Individual-Centered vs. Model-Centered Approaches to Consistency: A Dimension For Considering Human Rationality

Michael Ranney
University of California, Berkeley, USA
Email: ranney@cogsci.berkeley.edu

In many domains (e.g., "naive" physics), consistency represents an implicit test for human rationality. Unfortunately, consistency measures vary so widely that people may seem more or less rational as more measures are engaged. This article describes a dimension of consistency (and/or coherence) for considering these diverse measures. At the "model-centered" extreme, subjects' data are contrasted, regarding consistency, with a domain's formal models (e.g., Newtonian mechanics, impetus, etc.). Toward the opposite extreme, individuals' data are contrasted with their own consistency criteria. For instance, if someone predicts that two trajectories should be identical, then draws two paths with dramatically different forms, such performance would indicate inconsistency at even the individually-centered dimensional extreme. More than a dozen examples of consistency/coherence measures are herein located along this I-M ("Individual- vs. Model-centered") dimension. Although model-centered measures have some descriptive utility in assessing theories, sensitive individually-centered measures seem better for approximating answers regarding human rationality.

1 A Dimension for Considering Human Rationality

Human cognition about motion has been a highly popular field of research [e.g., Chi et al, 1981; Halloun and Hestenes, 1985; Larkin et al, 1980; White, 1993]. Adding to our entertainment as researchers are results that seem to indicate that certain configurations of incorrect responses are produced by laypeople and/or our scientific ancestors [e.g., Caramazza et al, 1981; Clement, 1983, diSessa, 1982; Nersessian and Resnick, 1989]. In a recent attempt to enrich the debate about whether subjects (naive or not) have consistent and/or coherent models [Ranney, 1994a], I briefly suggested a dimension that seems useful for considering patterns of subjects' behavior. This dimension is essentially a bipolar theoretical construct, anchored at the extremes by highly model-centered (i.e., historical or formal) measures of consistency and highly individually-centered ("personal") measures of consistency. In this article, I explicate this "I-M" (Individual- vs. Model- centered) dimension, by (a) illustrating its utility by placing various measures of consistency along its length and (b) explaining why we are wise to remember this continuum when considering questions of subjects' rationality. It is suggested that such considerations often (at least implicitly) dominate our conclusions about what we call "consistency."

2 Some Empirical Background

The I-M dimension largely developed from my past work on problem solving and reasoning about physical motion. Initially, I was intrigued by "non-physics" problems for which some subjects employed pendular constructs in their solution [e.g., Maier, 1930]. This led to some studies of middle school science teachers, which indicated that a consistent understanding of the simple harmonic motion of pendulums is quite difficult [Ranney, 1987, 1987/1988].

Later experiments focused on studying whether and how laypeople could relate the trajectories of pendulum bobs (which were released from their pendulums at various points in their swings [Ranney, 1994b]) to more common situations involving dropping and throwing. In an attempt to foster conceptual restructurings in these subjects, animated ballistic feedback was sometimes provided (a) for the pendular release situations (e.g., a good animation of a parabolic path from a midswing-release), and (b) on the relationships among the pendular and dropping/throwing items ("similarity judgments;" e.g., that something dropped by a walking man has the same basic trajectory as a bob released from the nadir of a swing). Transfer-of-training was assessed over both time and similarity to the pendular situations (e.g., with pendular isomorphs such as trapezes and wrecking balls), including more abstract queries regarding pendulums (and even various zero-gravity release situations) [Ranney, 1987/1988]. Subsequent studies have expanded on such work in a variety of ways [Ranney, 1998, 1994a, 1994b; Ranney et al, 1993; Ranney and Thagard, 1988; Schank and Ranney, 1992, etc.].

3 Consistency as a Contextually Bound Measure of Concordance

The I-M dimension is not meant to prove or disprove whether individuals are consistent on certain tasks, since consistency seems to be relative to the overarching context and temporal window a researcher chooses [cf. Collins and Gentner, 1987]. It would be absurd to expect the algebraic performance of a five-year-old to be consistent with his/her performance at age 35. As the context and time window narrow, though, we would expect more consistency; barring neuropathology or feedback, a 35-year-old will perform fairly consistently on an identical arithmetic problem when represented, for instance, with only a 30-second delay.

