Conceptual coherence of non-Newtonian worldviews in Force Concept Inventory data

Terry F. Scott and Dániel Schumayer[*]

Department of Physics, University of Otago, Dunedin 9106, New Zealand
(Received 1 February 2017; published 16 May 2017)

The Force Concept Inventory is one of the most popular and most analyzed multiple-choice concept tests used to investigate students' understanding of Newtonian mechanics. The correct answers poll a set of underlying Newtonian concepts and the coherence of these underlying concepts has been found in the data. However, this inventory was constructed after several years of research into the common preconceptions held by students and using these preconceptions as distractors in the questions. Their sole purpose is to deflect non-Newtonian candidates away from the correct answer. Alternatively, one can argue that the responses could also be treated as polling these preconceptions. In this paper we shift the emphasis of the analysis away from the correlation structure of the correct answers and look at the latent traits underlying the incorrect responses. Our analysis models the data employing exploratory factor analysis, which uses regularities in the data to suggest the existence of underlying structures in the cognitive processing of the students. This analysis allows us to determine whether the data support the claim that there are alternate non-Newtonian worldviews on which students' incorrect responses are based. The existence of such worldviews, and their coherence, could explain the resilience of non-Newtonian preconceptions and would have significant implications to the design of instruction methods. We find that there are indeed coherent alternate conceptions of the world which can be categorized using the results of the research that led to the construction of the Force Concept Inventory.

I. INTRODUCTION

It is the task of teaching to guide students from the physical conceptions which they have developed themselves to the physical conceptions that form the bedrock of physics. A number of common physical preconceptions have been reported in the literature [1–3] in several different fields of physics. The preconceptions that relate to the basic ideas of dynamics and kinematics have been particularly well studied and this research forms the basis of the Force Concept Inventory (FCI).

Traditionally, the FCI is administered twice: usually at the very beginning of the mechanics course and again at the end of the course. Individual results can be interpreted separately, or can serve as the basis for evaluating how much conceptual knowledge students (individually and as a cohort) have gained during their studies [4]. One may then view this gain parameter as an index of the effectiveness of the instruction process. Martín-Blas *et al.* [5] have recently performed such pre- and post tests on freshmen at two engineering schools; one which admits students without any restriction on their secondary school achievements and

another that enrolls students only with relatively high passing scores. This experimental setup naturally leads to significant differences in scores, however, the research also demonstrated that these two significantly different groups still shared common preconceptions triggering the students to choose similar distractor items in the Force Concept Inventory. Their findings seem to support the idea of historical parallelism [6], i.e., students' intuitively built concepts based on their perception of motion which seem to repeat widely accepted ideas in the centuries before Newton.

In two previous papers [7,8] we have investigated the structure of correct responses to the FCI. These papers have studied the underlying conceptual structure of the Newtonian conception of force and the representation of that complex conception in the minds of students. All incorrect responses are assigned a zero and contribute no further information to the analysis. Inevitably this approach must discard a great deal of the data collected when students respond to the FCI. This means that all four incorrect responses to a particular question provide the same bit of information in the analysis, namely, that the student answered this question incorrectly. Thus all information regarding these incorrect responses in the data is homogenized.

This has always seemed to be a counterintuitive approach to us since it is the students who answer these questions incorrectly who most need our assistance as educators. It is thus the students who answer incorrectly that we should be collecting the most information about.

---

[*]daniel.schumayer@otago.ac.nz

We would like to know whether incorrect responses to the questions in the FCI are based on a coherent underlying worldview or worldviews and, if so, how strong these worldviews are. Furthermore, we would like to investigate the relationship between the incorrect worldview(s) and the correct worldview. It is conceivable that an incorrect worldview is not entirely incompatible with all or part of the correct Newtonian worldview.

The present paper reports on our analyses of incorrect responses to the questions in the FCI. We use these data to extract information about those common preconceptions that are represented in the incorrect items in each question in the FCI. In the following two sections we discuss exactly how we will retain this data.

Finally, we note that Morris et al. [9,10] have performed an Item Response Curve analysis with which they have examined distractors in the FCI. In their paper, they quantitatively assess the effectiveness of these distractors, i.e., whether a particular distractor item performs a useful function in a question. Moreover Morris et al. offer [9,10] a practical methodology, which allows instructors to assess not only the students' abilities but also how the distractor items are functioning in a particular cohort of students. Their work complements Wang and Bao's analysis of dichotomous coding of FCI response data [11] and simultaneously confirms Dedic's conjecture [12] that not all incorrect answers for a given item are equivalent. The research reported by the paper of Morris et al. uses FCI data collected from students, however, Maries and Singh [13] have investigated the use of the FCI not only for assessing students' Newtonian thinking, but also for studying the pedagogical content knowledge of teaching assistants. Based on their findings it would seem to be beneficial to incorporate FCI surveying into professional development courses, and perhaps our discussion of typical student preconceptions would be useful in this context as well.

