentionally broad to maximize the chance of finding impetus responses.

In this experiment, 44% of the explanations were descriptive or contained omissions, as opposed to 48% in Experiment 1. Apparently the instructions to give detailed explanations helped only minimally. Therefore, as in Experiment 1, an ANOVA was performed that included error type (impetus, surface, or vague) as a within-subjects variable. The results presented in this section will be focused exclusively on significant effects involving the new factor, error type.

Across all nine problems, the number of impetus and surface errors ranged from 0 to 6 for the experienced subjects ($n = 50$) and from 1 to 29 for the inexperienced sub-

jects ($n = 98$). Vague errors were more plentiful for both groups (range, from 10 to 85), although they were again fewest for the easier tube problems (range, from 10 to 27). The results of the ANOVA indicated that the main effect of error type was significant [$F(2,288) = 133.61$, $MS_e = .32$, $p < .0001$], as were the interactions of error type and experience [$F(2,288) = 4.67$, $MS_e = .32$, $p = .01$], and error type and problem [$F(4,576) = 34.18$, $MS_e = .26$, $p < .0001$], as well as the three-way interaction between error type, experience, and problem [$F(4,576) = 5.95$, $MS_e = .26$, $p < .0001$]. Replicating the patterns found in Experiment 1, surface errors tended to occur more often in conjunction with the tube and cliff perceptual set problems. In general, most of the impetus and surface errors were made by inexperienced subjects in conjunction with the tube and cliff problems. Finally, the four-way interaction of error type, problem, version, and experience [$F(8,1152) = 2.57$, $MS_e = .07$, $p = .009$] indicates that this was especially the case for the cliff standard and cliff perceptual set problems. All other interactions with error type failed to reach significance.

As in the first experiment, a subject was classified as having a naive theory if he/she made the same error on all three versions within a particular problem context and if this error occurred across two or more of the three problems. Ignoring descriptive errors and errors of omission, only 1 of the 148 subjects had a naive theory by this definition. This subject made consistent surface errors on the tube and cliff problems. This can be contrasted with 30 subjects who produced consistent Newtonian explanations in this experiment. No subjects made consistent impetus errors on all versions of more than one problem. A total of 11 subjects made consistent impetus errors on all versions of one problem and 10 made consistent surface errors on all versions of a single problem. These results replicate the finding of explanation inconsistency in the previous experiment. Also, as in Experiment 1, less than half (44%) of the impetus explanations were associated with an incorrect trajectory.

Impetus questions. Although none of the subjects in this experiment explained their trajectory judgments in terms of a consistent impetus theory, all but 3 of the subjects (all experienced) answered one or more of the impetus-relevant questions on the postexperiment questionnaire (Questions 2, 4, 6, 7, 9, 10, and 12, Appendix C) in accord with impetus theory. A 2×2 (response type \times experience) ANOVA revealed that more impetus answers were given by inexperienced than by experienced subjects (means of 4.4 and 2.8 impetus answers out of seven, respectively [$F(1,144) = 40.56$, $MS_e = 1.88$, $p < .0001$]). No other effects were significant. These results suggest that the formal training of experienced subjects was beneficial not only in predicting and explaining trajectories, but also at distinguishing impetus ideas from Newtonian theory. However, it should be emphasized that 33 of the 50 (66%) experienced subjects answered three or more of the seven questions in accord with impetus theory. It is worth pointing out that informal comments from these subjects after the experiment indicated that many of the experienced sub-

jects felt that the impetus questions were too ambiguous to answer. They were particularly troubled by the word *impetus* and would have preferred a more scientific term such as *momentum*. Because the impetus questions were taken directly from accounts of impetus in the literature, these results raise questions about possible ambiguities in such accounts that may lead to spurious inflation of impetus prevalence.

Also, the total number of answers in accord with impetus theory correlated negatively with the total number of correct trajectories [$r(146) = -.48$] and positively with the total number of explanation errors [$r(146) = .60$]. Subjects who tend to do poorly on one measure do poorly on the other. Interestingly, however, the correlation between the number of impetus true-false answers and explanation errors is higher for surface explanation errors [$r(146) = .49$] than for impetus explanation errors [$r(146) = .26$]. An explanation for this pattern will be discussed later.

