In G.H. Marks (Ed.), *Proceedings of the 1994 International Symposium on Mathematics/Science Education and Technology* (pp. 142-146). Charlottesville, VA: AACE. (1994).

Assessing and Contrasting Formal and Informal/Experiential Understandings of Trajectories

MICHAEL RANNEY
Graduate School of Education, EMST Division, University of California, Berkeley, CA, USA, 94720
E-Mail: ranney@cogsci.berkeley.edu

Abstract: Do people necessarily predict a projectile's motion most accurately when they use appropriate, Newtonian (i.e., curvilinear or vertical) trajectory forms? The present analysis tests this "Joe Montana hypothesis" by contrasting the accuracies of pendular-release trajectories as predicted by subjects using either these "correct" paths or non-Newtonian (e.g., diagonal) paths. Both correlations and Euclidean deviations are employed as measures. Results include the findings that, while Newtonian responses *generally* yield higher accuracies than other responses, in a minority of cases, subjects can achieve superior accuracy by employing some of three aspects of impetus in their trajectory forms (e.g., dissipation, internal force, and curvilinear impetus). Further, the data reveal that women may fare better in naive physics comparisons when more ecological (and less formal) measures of accuracy are employed.

This study focuses on the degree to which differentially rich understandings of Newtonian physics functionally affect students' performance on tasks that involve the prediction of kinematic trajectories. Researchers in the field of Cognition and Instruction tacitly assume that students who have a more correct formal understanding of the physics of motion (say, regarding the "form" of trajectories) will also have a better sense of the *phenomena* of motion, for instance, than laypeople who have merely interacted with moving objects. This need not be the case, however, since relatively untutored individuals can obviously gain considerable kinesthetic facility with such objects. This possibility is what might be referred to (perhaps a bit facetiously) as the "Joe Montana hypothesis"—the suggestion that there is a sense in which the famous (American) football player, Joe Montana, is a "better physicist" than was Albert Einstein, particularly regarding his ability to accurately and quickly predict a projectile's position. (I.e., the "correct" may not necessarily coincide with the most "accurate.")

To empirically contrast these views, predictions about pendulum-bobs that break free from their supporting strings at various swing positions are analyzed (cf. Caramazza, McCloskey, & Green, 1981; Ranney, 1987/1988, 1988, in press; Ranney, Schank, Mosmann, & Montoya, 1993; Ranney & Thagard, 1988; Schank & Ranney, 1992). In this way, the formal properties (e.g., parabolic vs. diagonal, etc.; see Ranney, in press) of such trajectory predictions can be compared with the accuracy of these predictions (e.g., how far they deviate from the veridical trajectories). This study garnered such graphic predictions via mouse-click responses on a computer workstation—a method that greatly improves the accuracy, manipulability, and ease of analysis of the data.

With such data, we can ask—for instance—"Does the knowledge that ballistic objects move in curvilinear trajectories improve one's ability to predict the moment-by-moment positions of such an object, relative to subjects who have at least partially noncurvilinear, often impetus-like, models?" Consider a contrasting example: It is possible that subjects who have partially vertical or (at least partially) diagonal models of ballistic trajectories might have as good (or better) a sense of the magnitude of a projectile's motion as those who have more formally (e.g., parabolic-like) correct trajectory models. (Nb. that, obviously, a well-angled—if theoretically untenable—diagonal can better approximate an ecologically precise parabolic trajectory than can a parabola whose accelerative or linear vector components are either considerably exaggerated or highly underestimated.)

Other questions that can be addressed with this technology include: "Which sorts of pendular releases yield the highest response accuracy?" and "How similar are subjects' predicted trajectories when one compares situations that should have exactly the same basic (albeit sometimes reflected) form?" Due to space constraints, the experiment and analyses described below primarily focus on issues related to the first question; Ranney (1987, 1987/1988, 1988, in press) offers some answers to the second question.

Method

Subjects, Materials, and Apparatus

Forty undergraduate students (twenty male and twenty female) were paid to participate as subjects in this study. None of the subjects had ever taken either a high school or college physics course.