As in the studies mentioned above, our work also relies on the correlation between the FCI items. However, here we analyze the internal structure of these correlations and draw conclusions about the nature of the preconceptions a student might have rather than assessing the connection between particular items and the ability of a respondent. We conjecture that students are not choosing incorrect items randomly, but because they are choosing answers on the basis of a group of misconceived ideas which form a coherent worldview.

II. DATA COLLECTION AND PREPARATION

The data used in this study are the same as was used in two previous papers [7,8] and further details can be found in these articles. Briefly, these data were collected over two years from a large, algebra-based, physics service course and were collected prior to the mechanics section of the course was delivered. The students in this course have a wide variety of academic backgrounds and abilities.

The FCI was presented to students via the Blackboard online course management system [14] and students were required to view the FCI as part of their internal assessment. Careful consideration was given to the quality of data collected in this way and obviously frivolous attempts at the survey were removed from the sample [7,8].

The FCI contains thirty multiple-choice questions, each of which has five possible options, which we—for brevity— call *items*. In the standard analysis of FCI data each correct response is recorded as 1 and each incorrect response is coded as 0.

In the past we have looked for coherence between correct responses that allowed us to investigate the degree of coherence of correct conceptions, and the internal structure of these correct conceptions. At the heart of those analyses sits a 30×30 correlation matrix whose entries give the correlations between each pair of questions. In order to explore the coherent conceptual structures indicated within the data we performed an exploratory factor analysis [7] using this representation of the data. That previous work raised several questions regarding the students' preconceptions in mechanics, e.g., are the incorrect ideas coherent?

In the current research we thus wish to investigate exactly these incorrect responses, and the relationship between incorrect and correct conceptions of force. In order to study the conceptual structure of incorrect responses we have changed the way we marked the FCI responses. We treated each item as a true or false statement. As an example, option 1B is incorrect, consequently, in the standard marking procedure if a student chose this item then question 1 would be given a mark of 0. In the new procedure this student would receive a 1 for *item* 1B, meaning that this student considers
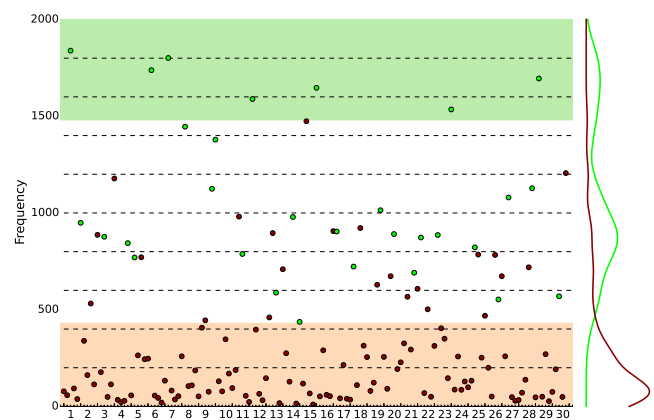

FIG. 1. The frequencies of each of the 150 items as observed in our corpus of 2109 answer sheets, similar to that published by Brewe et al. [15]. The light and dark circles represent the correct and incorrect items, respectively. The areas, where only one type of item is present (correct or incorrect), are shaded. On the right side, the smoothed density of the correct and incorrect answers are also depicted to demonstrate the clear separation of correct and incorrect answers. Note that these curves have different scales and are for illustrative purposes only.

this option to be true. At the same time items 1A, 1C, 1D, and 1E are marked to be 0. Their frequencies are given in Fig. 1. This means that the record associated with each student is a string of 150 1's and 0's and the correlation matrix is now a 150×150 matrix.

Some elements of this matrix, however, are spurious. Because of the way the FCI is usually delivered, students are only able to select one item per question. If a student selects option 1A, then they cannot select items 1B, 1C, 1D, or 1E. Thus, there is an automatic correlation of -1 between the selected and all nonselected items of the same question. These artificial anticorrelations do not indicate the existence of an underlying concept which produces these data, but rather reflect the way that the survey is delivered. Such strong anticorrelations would certainly invalidate the factor analysis and hide any genuine factors which do reflect latent traits underpinning student response.

For this reason all correlations between items within a question are manually set to zero (see Fig. 2). However, the linear dependence between each block of five items leads to a singular correlation matrix, and this causes the standard numerical methods of factor analysis to fail. We discuss this technical problem, and our solution to it, in the next section.