As argued elsewhere [Ranney, 1994a], consistency is not a step-function with which people's performances can be dichotomized as "consistent" or "inconsistent" [cf. Cooke and Breedin, 1994].[1] Furthermore, whenever measures fall between perfect consistency and randomness, we seem called upon to generate subjective ("gut") predictions and assessments about the individual or group. This seems true in many studies of concordance. In behavioral genetics, say, future scientists may tell us that one's genome determines 40% of the variance of all behavior (e.g., the highest academic degrees attained by identical zygotes that develop fully apart). We could then conclude either, "Well, genetics certainly determine a lot about behavior," or the equally defensible, "Well, genetics have only a minor influence on behavior." These reactions, when modulated by a priori expectations about a result, yield a 2 X 2 matrix of ways to interpret such data, namely with delight or dismay that the metaphorical "glass of concordance" is found partially "filled" or partially "empty." In many domains of study (e.g., regarding the statistical "base-rate fallacy"), researchers often seem to argue issues based upon which cell of the matrix their subjectivity has placed them in. In naive physics, the delight takes form in the "theory theorists," who suggest that laypeople have fairly consistent models of motion, while the dismay finds form in the "fragmentary cognition" camp, who suggest much less consistency [Ranney, 1994a].

[1] This is at least partially because, as Collins and others have similarly pointed out, as one employs increasingly fine levels of analyses, the result is a host of models that may appear, in aggregate, as far fewer global models [Collins, 1985; Collins and Gentner, 1987; Collins and Stevens, 1984]. The result is that inter-subject consistency seems less likely (and intra-subject consistency more likely) to be observed as researchers strive toward the ultimately impossible goal of fully specifying an individual's mental model (see [Rouse and Morris, 1986], on mental models' identifiability).

4 Uses for the I-M Dimension in the Face of the Relativism of Consistency

While a context-free measure of consistency seems rather useless in a dynamic world (and one with Heisenberg's "uncertainty principle," too [Rouse and Morris, 1986]), we can still find a relative notion of consistency useful. In contrast to researchers who take the "consistency hypothesis" to be a rejectable null hypothesis, absolute consistency should be viewed as beyond empirical rejection. Otherwise, we overstep our data; we cannot reject the possibility that subjects are perfectly consistent and are merely employing a set of organizing principles and/or situation features that elude us as researchers [cf. Collins and Stevens, 1984]. This aspect of the curve-fitting problem merely allows us to suggest that subjects do not seem to consistently apply a particular model (e.g., a favorite variety of impetus theory). However, we may be justified in rejecting various inconsistency hypotheses when (a) the hypotheses entail predictions of randomness and (b) the obtained data are actually significantly modeled by some organizing principle(s).

This article discusses how consistency can be considered regarding the degree to which a proposed organizing principle/model (1) is primarily imposed by the researcher (from pilot studies, the literature, etc.) or (2) emerges out of an individual's responses. Further, as illustrated below, the more one uses the individual's responses to develop consistency measures, the closer one might come to the (impossible) ideal of "rejecting" the consistency hypothesis.

4.1 A Sample of Findings from Measures Along the I-M Continuum

In various studies on motion, my colleagues and I have used measures that cover a good deal of the I-M continuum [e.g., Ranney, 1987/1988, 1988, 1994b; Ranney and Thagard, 1988; Schank and Ranney, 1992; cf. Chi, 1992, pp. 161-162], whereas most other experiments in the literature have focused on model-centered consistency measures. The following three sets of examples from our research are used to gradually illustrate progressively more individually-centered (and less model-centered) notions of consistency.[2] (As should become apparent, the results demonstrate to the author's "gut," perhaps that typical levels of subjects' consistency were rather low.) In each case, these young adults (largely naive of formal physics) provided both drawn trajectories and oral explanations for dozens of tasks involving pendular releases and various isomorphic situations (see above)including dropped, thrown, released, and swung objects (e.g., [Ranney, 1988], unless noted otherwise).