Discussion

The results of Experiment 2 indicate that the advantage of the selection condition found in Experiment 1 cannot be entirely attributed to the use of test-taking strategies. The overall benefit of the selection over the production condition in terms of trajectory judgments was replicated in Experiment 2 with two instead of four alternatives. In the case of two choices as opposed to four, subjects cannot use judged similarity between the alternatives as a means of selecting one that is most different from the rest or most prototypical. However, in this experiment the effect was much greater for the rocket problems than for the tube and cliff problems. Thus it may be the case that subjects can effectively use test-taking strategies to solve some, but not all, problems. The finding that performance varies greatly over different problem contexts emphasizes the importance of testing knowledge in a variety of problem scenarios.

Instructions to give more detailed explanations were unsuccessful in reducing the proportion of vague explanations. Rather than view the tendency to give descriptive or partial explanations as an artifact of the experimental instructions, it may be more instructive to view it as being representative of subjects' knowledge, which may often be incomplete, inaccurate, and descriptive. The product and process of explaining is an intriguing topic that is certainly worthy of more empirical investigation.

Finally, the inconsistencies found in the trajectory judgments and explanations of Experiment 1 were found again in this experiment. As before, problem context played a role in the type of explanation error. Unlike in Experiment 1, the effect of experience was more pronounced for some types of errors and problems than for others. This difference between experiments may be due to a slightly more knowledgeable group of experienced subjects in Experiment 2 than in Experiment 1. In fact, the experienced subjects in Experiment 1 made explanation errors on 60% of the problems, whereas the experienced subjects in Experiment 2 made explanation errors on 35%.

GENERAL DISCUSSION

The results of these studies indicate that individuals with little or no formal training in physics are more likely to make correct trajectory judgments in a selection response paradigm than in a production paradigm. Also, judgment accuracy was dependent on problem features, even on those features that were not indicative of different Newtonian or impetus principles. Thus, type of response and problem context can be added to the list of other factors that influence motion judgments, including problem familiarity (Kaiser et al., 1986) and the representation of motion (Kaiser, Proffitt, & Anderson, 1985). In general, the finding that performance on abstract problems can be affected by a variety of contextual factors raises questions about the generality of naive impetus theory.

Not only did trajectory judgments vary with context and response type, but the accuracy and type of explanations varied with these factors as well. Variations across explanations were also observed within individuals. For instance, subjects typically gave correct explanations for some problems, impetus explanations for others, and feature-based explanations for others. Again, these variations occurred even across isomorphic cases. In both experiments the degree of inconsistency within subjects was greater for naive explanations than for correct Newtonian explanations. Furthermore, this inconsistency was evident across different measures of beliefs. Specifically, the number of true-false questions that subjects answered in accord with impetus theory in Experiment 2 was more highly correlated with the number of their explanations that were based on surface features rather than on impetus theory. Similar contradictions and inconsistencies within explanations of a single individual have been reported by others for physics (diSessa, 1988; Halloun & Hestenes, 1985; Hojnacki, 1988; Ranney, 1987, 1988; Shanon, 1976) and algebra problems (Payne & Squibb, 1990). In sum, these inconsistencies weaken the proposal that many subjects possess a naive impetus theory of motion that is systematic, general, and coherent.

What is the role of naive impetus theory? These data do not support the proposition that naive impetus theory is general or that it plays a prominent role in the solving of motion problems. Impetus explanations made up only a small portion (15%-16%) of all explanation errors. However, the preceding proposal does not rule out the possibility that some subjects possess knowledge in the form of a well-developed, coherent, systematic naive theory that they apply under specific circumstances. Along these lines, Barsalou (1989) has suggested that intuitive theories, especially ones that are rarely relevant, may be dependent on context much more than on isolated or irrelevant facts or features that are often associated with a concept.

However, from a pragmatic perspective, the value of a concept that is purportedly responsible for only 15% of the errors in trajectory judgment is questionable. Its

value is further undermined if we cannot improve performance in these 15% of the cases by correcting subjects' misconceptions. Results from these studies imply that this would indeed be the case. Not only were impetus explanations relatively rare, but they were unrelated to performance on individual problems. In both experiments, less than half of the impetus explanations were associated with an incorrect trajectory. Because impetus explanations are unrelated to the accuracy of the associated trajectory, intervention to correct misconceptions in these few cases is unlikely to have a major impact on performance.