III. CORRELATION MATRIX

Before addressing the problem of singularity of correlation matrices we briefly summarize the idea behind factor analysis. We demonstrate the decisive role the correlation matrix plays in this method.

Exploratory factor analysis aims to explain the correlation between the observed variables under the assumption that a linear combination of a few, as yet unknown, unobserved (latent) factors, f, is the true origin of the observed variability in the data. Suppose that $Y = \{Y_1, Y_2, ..., Y_n\}^T$ is a vector of random variables that represents the observed data. The model equation determining the factors is then

$$Y - \mu = \Lambda f + \epsilon.$$

Here μ is a vector, such that $\mu_k = \mathbb{E}[Y_k]$, i.e., μ_k is the expectation value of the kth random variable. The matrix Λ contains the "loadings" of the factors on each variable in Y, while ϵ is a random "noise" vector. By explicitly subtracting the expectation values μ, we have ensured that $\mathbb{E}[f] = \mathbb{E}[\epsilon] = 0$. It is also assumed that the factors are uncorrelated with each other and with the random error. In formulas $\mathbf{cov}(f,f) = I$ and $\mathbf{cov}(f,\epsilon) = 0$, while each component of the error term may have a specific variance, $\mathbf{cov}(\epsilon,\epsilon) = \Psi$, where Ψ is a diagonal matrix. With these quite natural assumptions, the covariance matrix, Σ, of observed data is given by

$$\Sigma \stackrel{\text{def}}{=} \mathbf{cov}(Y,Y) = \mathbb{E}\{(Y - \mathbb{E}[Y])(Y - \mathbb{E}[Y])^T\}$$
$$= \Lambda\Lambda^T + \Psi. \tag{1}$$

This result is revealing: since the errors are uncorrelated, i.e., Ψ is a diagonal matrix, all the off-diagonal covariances, Σ_{ij}, $(i \neq j)$ should be described purely by the loadings λ_{ij}, while the covariances in the diagonal Σ_{ii}, are attributed both to the factors, $\sum_j \lambda_{ij}^2$, and to the error ψ_i.

If Ψ was known, one could directly determine Λ by rearranging Eq. (1) to obtain

$$\Sigma - \Psi = \Lambda\Lambda^T. \tag{2}$$

Solving Eq. (2) for Λ is equivalent to an eigenvalue-eigenvector problem [16]. Indeed, if Q and D denote the matrices of the eigenvectors and eigenvalues of $\Sigma - \Psi$, then $\Lambda = QD^{1/2}$. In practice, however, the error covariances Ψ are not known, rather they are estimated from the observation.

In our study two technical difficulties have arisen due to the marking scheme chosen and to the categorical nature of variables.

A. Removal of singularity

When students choose from the five items offered in a question, they consider four incorrect items which may poll different preconceptions. We wish to investigate these preconceptions, therefore we split the five possible items into five dichotomous variables. On the one hand this structure allows us to measure the correlation between any two items, while on the other hand this structure inherently carries a deterministic linear relationship between the five items, i.e., we know that only one of them is chosen and four of them are not. If, for example, 1A—1E denote the five dichotomous variables (coded with 0 and 1) corresponding to the five items of question 1, then the linear relationship is $1A + 1B + 1C + 1D + 1E = 1$. It is easy to show that similar linear relationships would emerge, with a different constant on the right-hand side, from any such coding scheme. Such a linear relationship inevitably manifests itself in a singular correlation matrix (see the Appendix at the end of this paper for a detailed
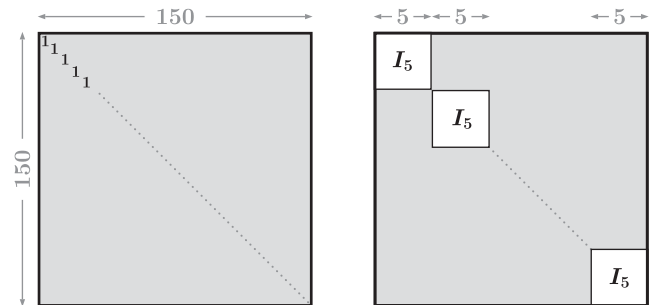
FIG. 2. We have manually adjusted the 150×150 correlation matrix by setting all 5×5 block diagonals to be equal to the 5×5 identity matrix, I_5. This adjustment nullifies the artificial anticorrelation between the five items corresponding to the same question.

explanation), with which the standard factor analysis cannot be carried out. Consequently, the correlation matrix between all 150 items must be adjusted to nullify these artificial anticorrelations.