4.1.1 Generally Model-Centered Consistency Measures

Spotty Newtonianism
Regarding subjects' Newtonian accuracy, it was observed that only two of six possible pairwise correlations among four sets of normal gravity (pendular-release, swinging-transfer, similarity-judgment, and dropping/throwing) tasks were statistically significant. (An "accurate" response is rather objective here i.e., it exhibits an appropriate trajectory form, such as an arching or vertical or downward- curving path; see Figure 1, and [Ranney, 1994b].) The significant correlations were between (a) the pendular-release tasks and the dropping/throwing tasks ($r = .37$), and (b) the pendular-release tasks and the (wrecking ball or trapeze-artists) near-transfer tasks ($r = .42$). Higher correlations among these (and the other) pairs of accuracy metrics would have been indicative of greater consistency in subjects' responding; but the results show that an individual's performance across one set of tasks was generally either a weak or poor predictor of his/her performance on other task-sets. Obviously, the fully Newtonian orientation for the accuracy metrics in these analyses mean that they are model-centered measures of consistency.

Limited ability to use transitive reasoning

After some subjects received nontheoretical feedback for a situation (i.e., the correct trajectory and a list of the isomorphic, "fundamentally similar," relations among a set of situations), only half of the time could subjects use the information in a transitive, quasi-syllogistic, consistent; way (e.g., by subsequently correctly predicting the trajectories for isomorphic situations). For instance, given feedback (a) that a release from the nadir of a pendulum yields a downward-curving trajectory, and (b) that such a release is fundamentally similar to a heavy ball dropped by a walking man, subjects should conclude that the heavy ball should also curve down [Ranney, 1987/1988]. Such reasoning is also based upon a model-centered, Newtonian, perspective.

Low consistency over isomorphic pendular and dropping / throwing tasks

Only 20% of the time did the forms of subjects' pendular release predictions (e.g., coded as in Figure 1) match their trajectory predictions for isomorphic dropping/throwing problems in which "match" means "contain the same features." ([Ranney, 1994b] describes the relationships between formally accurate and formally

[2] The "intermediately placed" consistency measures employ models that are less clearly Newtonian (or even impetus-like), and thus often involve less correct modelsor even aspects of seemingly bizarre situation-, apparatus-, or speed/magnitude-specific models (cf. "high nonvertical release velocity" below [Hojnacki, 1988; Ranney, 1988]).

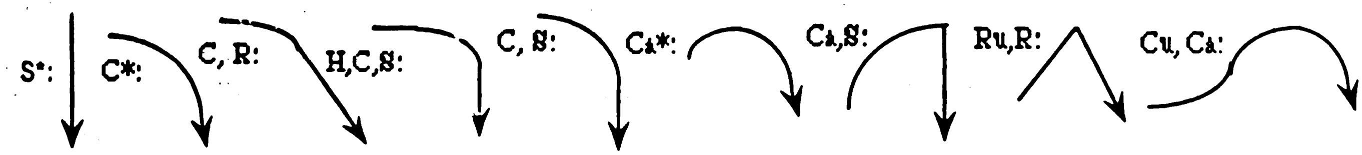

Fig 1 Some subjects' trajectory predictions, with codes that represent the features of the trajectories (* = an accurate trajectory form for some dropping, throwing, and pendular/swing release situations).

inaccurate trajectories and their Euclidean deviations from veridical parabolic trajectories.) This measure is primarily based upon a Newtonian perspective of isomorphisms.

4.1.2 Generally Intermediately Placed Consistency Measures

Low consistency among swinging objects

Strikingly, when compared to a "trapeze" problem, only 19% of the time did isomorphic pendular-release situations elicit the same trajectory from the same subject. (eg., s/he may predict that a trapeze artist released by another during midswing will curve downward yet later indicate that a pendulum bob released from midswing falls purely vertically.) This measure largely employs a Newtonian notion of isomorphism, but is consistent with other models, as is true of the following measure.

Inconsistencies among descriptions of pendular speed and responses to isomorphs

Oral, abstract, descriptions of variations in a (mentally animated) swinging pendulum's speed were inconsistent with predictions subjects made for near-transfer trapeze and wrecking ball problems 29% of the time. (E.g., subjects sometimes claim that a wrecking ball maximizes its impact by striking a wall at the endpoint of its natural swing yet also assert that a pendulum has no speed at its endpoints.) This result, like the following result, yields a rare encouragement for readers hoping to find relatively high response consistency.