Where do impetus explanations come from, if not from some naive theory that is also responsible for erroneous trajectory judgments? Some impetus explanations may be linguistic slips, rather than results of genuine misconceptions. The fact that even very experienced individuals occasionally produced impetus explanations supports this possibility. Payne and Squibb (1990) argue in the domain of algebra problem solving that there is a need to distinguish between misconceptions genuinely believed and errors "generated by a desperate student who simply has no idea how to solve the problem" (p. 462).

We propose that explanations for motion are constructed on the fly, with the use of a variety of information. In a few cases, subjects may rely on naive impetus theory, but according to our data this tendency is not widespread. Specifically, we view both the trajectory judgment and explanation tasks as problem-solving tasks in which subjects actively construct a solution from two main sources of information: (1) contextual cues surrounding the problem—including the diagram, the words in the problem description, the question, instructions, other alternatives; and (2) relevant knowledge—including familiar cases, theory (e.g., Newtonian, impetus), general problem-solving strategies, and relevant terminology (e.g., velocity, gravity). Thus, performance is better in the selection condition than in the production condition because the former conveys additional contextual information in the form of alternatives. This constructive view can be contrasted with the view that the solution is directly retrieved from memory in the form of either specific cases or theories (naive or formal).

A similar proposal has also been offered by diSessa (1988), who claims that instead of a theory subjects have "a fragmented collection of ideas, loosely connected and reinforcing, having none of the commitment or systematicity that one attributes to theories" (p. 50). This possibility is appealing, given that there seems to be little need to form a coherent theory to explain object motion. For that matter, theories may only develop through formal training or when phenomena need to be understood well enough for one to make precise predictions or diagnose problems. In the case of everyday motion, such needs seldom arise.

The analysis of the types of explanation errors supports our proposal. Not counting errors of omission or description, explanation errors were based either on the surface features of the problem or on impetus-like misconceptions. Surface errors tended to occur most frequently for the tube

problems and the cliff perceptual set problems. These problems seem to have the most salient surface features of all of the problems and therefore may encourage subjects to rely heavily on contextual cues when they formulate their explanations.

What are the implications of this constructive view? First, this view emphasizes the importance of contextual information, in addition to knowledge. It should not be surprising that context affects trajectory judgments, just as it has been found to affect any number of other judgments. However, this emphasis does allow trajectory responses (both correct and incorrect) to be explained by factors other than experiential or theoretical knowledge.

Research on physics problem solving (e.g., Chi, Feltovich, & Glaser, 1981) suggests that novices are more dependent on context than experts. Indeed, in Experiment 1, experienced subjects made only 14% of the surface explanation errors, but they made 31% of the impetus errors. In Experiment 2, they made 4% of the surface errors and 6% of the impetus errors. The emphasis on contextual information also helps to explain the performance of untrained individuals outside of the laboratory. Just as the selection paradigm provides additional information that facilitates the performance of inexperienced individuals, events in the world are accompanied by an even richer array of features that can serve the same purpose. Thus, the intuition that people know more than prior studies of naive physics have indicated is probably due to observations of people interacting in a contextually-rich world. As diSessa (1988) explains, "students come to physics classes with no theory at all, but instead are used to dealing with the world on a catch-as-catch-can basis, where it is quite fair to change tactics whenever the problem is minutely varied" (p. 61). Finally, the reliance of inexperienced subjects on contextual cues has implications for education. It suggests that instruction should target not only the transfer of correct knowledge and the correction of misconceptions, but also, the distinction between relevant and irrelevant contextual cues.

This constructive view also implies that the trajectory judgment and the explanation tasks can be thought of as two separate tasks and, as a consequence, that they need not be related. The assumption that subjects reason through a problem to make a trajectory judgment and then convey this reasoning process in their explanation is questionable. Trajectory explanations, like trajectory judgments, may be created on the fly as they are needed. A subject may use contextual information in the diagram to make a judgment and may rely on contextual information in the verbal description of the problem to construct an explanation. Therefore, in some cases, the explanation may reveal very little about the basis of the trajectory response. This proposal is consistent with the dissociation found between performance and impetus explanations in these studies, as well as other findings of dissociations between verbal reports and performance (e.g., Broadbent, Fitzgerald, & Broadbent, 1986).

Future research in this area should be directed at identifying the features of problems that are salient to sub-

jects and at determining how contextual features map onto trajectory judgments and explanations. This work would have prescriptive implications for overcoming the bias that untrained individuals have to attend to irrelevant surface features. Also, the results of these studies demonstrate that people not only lack explanatory coherence (Ranney, 1988), but have difficulty providing complete explanations and often provide descriptions rather than explanations. There seems to be surprisingly little research done on the generation of explanations.