We have suppressed this anticorrelation of items corresponding to the same question by manually setting their correlations to zero. In other words, the diagonal blocks of size 5 are set to be the identity matrix, I_5 (see Fig. 2). This is equivalent to saying that these items are independent of each other. While admittedly *ad hoc*, this approach does seem quite natural. While this procedure nullifies the artificial anticorrelation between certain items, one side effect is that it may also introduce small negative eigenvalues, thus the correlation matrix may not be a positive definite matrix (and often is not). However, these values can be easily eliminated as we discuss in the next section.

B. Regularization

The brief, mathematical introduction of the core ideas of factor analysis given above indicates that the covariance matrix Σ plays a decisive role in the statistical method. We will now discuss a potential technical difficulty related to the form of this matrix, and the solutions we have employed to overcome this issue.

In educational research we often collect data using multiple-choice questionnaires; we ask students to choose the correct answer from several offered options. The data obtained in this way is naturally categorical, it is dichotomous if the answers are marked correct or incorrect but it may not be. The commonly used Pearson's correlation function is appropriate for continuous data and should not be used for categorical data. In our analysis we have employed either tetrachoric or polychoric correlations as these functions are better suited for the analysis of categorical data, especially if the variables in question can be regarded as sampling an underlying continuous (normally distributed) latent trait.

Polychoric and tetrachoric correlation functions are calculated from marginals distributions (see Ref. [17] for more details), and sometimes produce correlation matrices which are not positive definite [18,19]. Data with inadequate sample size and/or many possible options for questions are especially prone to the inference of "incorrect" correlation matrices. Alternative methods for performing factor analysis that do not employ correlation matrices do exist (e.g., structural equation modeling or maximum likelihood methods [20,21]). However, these methods generally require the numerical evaluation of multiple integrals and thus these techniques are often prohibitively computationally intensive. In this study we have taken a different approach. We have replaced ill-conditioned correlation matrices with new matrices. These, in some sense, are the "closest" well-formed matrices to the sample correlation matrix, which is, after all, itself an estimate of the actual correlation matrix. The determination of the

closest matrix is not a fully solved problem, however, specific cases have been thoroughly analyzed in the mathematical literature [22].

The eigendecomposition of a square symmetric matrix, S, guarantees that it can be written as $S = QDQ^T$, where the diagonal of D contains the possibly complex eigenvalues d_i of S while the columns of Q are the corresponding eigenvectors, q_i ($i = 1, 2, \ldots, n$). We may write this decomposition in the form

$$S = QDQ^T = \sum_{i=1}^{n} d_i q_i q_i^T \approx \sum_{0 < d_t \leq d_i} d_i q_i q_i^T = \tilde{S}.$$

In the last term, we have truncated the finite sum based on the magnitude of the eigenvalues. It seems reasonable to assume that we can capture the most important effects by keeping the terms with eigenvalues larger than an arbitrarily chosen positive threshold value, $0 < d_t$. Furthermore, such truncation of the eigendecomposition of a symmetric, not positive semidefinite S naturally leads to a symmetric, but positive semidefinite $\tilde{S}$ matrix. In this sense, if S denotes the ill-conditioned correlation matrix calculated from the categorical data, then $\tilde{S}$ can be considered as the regularized correlation matrix which one may use in a factor analysis. It seems reasonable to suppose that $\tilde{S}$ is, in some sense, the closest positive semidefinite matrix to the actual correlation matrix of the sample.

We have followed this approach in the analysis presented below and we used a regularized 150×150 correlation matrix. We have checked that the choice of threshold (in the range from 10^{-6} to 10^{-2}) value does not significantly alter the item-factor assignment.

IV. FACTOR ANALYSIS

The main purpose of this paper is to present a factor analysis of the full data set, including incorrect responses. Equally importantly we intend to identify any relationships between the Newtonian and alternate worldviews, if they exist. To this end we look at the loadings of each item into a particular factor, but we also need to employ some simple clustering techniques. The need for these techniques becomes clear when we discuss the results of the factor analysis.

The great value of factor analysis is that it provides a model of the conceptual structure underlying students' responses to the FCI. This is apparent from the fact that the factor loadings represent the correlation between a particular variable and the underlying factor, rather than a correlation between the variables themselves. Furthermore, the factor loadings indicate the strength of connection between the underlying worldview held by the student and their subsequent response to a given question in the FCI.