Errors regarding the symmetry of reflecting pendular releases

If a subject drew a rectilinear-diagonal ('R') trajectory for a pendulum bob released from the nadir of a leftward swing, s/he presumably should draw the trajectory's mirror image for a nadir release during a rightward swing. In fact, 26% of the time subjects responded asymmetrically (i.e., with different trajectory forms for release situations that should yield mirror-image trajectories from a release-velocity-vector perspective; see Figure 2). Note that subjects even had their prior pendular predictions saliently visible on the CRT used for responses. This metric employs a Newtonian perspective even less than the preceding ones, because many alternative models also employ notions of such symmetry. Hence, this measure seems to represent a fairly intermediate position on the I-M dimension.

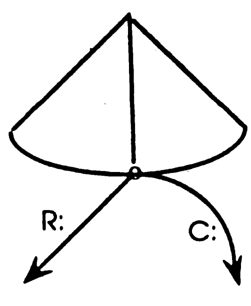

Fig 2 Predictions that violate the right/left symmetry of two pendular release situations.

Inconsistencies between judgments of velocity and non-vertical trajectories

Most of the time (75%), subjects who drew vertical ("straight-down;" [McCloskey et al, 1983]) trajectories from a pendulum-swing's nadir also (surprisingly) chose the nadir as the optimal point for the wrecking ball to strike and damage a building as the "point of maximum velocity." Such measures involve violations of any models that assume the relation between a high nonvertical release velocity and subsequent lateral motion.

4.1.3 Generally Individual-Centered Consistency Measures

Evidence of inconsistency across comparable kinematic situations

More than half of the subjects who (incorrectly) predicted apex/endpoint release trajectories as being other than just "straight down" also mentioned (in a more neutral context) that the pendulum is at rest at the apex. (Cf. the initial parts of the "Hal" simulation in [Ranney and Thagard, 1988].) Note that this apparent incoherence is with respect to any models that presume that lateral trajectories can only result from (non-zero) non-vertical release velocities.

A confused organization of kinematic phenomena

Based upon similarity judgments from 42 subjects, a low-stress multidimensional scaling (MDS) analysis revealed two largely uninterpretable dimensions regarding relationships among the dropping and throwing phenomena (see Figure 3) especially relative to the dimensions yielded by subjects who received some feedback. As a group, then, subjects viewed the situations as fairly isolated entities, again, relative to post-feedback (and certainly, experts') multidimensional scalings. Such a finding is quite centered on subjects (albeit not individual subjects), as opposed to models; it involves no commitments to accuracy, symmetry, release vectors, coding schemes, or notions of isomorphic trajectories.

Low "behavioral agreement"

Only 31% of the time did subjects draw the same trajectory form (e.g., coded as in Figure 1) for pairs of situations that subjects themselves claimed were isomorphic ("fundamentally similar"). This measure seems highly individually-centered in nature, as it even welcomes pairs that are isomorphic from neither a Newtonian nor a typical impetus perspective. For instance, even if some subjects drew an arching trajectory for both an apex release and a

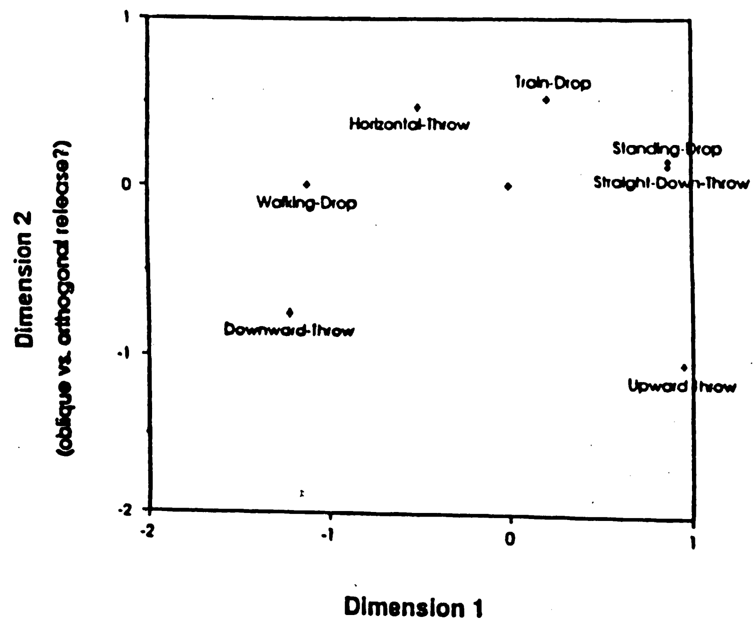

Fig 3 A multidimensional scaling of some (pre-feedback) dropping-and-throwing tasks' similarity data.

downward throw, their behavioral agreement would be fine if they also claimed that the two situations were isomorphic or fundamentally similar. (The measure is, thus, a kind of analog to social psychological metrics that relate attitudes to behaviors.)