In conclusion, these results demonstrate that trajectory responses, even on abstract problems of object motion, can be affected by contextual influences. These effects, along with the infrequent and inconsistent application of impetus ideas, raise questions about the importance and generality of naive theories of motion. In addition, the finding that impetus ideas are unrelated to the accuracy of the associated trajectory judgments undermines the pragmatic value of naive impetus theory. Instead, these results suggest that trajectory judgments and explanations are constructed on the fly, with the use of a variety of problem features and knowledge fragments.

REFERENCES

- ANZAI, Y., & YOKOYAMA, T. (1984). Internal models in physics problem solving. *Cognition & Instruction*, *1*, 397-450.
- BARSALOU, L. W. (1989). Intraconcept similarity and its implications for interconcept similarity. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 76-121). New York: Cambridge University Press.
- BROADBENT, D. E., FITZGERALD, P., & BROADBENT, M. H. P. (1986). Implicit and explicit knowledge in the control of complex systems. *British Journal of Psychology*, *77*, 33-50.
- CARAMAZZA, A., MCCLOSKEY, M., & GREEN, B. (1981). Naive beliefs in "sophisticated" subjects: Misconceptions about trajectories of objects. *Cognition*, *9*, 117-123.
- CHI, M. T. H., FELTOVICH, P. J., & GLASER, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121-152.
- CLEMENT, J. (1982). Students' preconceptions in introductory mechanics. *American Journal of Physics*, *50*, 66-71.
- CLEMENT, J. (1983). A conceptual model discussed by Galileo and used intuitively by physics students. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 325-340). Hillsdale, NJ: Erlbaum.
- DISSA, A. A. (1988). Knowledge in pieces. In G. Forman & P. B. Pufall (Eds.), *Constructivism in the computer age* (pp. 49-70). Hillsdale, NJ: Erlbaum.
- DONLEY, R. D., & ASHCRAFT, M. H. (1992). The methodology of testing naive beliefs in the physics classroom. *Memory & Cognition*, *20*, 381-391.
- FELDMAN, J. M., & LYNCH, J. G., JR. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Applied Psychology*, *73*, 421-435.
- 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.
- JOHNSON-LAIRD, P. N., LEGRENZI, P., & LEGRENZI, M. (1972). Reasoning and a sense of reality. *British Journal of Psychology*, *63*, 395-400.
- KAISER, M. K., JONIDES, J., & ALEXANDER, J. (1986). Intuitive reasoning about abstract and familiar physics problems. *Memory & Cognition*, *14*, 308-312.
- KAISER, M. K., & PROFFITT, D. R. (1984). The development of sensitivity to causally relevant dynamic information. *Child Development*, *55*, 1614-1624.
- KAISER, M. K., PROFFITT, D. R., & ANDERSON, K. (1985). Judgments of natural and anomalous trajectories in the presence and absence of motion. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *11*, 795-803.
- KAISER, M. K., PROFFITT, D. R., & MCCLOSKEY, M. (1985). The development of beliefs about falling objects. *Perception & Psychophysics*, *38*, 533-539.
- KEMPTON, W. (1986). Two theories of home heat control. *Cognition*, *10*, 75-90.
- KREITLER, S., & KREITLER, H. (1981). Test item content: Does it matter? *Educational & Psychological Measurement*, *41*, 635-642.
- LEEPER, R. W. (1935). A study of a neglected portion of the field of learning: The development of sensory organization. *Journal of Genetic Psychology*, *46*, 41-75.
- MCAFEE, E. A., & PROFFITT, D. R. (1991). Understanding the surface orientation of liquids. *Cognitive Psychology*, *23*, 483-514.
- MCCLOSKEY, M. (1983a, April). Intuitive physics. *Scientific American*, pp. 122-130.
- MCCLOSKEY, M. (1983b). Naive theories of motion. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 299-324). Hillsdale, NJ: Erlbaum.
- MCCLOSKEY, M., CARAMAZZA, A., & GREEN, B. (1980). Curvilinear motion in the absence of external forces: Naive beliefs about the motion of objects. *Science*, *210*, 1139-1141.
- 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.
- MCCLOSKEY, M., WASHBURN, A., & FELCH, L. (1983). Intuitive physics: The straight-down belief and its origin. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *9*, 636-649.
- PAYNE, S. J., & SQUIBB, H. R. (1990). Algebra mal-rules and cognitive accounts of error. *Cognitive Science*, *14*, 445-481.
- RANNEY, M. A. (1987). *Changing naive conceptions of motion*. Unpublished doctoral dissertation, University of Pittsburgh.
- RANNEY, M. [A.] (1988, November). *Contradictions and reorganizations among naive conceptions of ballistics*. Paper presented at the meeting of the Psychonomic Society, Chicago.
- SHANON, B. (1976). Aristotelianism, Newtonianism and the physics of the layman. *Perception*, *5*, 241-243.
- TOWNSEND, J. T., & ASHBY, F. G. (1984). Measurement scales and statistics: The misconception misconceived. *Psychological Bulletin*, *96*, 394-401.
- TIPLER, P. A. (1976). *Physics* (2nd ed.). New York: Worth Publishers.
- TVERSKY, A., & KAHNEMAN, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453-458.
- VOSNIADOU, S., & BREWER, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology*, *24*, 535-585.