The materials included written instructions and a set of problem-sheets, each of which described a pendular situation that was about to be animated on the video monitor of a Xerox 1108 Lisp computer. The materials also explained how subjects could "draw" trajectories with the computer's mouse, and that they would be watching simulations of the pendular swing prior to the point at which the pendulum bob was supposed to be released from its supporting string. Eight pendular releases were described and animated (with the simulation "freezing" at the point of release): one from both endpoints, two from the midpoint/nadir of the swing (with prior motion both right and left), and two each from both an upswing and a downswing (also with prior motion both right and left).

Design and Procedure

Each subject was asked to predict the trajectories of the eight pendular releases by (a) watching a simulation until the release point and (b) placing "knots" on the monitor's screen with the mouse, such that the knots indicated the released pendulum-bob's position at various times after its release. Each subject was asked to put enough knots on the screen to follow the bob to the ground (about 'four meters' below; subjects usually generated ten or more knots). They could then edit their knots, moving each to another location—or even add new knots. Following this, lines on the screen were interpolated between the knots to let each subject view each trajectory in a more fluid form. Verbal protocols were also garnered, so subjects yielded converging descriptions about their graphically depicted trajectories (cf. Ranney, 1987/1988). When a subject was satisfied, s/he moved on to the next prediction, although miniature images of past predictions (at the side of the screen) were available for inspection, comparison, and change. All influences of any friction or air resistance were, by explicit instruction, to be ignored; hence, the ecologically veridical trajectories represent an idealized, "vacuum-contexted," Newtonian set.

Results

Coding the Data

The coordinates of each knot placed by a subject were recorded for the last trajectory that the subject predicted for a given release situation (i.e., after any modifications), and these were combined into a set. The set was then labelled as representing a particular trajectory "form," based upon the convergence of both the subject's verbal protocols and the appearance of the depicted trajectory. (Inter-rater reliability in the coding of such forms is quite high; see Ranney, 1987/1988.) The correct form for the endpoint releases are those coded as "S" for "Straight-down," while the correct form for the upward-releases is "Ca" for "Curvilinear-arch." "C," for "Curvilinear-down" is the correct form for the remaining (midpoint/bottom- and downward-) releases. (A finer distinction between parabolic and other curvilinear trajectories is unnecessary, as most of the subjects could not generate one.) See Figure 1 for some examples of complex trajectories that result when subjects combine these and other forms.

In contrast to the few correct forms, subjects generated dozens of trajectories—about forty unique codes in total. These include impetus-like variants that both curve down for a while and then fall straight down (C,S, which exhibits "dissipation"), others that initially move horizontally for some time (e.g., H,C, exhibiting "internal force"), and even more unusual ones that involve, for instance, trajectory segments that initially continue to follow a concave-up path (e.g., "Cu,Ca," highlighting "curvilinear impetus"). Some depicted paths even move retrograde relative to the lateral motion at the point of release (e.g., "R,<->" for "diagonally downward, but directionally reversed"). Ranney (1987/1988 & in press) describes these in more detail.

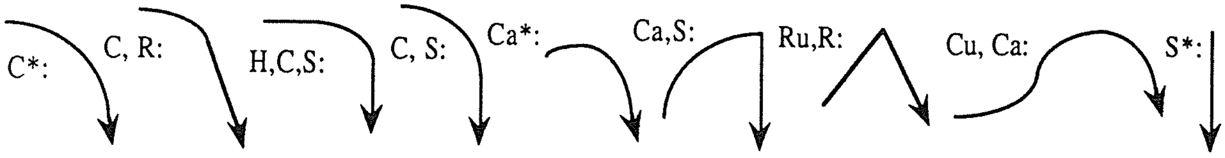

Figure 1. A small sample of subjects' dozens of inappropriate and three appropriate (*) trajectory forms, with their associated codes. (Ca is appropriate for upswing-releases, S for endpoints, C for the rest.)