Coherence measured via ECHO models of the Theory of Explanatory Coherence

In addition to the preceding exemplars of measures, my colleagues and I have been simulating subjects' beliefs about motion largely without relying on any physics theory (e.g., [Ranney and Thagard, 1988], using the ECHO model). For instance, rather than adopting a physical framework, we might employ our "bifurcation/bootstrapping" method, in which we elicit oral explanations for each belief, prediction, or alternative that a subject considers and verbalizes [Ranney and Schank, 1995; Ranney et al, 1993; Schank and Ranney, 1992]. Each explanation specifies which beliefs are (a) explained, (b) do the explaining, and (c) contradict one another. Although this method is still developing, we have used it to predict subjects' believability ratings for both their verbalized propositions and a set of alternative trajectories (including their initial prediction). We are testing ways to extend, streamline, and automate the bifurcation/ boot- strapping method with the development of a computer-based "reasoner's workbench" [Ranney, in press; Ranney et al, 1995] that, among other things, assists subjects in explicating their physics arguments. (The computer program is called Convince Me; e.g., [Ranney and Schank, 1995; Ranney et al, 1994; Schank and Ranney, 1993].) Thus, we combine the subjects' on-line theorizings (such as they are) with a general model of belief evaluation (ECHO; [Ranney and Thagard, 1988; cf. Hoadley et al, 1994, and Thagard, 1989, 1992]).[3]

[3]Almost all of this section's many measures of subjects' consistency (or inconsistency) utilized verbal protocols from semi-structured interviews [cf. Gutwill et al, 1996]. Even in using accuracy measures, oral descriptions were often critical. For instance, they disambiguated potential interpretations of the graphical depictions offered by subjects, yielding high intercoder reliabilities [e.g., Ranney, 1987/1988]. The disambiguations were also critical in comparing the predicted trajectories of subject- alleged, Newtonian, and "mirror-image" isomorphs. Further, verbal responses to semi- structured interview probes allowed for more precise tallies of the types of impetus evidenced by subjects (e.g., curvilinear impetus, dissipation, and internal force, as in [Ranney, 1987]; cf. [Ranney, in 1994a]). This convergence of measures is especially critical in the domain of motion, in which difficult-to-verbalize perceptual (and other) characteristics seem unavoidable (cf. stimulating work by Diane Halpern, Timothy Hubbard, Michael McCloskey, Dennis Proffitt, Catherine Reed, Benny Shannon, Norman Vinson, and colleagues; e.g., [Freyd et al, 1988; Kaiser et al, 1986; Ranney, 1989]).

5 Examples of Other Measures to Place Along the I-M Dimension

To further illustrate the I-M continuum, results from Hojnacki and Resnick's work [e.g., Hojnacki, 1988] are useful. They also looked to assess consistency in subjects' naive conceptions of motion; they considered many problem situations and features, as well as (a priori and a posteriori) dimensions that might relate to subjects' consistency. Together, we generated a metric of random consistency for their data, which indicated that the subjects exhibited significantly non-random performance although they were only a fraction of the way from random consistency to perfect consistency. This measure, involving Newtonian isomorphs, near the "model-centered" end of the dimension not at the extreme, perhaps, but beyond the (intermediate) symmetry measure mentioned above.

Another measure from this study involved temporal consistency, which falls toward the "individually-centered" end of the I-M dimension. After nearly a month's delay, Hojnacki's subjects were unexpectedly presented with the same (and new) qualitative physics problems. In contrast to the low, barely-above-random consistency from the modeling perspective, the subjects were strikingly more temporally consistent closer to perfection than to random responding [cf. Ranney, 1994a].