NOTES

1. The two types of responses manipulated in this paper are referred to as *production* and *selection* responses as opposed to *recall* and *forced choice recognition* tests in order to distinguish the former—tests of existing knowledge—from the latter—tests of material learned in the laboratory.
2. There is general agreement that parametric statistics can be used on ordinal data (Townsend & Ashby, 1984).
3. Kaiser, Proffitt, and Anderson (1985) note that they obtained no difference between selection and production conditions in an unpublished pilot study in which adults made judgments about the tube problem. This is in line with the results of Experiment 1 (see Table 1) for the tube problem, but not for the other three problems.
4. Thanks to Mary Kaiser for pointing out these possibilities.

APPENDIX A
Problem Descriptions Used in Experiments 1 and 2

Tube (Standard, Perceptual Set, and Orientation)

This display shows a thin curved tube. Note that the tube is smooth, even though the graphics may tend to appear rough. In the display you are looking down on the tube. In other words, the tube is lying flat on a horizontal surface. Therefore, gravity is not a factor. A metal ball is put into one end of the tube and the ball is shot out the other end of the tube at high speed. Ignore air resistance and any spin the ball may have. Your task is to determine the correct path the ball will follow after emerging from the tube. After you have finished, please provide a rating indicating your degree of confidence in your answer.

Cliff (Standard, Perceptual Set, and Orientation)

This display shows the side of a cliff. A metal ball is rolling along the top of the cliff and is traveling 50 mph at the point that it leaves the cliff. Consider the path that the ball would follow as it leaves the cliff, ignoring air resistance. After you have indicated the correct path, please provide a rating indicating your degree of confidence in your answer.

Pendulum (Standard and Perceptual Set)

This display shows a metal ball suspended by a string. The ball and the string move in an arc as a pendulum from point A to point B. While the ball is in motion, the string is cut at point C. Your task is to determine the correct path that the ball would take after the string is cut. After you have finished, please provide a rating indicating your degree of confidence in your answer.

Pendulum (Orientation)

This display shows a metal ball suspended by a string. The ball and the string move in an arc as a pendulum from point A to point B. While the ball is in motion, the string is cut at point B. Your task is to determine the correct path that the ball would take after the string is cut. After you have finished, please provide a rating indicating your degree of confidence in your answer.

Rocket (Standard, Perceptual Set, and Orientation)

This display shows a rocket moving along sideways in space. It is moving, with its engine off, from point A to point B. It is not subject to any outside forces or gravity. Its engine is fired at point B and left on for 2 seconds while the rocket travels from point B to some point C. At point C the engine is turned off again and the rocket continues to travel to some point D. Your task is to determine the correct pair of paths (B to C and C to some point D) that the rocket would take. After you have finished, please provide a rating indicating your degree of confidence in your answer.