However, before we delve into the factor analysis of all items, we remark on question 1 and its use in interpreting

the result of factor analysis. Question 1 asks the student to consider a situation in which two balls of equal size but different masses are dropped from the same height at the same time. The student is then asked about the time it takes for each ball to reach the ground. The question can be seen as asking the student whether or not the balls hit the ground at the same time. This scenario is a staple of early education in science, both in the various popular science documentaries which are now widely available and in early school education. Indeed this scenario is so well known that a correct answer to this question provides little information about the nature of the force concept held by the student. The answer to this question could be seen as indicating the level of general scientific knowledge of the student. For this reason we see this question as being of additional value to researchers; the answer to this question automatically segments the respondents into two cohorts. The presence of incorrect responses to this question may be taken as indicating a lower level of general scientific background compared to students who answer this question correctly. Some support for this view may be seen in the fact that this question is answered correctly by 87% of the students polled in the current study. We have calculated Welch's t test [23] for comparing the overall scores of students answering question 1 correctly (mean = 15.93, sd = 5.85, $n = 1839$) with that of students choosing an incorrect response (mean = 10.72, sd = 4.74, $n = 268$). There is a significant difference between these groups; $t = 16.28$, df = 395.84, $p < 2.2 \times 10^{-16}$. While we do not rely on this interpretation of question 1 in the analysis that follows, it is worth noting. We will mention the appearance of question 1 preconceptions in some of the factors which we analyze.

In all factor analyses it is important to decide how many factors to keep. One popular, but slightly *ad hoc* approach is the scree plot technique [24]. The scree plot in Fig. 3 shows the eigenvalues as determined from factor analysis and principal component analysis. Although we do not present the outcome of this latter technique, we show its scree plot together with that of the factor analysis in order to support our findings. Both curves suggest that we should set the number of factors in the range of 5–10 factors. We emphasize again that the selection of the number of factors is not a settled issue in statistics and the process involves some degree of arbitrariness. We use the scree plot in Fig. 3 to suggest an appropriate number of factors to retain, however we investigate this choice further below.

We first summarize the use of the scree plot. In Fig. 3 the straight lines correspond to the eigenvalues generated for uncorrelated random normal variables. These lines thus represent a baseline set of eigenvalues in that, if the eigenvalues of a factor generated by our data falls on or below these lines, then that factor is indistinguishable from random, normal noise. The curves generated by our data cross the straight lines around $n \approx 15$ and $n \approx 80$. Thus,
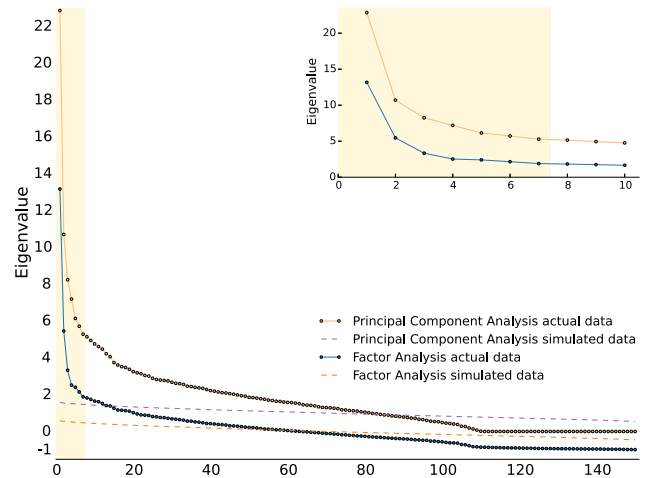

FIG. 3. The result of a parallel analysis based on the tetrachoric correlation matrix is depicted. The dark and light circles represent the eigenvalues as returned by the exploratory factor analysis and principal component analysis, respectively. Each dashed line shows a linear fit of the eigenvalues of correlation matrices generated from random uncorrelated, standardized normal variables. This suggests models with relatively large numbers of factors, in excess of 20 (crossings of the corresponding actual and simulated data). However, choosing the number of factors also depends on the interpretability of the solution. Note that the "kink" in the curves at around ~110 is an artifact of our regularization procedure and immaterial to our analysis.

eigenvalues falling below these straight lines could be attributed to the particular random sample we have collected and not to any inherent, but unknown structure in the data. As a consequence, one should perhaps neglect these eigenvalues. Nevertheless, such simulation does not mean that we should blindly accept all eigenvalues above these lines.

Another commonly used approach employing the scree plot is to consider the point at which the scree-plot curve exhibits a sharp change in its slope. This point is often taken as indicating the optimal number of components required. We have followed this tradition, both here and in our previous publication [7]. The graph shows a number of sharp changes in slope, however, the relevant point in the curve is the point at which the initial steep slope flattens out. This point occurs in our graph somewhere after the fourth eigenvalue and the curve has definitely flattened out after about the tenth eigenvalue. On the basis of this graph we should choose between five and ten factors, clearly the scree plot in this case is not particularly definitive. However, the scree plot is still useful as an initial guide.