Overall Performance on the Various Tasks

As subjects were told that the animated pendulum was a meter in length, and since its animation (e.g., period) was veridical, one can compare the set of each subject's knots, for each pendular situation, with the (respective) closest points on the veridical trajectory for the eight pendular releases. Overall, subjects placed their knots (totaling 2,612 coordinate pairs) an average of 261 pixels from a trajectory's origin (range: 4 to 789). The knots deviated from their veridical paths by an average of 93.3 pixels, about 35.8% of the mean distance from the origin. Releases stemming from a rightward swing (i.e., excluding leftward and endpoint releases), which were elicited first, produced less accurate predictions than those from a leftward swing (98.0 vs. 85.4 pixels). (Nb. all differences described in this paper are significant at least at the $p < .05$ level, unless noted otherwise.) Although confounded with primacy, rightward releases also consistently yielded response trajectories that exhibited more lateral motion, compared to leftward releases (for both genders); such directional asymmetries are not uncommonly observed in perceptually related phenomena (e.g., Ranney, 1989, and associated articles on representational momentum and similar effects). Releases during a downswing produced more accurate knots (79.7 pixels' deviation) than those of either (a) lateral releases from the nadir (99.5 pixels) or (b) releases from an upswing (97.5 pixels) or (c) releases from the endpoints (97.1 pixels). These differences are not due to differences in the mean displacements of subjects' knots from the points of release, as these were all comparable (indeed, slightly greater for the more accurate downward-releases). This advantage for accuracy on the downward releases was also mirrored in the likelihoods with which subjects yielded paths of the correct form (cf. Ranney, 1987/1988).

Since these average deviation measures are fairly aggregative (cf. Ranney, in press), more telling are comparisons between instances of subjects who generated the appropriately correct forms and subjects who generated particular *in*correct forms. To illustrate the general frequency of this latter set of forms, it seems worth noting that half (four) of the tasks should have yielded "C" trajectories—yet various subjects generated 21 incorrect ("non-C") sorts of trajectory forms for them. The following sections contain analyses of such diverse responses.

Accuracies Among Formally Correct vs. Formally Incorrect Predictions

On average, only 21% of subjects' responses yielded formally correct predictions (although performance and consistency varies widely between subjects; Ranney, 1987/1988, in press). These subjects placed their knots only about 56.4 pixels away from where they should have been, while the overwhelming majority of subjects' paths, which exhibited incorrect forms, yielded nearly twice that average—a mean deviation of 102.3 pixels. (Again, the effect is not due to variations among the subjects' overall displacements of their knots from the origin.)

Several other patterns emerge from these deviational data. First, much of the correct vs. incorrect difference stems from the advantage one might get from knowing that endpoint releases yield vertical trajectories (Ranney & Thagard, 1988). Subjects who predicted "S" for these releases yielded a trivial average deviation of only 9.5 pixels, whereas subjects who predicted nonvertical trajectories exhibited a mean deviation of 116.4 pixels. Excluding these tasks (i.e., focusing only on the six interior pendular releases) yields a still-significant, but more modest advantage for subjects who provided correct-form trajectories (69.5 pixels vs. 97.8 pixels).

A second pattern relates to subjects who predicted paths that were retrograde with respect to the motion of the pendulum prior to the bob's release. (Verbal protocols show that most, if not all, of these responses were intended to be retrograde; they were not generally "mental slips" regarding the pendulum's prior motion.) Although relatively few in number (about 15 paths), retrograde trajectories yielded an average deviation of 196.9 pixels, and significantly increased the average deviation of paths with incorrect forms from what would have been only 86.3 pixels for the interior releases (and 94.1 pixels if the endpoint releases are included). Excluding both the endpoint releases and any retrograde trajectories, subjects who predicted the correct forms (C and Ca in this restricted set) at the appropriate times generally performed more accurately than those who did not (i.e., 69.5 pixels vs. 86.3)—although the former group's mean deviation is only about 20% less than that of the latter.