6 Consistency and Naive Physics

The case of motion seems instructive in discussing the I-M dimension. One noteworthy feature of the continuum is that group-derived data need not necessarily be placed at the model-centered extreme. For example, the group-derived multidimensional scaling is clearly not a strongly model-centered measure. Another feature of the dimension is that data that indicate more consistent responses do not necessarily cluster at either extreme (but recall footnote 1). From the model-centered end of the continuum, we see that (a) a fairly small proportion of the variance of subjects' (Newtonian) accuracy is accounted for by their accuracy on related problem-sets, (b) quasi-syllogistic reasoning is correctly performed only about 50% of the time in these ballistic contexts, and (c) four measures (including one from [Hojnacki, 1988]) involving fairly Newtonian notions of isomorphism generally yielded considerable inconsistency. On the other hand, the fairly medially placed measure regarding mirror-image symmetry and Hojnacki's highly individually-centered temporal measure indicate considerably greater consistency in subjects' responding. Even so, most of the remaining, largely individually-centered measures, continued to exhibit the overall pattern of considerable inconsistency in responses to probes about physical motion (e.g., "low behavioral agreement").

7 Rationality and Consistency

Many researchers from various fields (e.g., economic, behavioral, evolutionary, and statistical domains) have proposed differing criteria for what constitutes rationality. I suggest that consistency plays a major role in many of these (and more informal) notions of rationality. Therefore, we might ask, "Which end of the I-M dimension captures the best measures of human rationality?," or "How might dimensionally divergent measures complement each other?"

Clearly, model-centered measures have their uses; as with the examples above, we can assess Newtonian consistencies of novices or experts with several metrics from that extreme of the continuum. Such measures are fine for a limited class of questions and theories, for instance, "Are experts consistent over distal isomorphs of a problem?," and "Will novices use basic feedback from one task to deduce another task's answer?" Addressing such questions demonstrates such measures' descriptive power. However, it seems a leaping inference to assert that either physicists or novices are irrational regardless of the data obtained.

Although studies involving measures from the individual-centered extreme are still logically powerless to fully rule out the possibility of either general consistency or general rationality ([Ranney, 1994a]; see [Rouse and Morris, 1986], on fundamental limits), they provide better approximations to attempted rejections of a consistency hypothesis. This is especially true when consistency is considered with respect to individuals working within a short temporal window, in the same context , and on proximally isomorphic (or even identical) tasks. Of course, such individuals may still be "dancing to a drummer" that the experimenter has not yet heard (cf. [Collins, 1985], etc., on combining components in mental models), but rationality seems least evident when an individual seems to act most capriciously in a constant environment. (Of course, variability clearly may have some inherent evolutionary advantages e.g., in foraging or in foiling predators and competitors.)

8 Conclusions

As with other areas of study, describing the relative "theoryness" of naive conceptions of motion is a multifaceted, difficult, and fairly ill-formed task. We might attribute a naive (impetus or other) theory to an untrained subject if his/her data were well-predicted by a proposed model that we researchers can verbalize. But much of the data cited above (and elsewhere) indicate that no such model is yet satisfactory. Still, it may be that some people have (at least some aspects of) well-entrenched and consistent belief systems, but that features of researchers' models are just, so far, largely orthogonal to those of such subjects' "theories" [see Collins and Stevens, 1984]. Hence, researchers should strive to better develop models that either (a) entail principles that, from the start, guide the design of an empirical study and its contrasts, or (b) emerge from a sufficiently dense corpus of subjects' data through the adoption of both sophisticated and fine- grained analyses [Ranney, 1994a].

The individual-centered versus model-centered continuum represents a dimension upon which we may place consistency measures from either of these two sorts of ventures, regardless of the target domain (e.g., physics) involved. The model-centered extreme involves organizing principles that are imposed externally by the researcher (based on past results, hunches, or historical and scientific theories, etc.), while the individually-centered extreme highlights more idiographically fundamental notions of consistency. It would seem that the measures that center on individuals (e.g., temporal consistency and behavioral agreement) are primarily those that offer the clearest glimpse at context-free rationality.

References

[Caramazza et al, 1981]
Caramazza A, McCloskey M and Green B. Naive Beliefs in "Sophisticated" Subjects: Misconceptions about Trajectories of Objects. Cognition, Vol 9, pp. 117-123, 1981.

[Chi, 1992]
Chi MTH. Conceptual Change Within and Across Ontological Categories: Examples From Learning and Discovery in Science. In RN Giere (ed.), Cognitive Models of Science, University of Minnesota Press, pp 129-186, 1992.

[Chi et al, 1981]
Chi MTH, Feltovich PJ and Glaser R. Categorization and Representation of Physics Problems by Experts and Novices. Cognitive Science, 5, pp 121-152, 1981.