In order to examine the robustness of our model we have examined alternative models using 3 to 30 factors and have found that (a) selection of less than six factors *does* result in significant changes in the assignment of items to factors, and (b) that selecting more than six factors *does not* significantly change the first six factors. The first of these findings is unsurprising, since one has to put the same

number of items into a smaller number of "buckets." However, the second observation is perhaps a little unusual. In the second case factors after the sixth are very small, i.e., contain small numbers of items, and the structure of the first six factors does not change much. For this reason we have performed our analysis with six factors.

The most interesting results that we find are in factors 1 and 2. These factors contain all of the correct items along with a similar number of incorrect items. Most interesting is the sign of the factor loadings in these two factors.

A. Factor 1

Factor 1 contains 21 correct items and 25 incorrect items. These are given in Table I together with their loadings. More than two-thirds of the correct answers appear in this factor and we interpret them as essentially representing the Newtonian worldview.

Of more interest is the sign of the factor loadings. All of the correct items have negative factor loadings and are thus negatively correlated with the underlying factor. This factor should therefore be interpreted as anticorrelated with Newtonian thinking, and thus this factor represents an alternate worldview to the Newtonian world view. However, there is another possibility which should be considered before going further with interpretation.

In our earlier paper [7] we have shown that the correct answers to the FCI represent an underlying factor which we have called Newtonian-ness. That research shows that the correct answers are strongly correlated with this underlying factor and with each other. Given the reliance of factor analysis on the correlation matrix it is possible that factor 1 is actually just this Newtonian-ness factor and the incorrect items which are attached to it are *individually* anticorrelated with one or some of the correct answers. In other words, it is possible that the incorrect answers do not constitute an alternate coherent worldview but that each of them

independently anticorrelates with the Newtonian world-view. Following Sabella and Redish's model concerning knowledge structure [25], one could visualize the situation as a cloud of strongly interconnected concepts (correct items) and on the periphery lie some preconceptions that are linked to one or only a few central correct concepts. Therefore, if a node of a correct concept is activated there is still some likelihood of triggering a node corresponding to an incorrect concept.

In order to discount this possibility we need to show that the incorrect items are correlated with each other as well as being anticorrelated with the correct items. The simplest and most easily interpretable way to do this is using a clustering technique. In Fig. 4 we show the cluster diagram of items appearing in factor 1. The horizontal axis in this diagram represents dissimilarity as traditionally used in cluster analyses. The further a join is from the right-hand side of the diagram, the more dissimilar the items joined are. Similarity and dissimilarity in this clustering analysis are determined by the size of the correlation between those items.

From this diagram it is clear that the correct items are all more closely correlated with each other than they are with the incorrect items, and that the incorrect items are more closely correlated with each other than with the correct items. There is a single tree containing all the incorrect items and none of the correct items, and a single tree containing all of the correct items and none of the incorrect items. Thus, we may safely conclude that the correct items in this factor represent the Newtonian worldview and the incorrect items represent a seemingly coherent conceptual worldview, which is anticorrelated with that of the Newtonian one.

TABLE I. Items in factor 1 with higher absolute loading than 0.290 are listed. The items appear in the same order as in the FCI test. Correct items are indicated with an asterisk.

| Item	Loading	Item	Loading	Item	Loading	Item	Loading
*3C	−0.317	*11D	−0.597	18D	0.346	24E	0.315
4A	0.412	*12B	−0.350	18E	0.370	25C	−0.525
*4E	−0.371	12C	0.378	19D	0.328	25D	0.314
*5B	−0.556	13B	0.384	*19E	−0.391	*25E	0.336
6A	0.361	13C	0.349	20C	0.293	27A	0.328
7A	0.304	*13D	−0.708	*20D	−0.453	*27C	−0.348
8A	0.320	14A	0.346	*21E	−0.360	28D	0.432
9B	0.333	*14D	−0.422	*22B	−0.522	*28E	−0.407
*9E	−0.367	16C	0.384	22D	0.342	*30C	−0.617
*10A	−0.390	17A	0.411	*23B	−0.485	30E	0.484
10D	0.428	*17B	−0.494	23D	0.308		
11C	0.522	*18B	−0.585	*24A	−0.329		
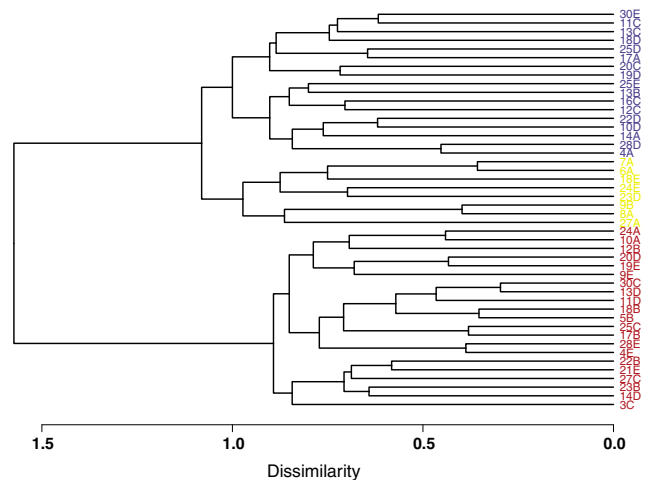
FIG. 4. The tree structure of factor 1 as returned by the hierarchical cluster analysis. The dissimilarity of items i and j is defined as $r_{ij} = 1 - c_{ij}$, where c_{ij} is their tetrachoric correlation. The colors are used to guide the eye and distinguish between groups of correct (red) and incorrect items (group A is blue and group B is yellow).