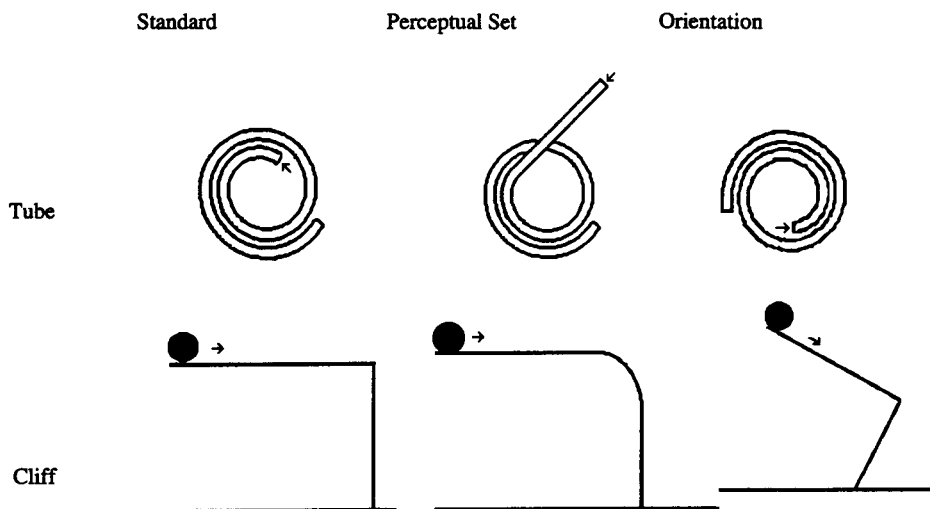


Figure A1. Problem diagrams used in Experiments 1 and 2.

APPENDIX A (Continued)