[Clement, 1983]
Clement JA. Conceptual Model Discussed by Galileo and Used Intuitively by Physics Students. In D Gentner and AL Stevens (eds), Mental Models, Erlbaum, pp 325-339, 1983.

[Collins, 1985]
Collins, A. Component Models of Physical Systems. In Proceedings of the Seventh Annual Conference of the Cognitive Science Society, Erlbaum, 1985.

[Collins and Gentner, 1987] Collins A an Gentner D. How People Construct Mental Models. In D Holland and N Quinn (eds), Cultural Models in Language and Thought Cambridge, pp 243-265, 1987.

[Collins and Stevens, 1984]
Collins A and Stevens AL. Mental Models of Complex Systems: Project Summary. Tech. Rep. No. 5788, Bolt Beranek and Newman Inc., 1984.

[Cooke and Breedin, 1994]
Cooke NJ and Breedin SD. Constructing Naive Theories of Motion On-the-fly. Memory and Cognition, Vol 22, pp 474-493, 1994.

[diSessa, 1982]
diSessa AA. Unlearning Aristotelian Physics: A Study of Knowledge-Based Learning Cognitive Science, Vol 6, pp. 37-75, 1982.

[Freyd et al, 1988]
Freyd JJ, Pantzer TM and Cheng JL. Representing Statics as Forces in Equilibrium. Journal of Experimental Psychology: General, 117, pp 395-407, 1988.

[Gutwill et al, 1996]
Gutwill J, Frederiksen JR and Ranney M. Seeking the Causal Connection in Electricity: Shifting Among Mechanistic Perspectives. International Journal of Science Education, Vol 18, pp 143-162, 1996.

[Halloun and Hestenes, 1985]
Halloun IA and Hestenes D. Common Sense Concepts About Motion. American Journal of Physics, VOl 53, pp 1056-1065, 1985.

[Hoadley et al, 1994]
Hoadley CM, Ranney M and Schank P. WanderECHO: A Connectionist Simulation of Limited Coherence. Proceedings of the Sixteenth Annual Conference of the Cognitive Science Soc. Erlbaum, pp. 421-426, 1994.

[Hojnacki, 1988]
Hojnacki SK. Consistency in Naive Physical Reasoning. Unpublished master's thesis, University of Pittsburgh, Learning Research and Development Center, 1988.

[Kaiser et al, 1986]
Kaiser MK, Jonides J and Alexander J. Intuitive Reasoning about Abstract and Familiar Physics Problems. Memory and Cognition, 14, pp 308-312, 1986.

[Larkin et al, 1980]
Larkin J, McDermott J, Simon DP and Simon HA. Expert and Novice Performance in Solving Physics Problems. Science, 208, pp 1335-1342, 1980.

[Maier, 1930]
Maier NRF. Reasoning in Humans: I. On Direction, Journal of Comparative Psychology, 10, pp 115-143, 1930.

[McCloskey et al, 1983]
McCloskey M, Washburn A and Felch L. Intuitive Physics: The Straight-Down Belief and its Interactions with Moving Objects. Journal of Experimental Psychology: Learning. Memory and Cognition, Vol 9, pp 636-649, 1983.

[Nersessian and Resnick, 1989]
Nersessian NJ and Resnick LB. Comparing Historical and Intuitive Explanations of Motion: Does Naive Physics Have a Structure? Proceedings of the Eleventh Annual Conference of the Cognitive Science Society, Erlbaum, pp. 412-417, 1989.

[Ranney, 1987]
Ranney M. Restructuring Conceptions of Motion in Physics-Naive Students. Report No. SE-053-227, Univ of Pittsburgh, Learning Research and Development Center (ERIC Document Reproduction Service No. ED 357 965; pp. 1-15), 1987.

[Ranney, 1987/1988]
Ranney MA. Changing Naive Conceptions of Motion

(Doctoral dissertation, Univ of Pittsburgh, Learning Research and Development Center, 1987), Dissertation Abstracts Intl., 49, p 1975B, 1988.

[Ranney, 1988]
Ranney, M. Contradictions and Reorganizations Among Naive Conceptions of Ballistics. Paper presented at the annual meeting of the Psychonomic Society, Chicago, November,1988.