"Joe and 'Joanna' Montana Meet Einstein:" The Accuracies of Specific Forms

In general, male subjects were much more likely to depict paths with the correct form (about 33% vs. about 8% for females; cf. other such gender effects in Ranney, 1987/1988). This finding conflicts with conclusions from Kaiser, Proffitt, Whelan, and Hecht (1992), whose forced-choice data led them to suggest that simply animating pendular releases eliminates gender differences. Despite this difference, both genders were statistically equivalent in terms of their average deviation from the veridical paths (93.7 pixels for men, 92.9 for women). This indicates that, in terms of the difference between one's formal knowledge and one's kinesthetic intuition, there may be more "Joanna Montanas" (or perhaps "Martina Navratilovas") than "Joe Montanas."

Of particular interest is whether or not any incorrect forms yielded more accurate predictions than correct forms. Such forms existed for paths that should have yielded either C or Ca forms, but not for S forms. Specifically, subjects correctly yielding Curvilinear responses manifested a mean deviation of 77.5 pixels. Of the two (impetus-laden) incorrect trajectories that yielded significantly smaller deviations, only the C,S path (see Fig. 1) was reasonably frequent—i.e., about as common as the C response itself. The mean deviation for "C,S" depictions was 48.8 pixels, suggesting that people who correctly draw C responses actually overestimate the continuing lateral motion relative to the accelerative downward motion due to gravity. Although the "C,S responders" fall prey to the notion of dissipation (Ranney, 1987/1988, 1988), it actually seems to serve them better than does its absence serve their more Newtonian peers. Still, the "C responders" outperformed the accuracy evidenced from seven other path-forms. Statistical power was not high enough to reject equivalent-mean hypotheses for twelve other forms (if they are indeed inferior). Note that R responses were plentiful, yet, their average deviation (77.7 pixels) was virtually identical to that of the correct C corpus of data; the less frequent H,C and H,C,S responses evidenced even (again, nonsignificantly) lower average deviations (67.4 and 73.0 pixels). The linear segments of these responses are indicative of the internal force type of impetus.

For the upswing releases, which should have yielded arching Ca paths, "Ca responders" outperformed 13 of 24 sets of alternative-form responders, and were themselves outperformed by three such sets. (Statistical power was lacking for determining the remaining eight contrasts, although four each were numerically lower than and higher than the mean Ca deviation.) In particular, the proper Ca paths yielded a mean deviation of 51.4 pixels, in contrast to those who offered the more accurate C paths (again, for Ca-appropriate releases; 14.3 pixels), "H,C" paths (14.4 pixels), and "C,R" paths (21.7 pixels). Although each of these incorrect forms were low in frequency, they each benefit from the fact that a release during an upswing generally has a fairly minimal segment of upward motion before the acceleration due to gravity moves the resultant velocity vector downward. Subjects responding with a Ca path generally overdramatized this upward segment, so some subjects who included either horizontal segments (e.g., H,C), well-angled diagonal segments (C,R—which is also indicative of the "internal force" type of impetus; Ranney, 1987/1988, 1988), or even fairly broad curvilinear segments (C) could outperform the subjects who responded with a more appropriate (but often over-) arching Newtonian form.

For the endpoint releases, which prove very difficult, no incorrect path-forms could match the vertical S form. Of 19 alternative forms, all but one had higher average deviations than the 9.5 pixels evidenced by subjects drawing the S trajectories. This obviously reflects the fact that nonvertical trajectories deviate considerably from easy-to-draw vertical trajectories. The only form that could not be rejected as inferior (again, probably due to the low power afforded by infrequent responses), included trajectories that curved rapidly into verticality (C,S).