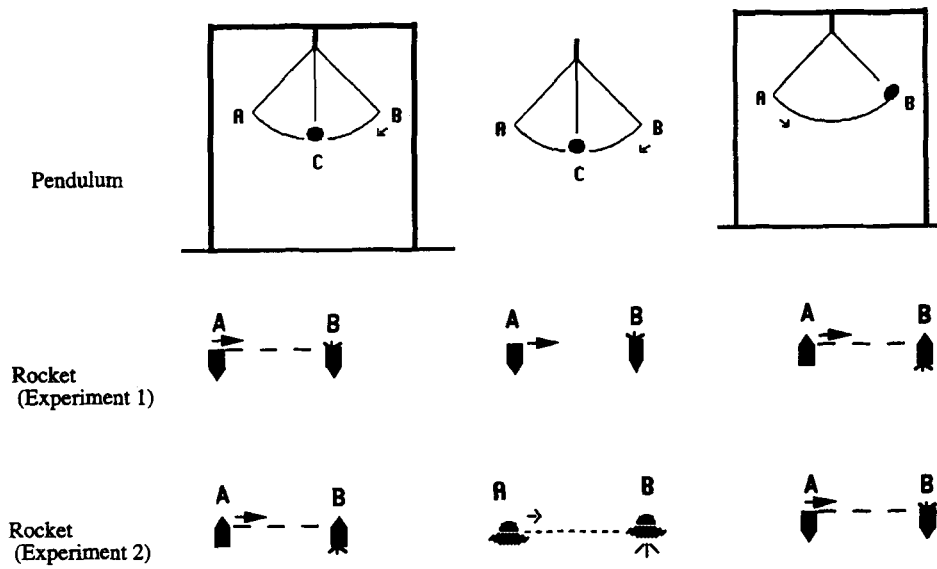
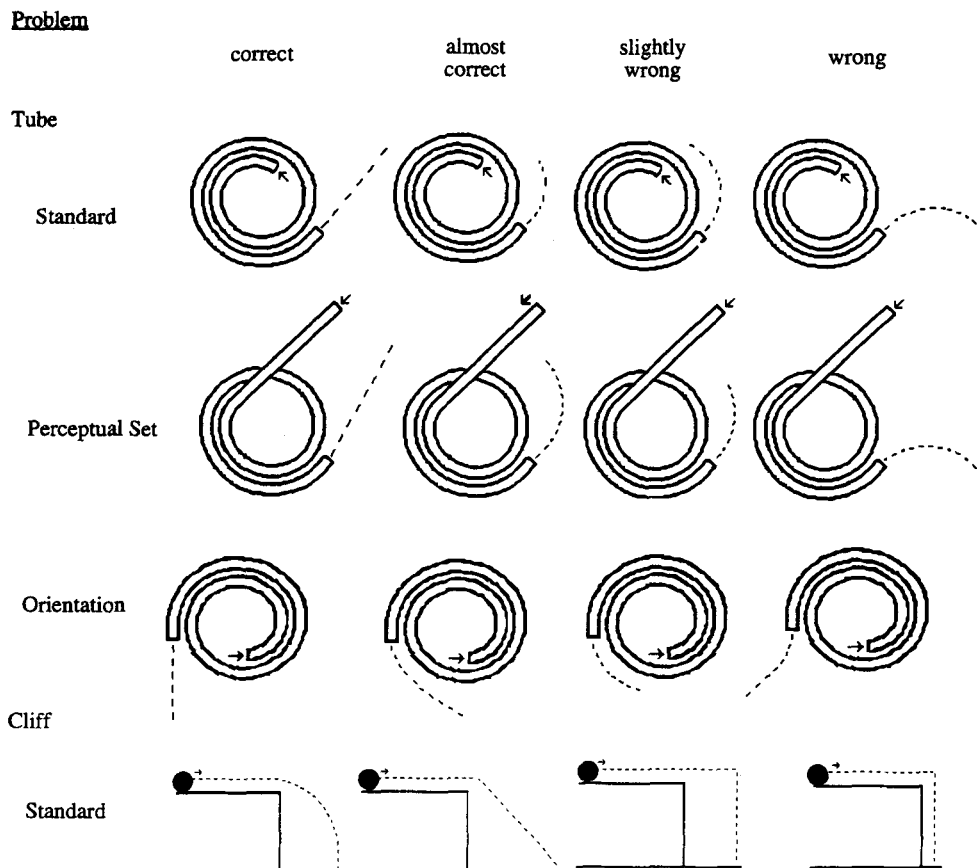
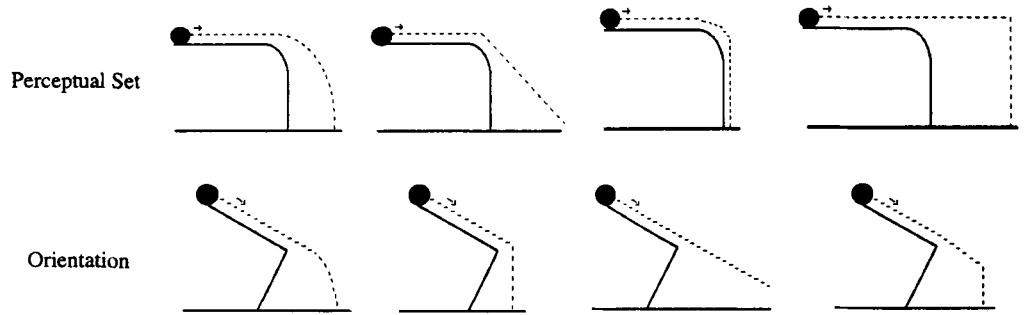