In order to interpret exactly what this anti-Newtonian worldview represents we consider the items which are positively loaded onto this factor, in particular those items with the largest positive loadings. The original FCI paper [3] contained a list of the well-known preconceptions relating to the concept of force and the items in the FCI which were related to these preconceptions. We use these labels in our analysis of the anti-Newtonian worldview identified in this factor.

The incorrect items in factor 1 are grouped together in the cluster diagram. However, this diagram also shows that the incorrect items are divided within this group into two subgroups. We call them group A and group B. The items in each of these groups and their loadings into factor one are shown in Table II ordered by decreasing factor loadings.

We now interpret these clusters using the factor loading as a guide. The most important item for the interpretation of each cluster is the one with the highest loading as this item is the most strongly correlated with the underlying factor. We now begin with Group A as these items have the largest factor loadings.

The largest factor loading is associated with item 11C indicating an "impetus" conception, i.e., the belief that the cause of an object's motion is some inherent property of the object itself, rather than being an interaction between objects [26–28]. In this item, the impetus conception is indicated by the belief that the force causing the motion must be in the direction of the motion. Items 30E, 10D, 13B, 12C, and 13C also support an impetus based conception of force. These are all in the top end of factor loadings and thus there is strong evidence that this cluster represents an impetus conception of force. Furthermore, most of the other items are also related to the impetus preconception. This is clearly a very powerful and natural idea to novices.

Items 28D, 4A, and 16C, which are also in the top end of the factor loadings in this cluster do not, however, directly relate to the impetus preconception. These are all items

TABLE II. Factor 1 comprises two strikingly different groups: A and B. The membership of items in these groups are listed together with their loadings on to factor 1. The items are ordered by decreasing loadings.

Group A				Group B	
Item	Loading	Item	Loading	Item	Loading
11C	0.522	13C	0.349	18E	0.370
30E	0.484	18D	0.346	6A	0.361
28D	0.432	14A	0.346	9B	0.333
10D	0.428	22D	0.342	27A	0.328
4A	0.412	25E	0.336	8A	0.320
17A	0.411	19D	0.328	23D	0.308
16C	0.384	25D	0.314	7A	0.304
13B	0.384	20C	0.293	24E	0.315
12C	0.378				

which indicate preconceptions relating to Newton's third law. Students with an impetus based conception of force are not in a position to answer these questions in the Newtonian manner. Newton's third law is only intelligible if forces are not properties of objects, or in any way contained in an object. The third law relies on the fact that forces are interactions between objects rather than properties of single isolated objects. A student who has an impetus conception of force therefore requires some extension of this concept in order to answer a question involving the interaction of two (or more) objects. The student needs to form some sort of *ad hoc* rule that relates the forces produced and experienced when the situation involves two objects that interact. In other words they need to come up with an "impetus third law" on the fly (as it were). The items in this cluster—we conjecture—represent the attempts by students to understand the interaction of objects within an impetus conception of force. Two of the three items in this cluster which are related to Newton's third law preconceptions represent the view that the object with the greater mass exerts the greater force. The other item represents the view that the most active agent exerts the greatest force.

The belief that the greater mass produces the greatest force is the most natural idea since everyday experience teaches us that the more massive an object is the more effective it appears to be in changing the motion of smaller objects. The active force idea is in some ways more sophisticated as it is less obviously connected to everyday experience and introduces a novel idea, namely, that of an active agent.

The second group of items, Group B, are also largely related to the impetus idea. In this group though there are several items related to impetus dissipation and to circular impetus. These items further reinforce our finding that this factor primarily indicates an impetus based worldview.

It is important to reiterate that this factor includes the correct Newtonian answers but with negative factor loadings. Thus it appears that the strongest hindrance to a Newtonian worldview is the impetus worldview and that this worldview strongly anticorrelates with the Newtonian worldview [25,29,30].

B. Factor 2

Factor 2 contains nine correct and one incorrect item which are negatively correlated with the underlying factor, while the remaining thirty-three incorrect items are positively correlated with the underlying factor, see Table III. Note that the correct answer to question 10, item 10A, appears in both factor 1 and factor 2.