[Ranney, 1989]
Ranney, M. Internally Represented Forces May Be Cognitively Penetrable: A Comment on Freyd, Pantzer, and Cheng (1988). Journal of Experimental Psychology: General, 118, pp 399-402, 1989.

[Ranney, 1994a]
Ranney M. Relative Consistency and Subjects' 'Theories' in Domains Such as Naive Physics: Common Research Difficulties Illustrated by Cooke and Breedin. Memory and Cognition, Vol 22, pp. 494-502, 1994.

[Ranney, 1994b]
Ranney, M. Assessing and Contrasting Formal and Informal/Experiential Understandings of Trajectories. In G Marks (ed), Proceedings of the International Symposium on Mathematics/Science Education and Technology pp 142-146, AACE, 1994.

[Ranney, in press]
Ranney M. Explorations in Explanatory Coherence. In E Bar-On, B Eylon and Z Schertz (eds), Designing Intelligent Learning Environments: From Cognitive Analysis to Computer Implementation, Ablex, in press.

[Ranney and Schank, 1995]
Ranney M and Schank P. Protocol Modeling, Textual Analysis, the Bifurcation/Bootstrapping Method, and Convince Me: Computer-Based Techniques for Studying Beliefs and Their Revision, Behavior Research Methods. Instruments and Computers, Vol 27, pp 239-243, 1995.

[Ranney et al, 1995]
Ranney M, Schank P and Diehl C. Competence and Performance in Critical Reasoning: Reducing the Gap by Using Convince Me. Psychology Teaching Review, Vol 4, pp 153-166, 1995.

[Ranney et al, 1994]
Ranney M, Schank P, Hoadley C and Neff J. "I Know One When I See One": How (Much) Do Hypotheses Differ From Evidence? Proc. of the Fifth Annual American Society for Information Science Workshop on Classification Research, pp 139-156, 1994.

[Ranney et al, 1993]
Ranney M, Schank P, Mosmann A and Montoya G. Dynamic Explanatory Coherence with Competing Beliefs: Locally Coherent Reasoning and a Proposed Treatment. In TW Chan (ed), Proc. of the Intl. Conf. on Computers in Education: Applications of Intelligent Computer Technologies, pp 101-106, 1993.

[Ranney and Thagard, 1988]
Ranney M and Thagard P. Explanatory Coherence and Belief Revision in Naive Physics, Proceedings of the Tenth Annual Conference of the Cognitive Science Society, Erlbaum, pp 426-432, 1988.

[Rouse and Morris, 1986]
Rouse WB and Morris NM. On Looking into the Black Box: Prospects and Limits in the Search for Mental Models. Psychological Bulletin, 100, pp 349-363, 1986.

[Schank and Ranney, 1992]
Schank P and Ranney M. 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 Soc, Erlbaum, pp 599-604, 1992.

[Schank and Ranney, 1993]
Schank P and Ranney M. Can Reasoning Be Taught? Educator, Vol 7(1), pp. 16-21 [Special issue on Cognitive Science and Education], 1993.

[Thagard, 1989]
Thagard P. Explanatory Coherence. Behavioral and Brain Sciences, Vol 12, pp 435-502, 1989.

[Thagard, 1992]
Thagard P. Conceptual Revolutions, Princeton University Press, 1992.

[White, 1993]
White BY. Thinkertools: Causal Models, Conceptual Change, and Science Education. Cognition and Instruction, Vol 10, pp 1-100, 1993.

Acknowledgements

This article benefitted from comments by (among others) Patricia Schank, Jayashree Ramadas, KSR Anjaneyulu, Michelle Million, William Prinzmetal, and Bernadette Wilkin, and from discussions with Lauren Resnick, Susan Hojnacki, Seth Chaiklin, Barbara White, John Frederiksen, George Montoya, and UC-Berkeley's Reasoning Group. Preparation of this article was supported by the National Academy of Education and the Spencer Foundation, and by faculty research grants from UC-Berkeley.

Michael Ranney is Associate Professor at the Univ of California, Berkeley, in the Graduate School of Education, its EMST/SESAME (Mathematics, Science, and Technology Education) Programs, and its Cognition and Development Area. He is also a faculty member in both Psychology and Cognitive Science, and is Director of $TEA(CH)_2EM$, a high school chemistry credentialing program. A 1991 Spencer Fellow (awarded via the US National Academy of Education), Ranney heads Berkeley's interdisciplinary Reasoning Research Group.