Figure 1A (Continued)

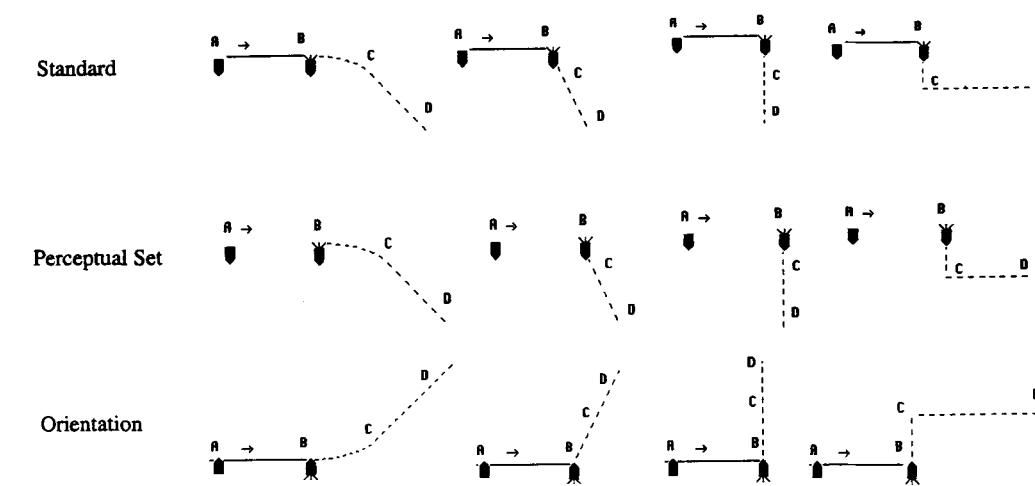
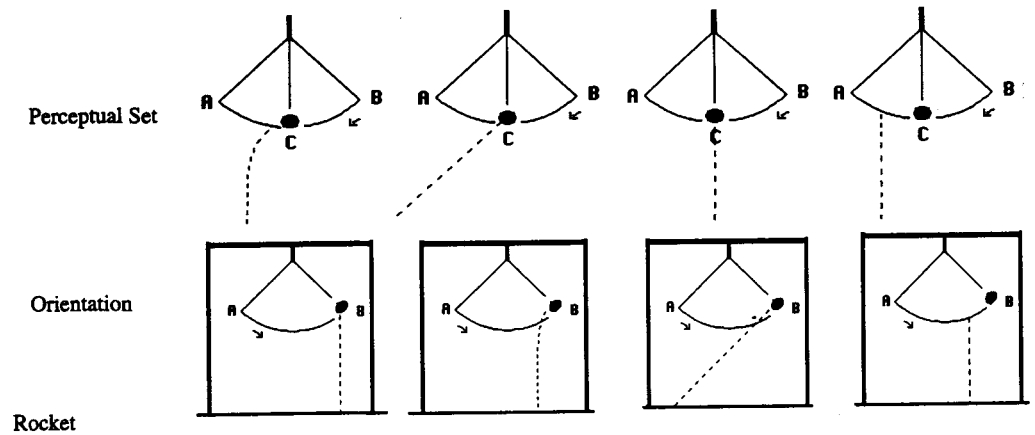
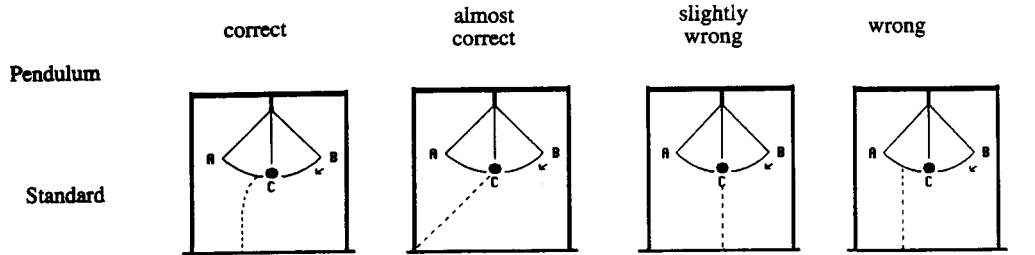
APPENDIX B
Selection Alternatives and Ranks



APPENDIX B (Continued)



Problem



APPENDIX C
Questions Used for Physics Knowledge Scale

1. What are Newton's three laws of motion?
 - a. *A body continues in its initial state of rest or motion with uniform velocity unless acted on by an unbalanced external force.*
 - b. *The acceleration of a body is inversely proportional to its mass and directly proportional to the resultant external force acting on it: $F = ma$.*
 - c. *Forces always occur in pairs. If body A exerts a force on body B, an equal and opposite force is exerted by body B on body A.*

2. Define a newton.

The force required to produce an acceleration of 1 m/s^2 .

3. What is the rate of acceleration due to gravity?

9.81 m/s^2 .

4. Define weight.

$w = mg$.

5. A 3-kg body moves with a speed of 4 m/s in the x direction. What is its momentum?

12 kgm/s

True or False

1. If there are no forces acting on a body, the body will not accelerate. *True*
2. The act of setting an object in motion imparts to that object an internal force or impetus that serves to maintain the motion. *False*
3. If a body is not accelerating, there must be no forces acting on it. *False*
4. There is no qualitative difference between a state of absolute rest and a state of absolute motion. *True*
5. The mass of a body depends on its location. *False*
6. A moving object's impetus gradually dissipates (either spontaneously or as a result of external influences) and as a consequence the object gradually slows down and comes to a stop. *False*
7. A ball rolling along a floor will gradually slow down and come to a stop because friction and air resistance absorb its impetus. *False*
8. The weight of a body depends on its location. *True*
9. A ball travelling through a circular tube acquires directional momentum. *False*
10. Gravity begins to affect an object when that object's internal force becomes weaker than the force of gravity. *False*
11. Action equals reaction only if the bodies are not accelerating. *False*
12. An object remains in motion only so long as it is in direct contact with an external mover. *False*

(Manuscript received September 25, 1992;
 revision accepted for publication August 13, 1993.)