Correlations Among Formally Correct vs. Formally Incorrect Predictions

A different measure of how well subjects' predictions approximate (most) correct trajectories is represented by the correlation between the y-coordinates of a subjects' knots and comparable y-coordinates from the veridical parabola (based upon the x-coordinates of subjects' knots). (This measure also controls for a potential, but not presently manifested, problem with the average-deviation measures reported above: that subjects may have placed more knots closer to the various release origins for some, especially nonlinear, predicted forms.) Since vertical (S) paths contain no variation in the horizontal dimension, correlations with such forms are improper; therefore, the following analyses are only with respect to the veridical parabolic trajectories for the six interior pendular releases (i.e., for releases other than those at the endpoint).

For the "appropriately-C" data (i.e., from subjects yielding "C" predictions for downward and lateral releases), the knots correlated at $r = .62$ with appropriate coordinates from the veridical, parabolic, "C" trajectories. This correct form was significantly better correlated than the inappropriate forms "Cu,S", "Ca", "S", and "C,S" (.14, .18, .20, and .39). The "C" form subjects' data were also marginally better correlated than those who predicted the diagonal "R" form ($r = .42$; $p = .06$), a finding that conflicts with one "Joe Montana" prediction explicitly mentioned above. Largely due to relatively few responses per "non-C" form, the remaining 16 inappropriate trajectory forms were not significantly more or less correlated with the veridical parabolic trajectory than the .62 from the "C" subjects. Even so, one odd and rare trajectory involving curvilinear impetus (e.g., Ranney, 1987, 1987/1988)—"Cu,R"—almost achieved such statistical significance, with a surprising $r = .99$ ($p = .076$).

The story was much the reverse for parabolic trajectories of the "Ca" form, largely because the period of upward movement from an upward pendular release is so short that subjects who focus on this portion often greatly exaggerate it. As a result, the "appropriately-Ca" subjects' data correlated only at .22 with the veridical, parabolic, "Ca" trajectories. In contrast, four of the 19 inappropriate forms obtained from the upswing tasks ("R," "C,R," the triangular "Ru,R," and the near-veridical "C") were significantly more highly correlated ($r = .61$, .76, .91, and .96); another form (H,C) was marginally superior ($r = .925$, $p = .07$).

Discussion, Conclusions, and Implications

These results suggest that subjects with more appropriate notions about the forms of ballistic trajectories generally—but certainly not always—yield more accurate and/or better-correlated predictions, compared to subjects who offer inappropriate (and sometimes almost bizarre) forms. Acting Newtonian was a distinct advantage for responding to the endpoint-release tasks, and was usually advantageous for the interior positions. However, by some measures, various paths indicative of one or more of three forms of impetus beliefs (dissipation, internal force, and curvilinear impetus) yielded responses that were either superior or equivalent to the correct forms. One modulating factor for these findings is the relative difficulty of the tasks involved. For instance, the situations that yielded the most evidence in favor of the alternative (and primarily disconfirmed) "Joe Montana hypothesis" was also one of the more difficult; subjects' trajectories for these upswing pendular releases yielded the lowest overall correlation between predicted knots and points from the veridical parabolae ($r = .14$, compared to .27 to .79 for the other types of releases), and manifested one of the highest deviation/displacement ratios (40%). Another factor is the overall approximate predictive similarity of many of the generated trajectories (e.g., a triangular vs.

an arching form) over distances that are perhaps modest in scale. Yet another factor involves relative expertise; more sophisticated subjects would probably increase their predictions' relative accuracies, especially if such subjects were offered measurements and less qualitative situations. Even so, given that some of these subjects who seem to best predict where a projectile will be at a given moment do so via considerably non-Newtonian drawings and explanations, we would do well to continue to seriously consider the question, "Does a theory rest in one's actions, or one's conceptions, or both?" (See Ranney, in press, for an elaboration on aspects of this question.)

Methodologically, this study demonstrates and highlights the utility of garnering graphically oriented "naive physics" data (in addition to other sorts of graphic data; Ranney & Reiser, 1989) via an interactive computer system. Many experiments continue to under-utilize the potential of such technology, often due to a variety of practical or theoretical apologia that may not hold up under closer scrutiny (Ranney, in press). The present method also offers other perspectives on real or alleged gender differences. Ranney (1987/1988) found many gender effects in the naive physics of motion (all favoring males—sometimes by a four-to-one ratio); still, these results also indicated that some disparities disappear (as shown by post-tests) with either more experience with the phenomena or nontheoretical feedback. Furthermore, the analyses reported presently show that women may perform as accurately as men in a rather ecological sense, even if their theoretical constructs may not be as formally "correct."

Finally, beyond the inherent interest to academic psychologists and cognitive scientists, the results of these analyses should interest science educators; the findings offer hints regarding the ways in which one should try to integrate a student's theoretical and pre-theoretical (e.g., informal, experiential, and kinesthetic) knowledge to yield both more optimal and more motivating educational curricula.

References

Caramazza, A., McCloskey, M., & Green, B. (1981). Naive beliefs in "sophisticated" subjects: Misconceptions about trajectories of objects. *Cognition, 9*, 117-123.

Kaiser, M.K., Proffitt, D.R., Whelan, S.M., & Hecht, H. (1992). Influence of animation on dynamical judgments. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 669-689.

Ranney, M. (1987). *Restructuring Conceptions of Motion in Physics-Naive Students.* Manuscript of the paper presented at the annual meeting of the American Educational Research Association.

Ranney, M.A. (1988). Changing naive conceptions of motion. (Doctoral dissertation, University of Pittsburgh, Learning Research and Development Center, 1987.) *Dissertation Abstracts International, 49*, 1975B.

Ranney, M. (1988). *Contradictions and Reorganizations Among Naive Conceptions of Ballistics.* Manuscript of the paper presented at the annual meeting of the Psychonomic Society.

Ranney, M. (1989). Internally represented forces may be cognitively penetrable: Comment on Freyd, Pantzer, & Cheng (1988). *Journal of Experimental Psychology: General, 118*, 399-402.

Ranney, M. (in press). Relative consistency and subjects' "theories" in domains such as naive physics: Common research difficulties illustrated by Cooke and Breedin. *Memory & Cognition.*

Ranney, M., & Reiser, B.J. (1989). Reasoning and explanation in an intelligent tutor for programming. In G. Salvendy & M.J. Smith (Eds.), *Designing and using human-computer interfaces and knowledge based systems* (pp. 88-95). New York: Elsevier Science Publishers.

Ranney, M., Schank, P., Mosmann, A., & Montoya, G. (1993). Dynamic explanatory coherence with competing beliefs: Locally coherent reasoning and a proposed treatment. In T.-W. Chan (Ed.), *Proceedings of the International Conference on Computers in Education: Applications of Intelligent Computer Technologies* (pp. 101-106).

Ranney, M., & Thagard, P. (1988). Explanatory coherence and belief revision in naive physics. *Proceedings of the Tenth Annual Conference of the Cognitive Science Society* (pp. 426-432). Hillsdale, NJ: Erlbaum.

Schank, P., & Ranney, M. (1992). Assessing explanatory coherence: A new method for integrating verbal data with models of on-line belief revision. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society* (pp. 599-604). Hillsdale, NJ: Erlbaum.

Acknowledgements

Special thanks go to Lauren Resnick, Patricia Schank, Mary Kay Brooks, Susan Hojnacki, Ernest Rees, and Joshua Paley. Comments, suggestions, and help were also graciously provided by Stellan Ohlsson, Christopher Hoadley, George Montoya, Christine Diehl, Collin Mulquiney, Michelle Million, Bernadette Wilkin, Jonathan Neff, Elijah Millgram, Joshua Gutwill, the UC-Berkeley Reasoning Group, and others. Funding for this research, for which I am most grateful, was provided by the National Academy of Education and the Spencer Foundation (with a Spencer Postdoctoral Fellowship), the Department of Energy (with their Teacher Research Associate Program), the Office of Naval Research (with grant N-00014-85-K0337 to Lauren Resnick and Stellan Ohlsson), and the Committee on Research of the University of California, Berkeley (with a Faculty Research Grant).