Here, the correct items all relate to Newton's first law in the absence of external forces. In other words these items capture force balance occurring in rectilinear motion. The correct answer to question 16 is also in this list, but as has been noted in earlier studies [7,31,32], this is to be expected as students often incorrectly employ Newton's first law to
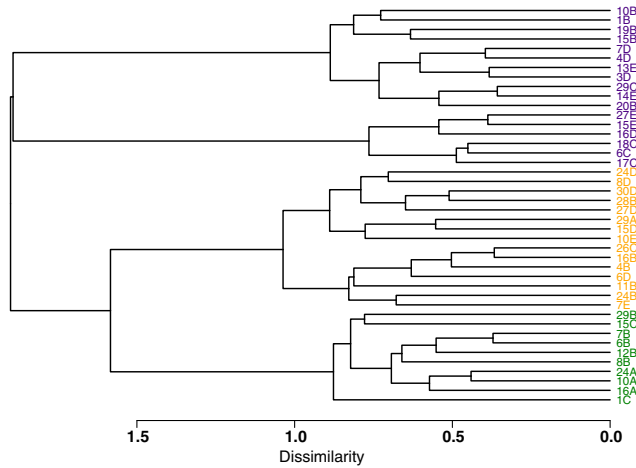
FIG. 5. The tree structure of factor 2 is shown analogously to Fig. 4, however, the colors do not represent any connection to the groups of factor 1. The two groups C and D identified within factor 2 are coded with colors orange and indigo, respectively. Correct items are green.

arrive at the correct answer to this question, even though the question is designed to poll their understanding of Newton's third law.

The incorrect items which load positively onto this factor appear to sample a wide range of preconceptions. We have again checked—using hierarchical cluster analysis—whether or not these incorrect items are correlated with each other rather than just individually anticorrelated with the correct ones.

The outcome of cluster analysis is shown in Fig. 5 and it reveals an interesting and complex structure. The incorrect items are again clustered together, corroborating the claim that they do indeed represent another coherent non-Newtonian worldview. However, in this case the correlation structure is considerably richer. The incorrect items fall into two distinct groups (named C and D) like the incorrect

TABLE III. All items in factor 2 with higher absolute loading than 0.290. Their order is the same as in the FCI test. Correct items are indicated with an asterisk.

Item	Loading	Item	Loading	Item	Loading	Item	Loading
1B	0.335	*8B	−0.379	15D	0.445	24D	0.3134
*1C	−0.293	8D	0.386	15E	0.300	26C	0.4444
3D	0.412	*10A	−0.508	*16A	−0.462	27D	0.3004
4B	0.498	10B	0.306	16B	0.377	27E	0.4594
4D	0.428	10E	0.324	16D	0.506	28B	0.5534
*6B	−0.429	11B	0.310	17C	0.342	29A	0.4014
6C	0.355	*12B	−0.365	18C	0.498	*29B	−0.3474
6D	0.345	13E	0.340	19B	0.304	29C	0.4064
*7B	−0.446	14E	0.456	20B	0.472	30D	0.4704
7D	0.550	15B	0.425	*24A	−0.461		
7E	0.328	15C	−0.449	24B	0.441		

TABLE IV. Factor 2 comprises two groups, called C and D. Membership of items in these groups together with their loadings onto factor 2 are given. Items are ordered by decreasing loadings.

Group C				Group D			
Item	Loading	Item	Loading	Item	Loading	Item	Loading
28B	0.553	16B	0.377	7D	0.550	29C	0.406
4B	0.498	6D	0.345	16D	0.506	6C	0.355
30D	0.470	7E	0.328	18C	0.498	17C	0.342
15D	0.445	10E	0.324	20B	0.472	13E	0.340
26C	0.444	24D	0.313	27E	0.459	1B	0.335
24B	0.441	11B	0.310	14E	0.456	10B	0.306
29A	0.402	27D	0.300	4D	0.428	19B	0.304
8D	0.386			15B	0.425	15E	0.300
				3D	0.412		

items in factor 1; see Table IV. In this case, one group, group C, is more closely associated with the correct answers than with the other group of incorrect items, group D.

We begin with the interpretation of group C. The item with the largest loading is 28B which is associated with the notion that only active agents exert forces, and a passive agent changes its motion in response to the force exerted by the active agent. Naturally then, a student who holds this active agent conception of force struggles in trying to determine the trajectory of an object that is not acted on by an active agent. A number of the top items (by their loading) are related to the impetus conception of force, in particular, to the way impetus is imparted to an object and how the impetus varies. Since impetus is a property of an object, the change in impetus during motion changes the way that the object moves. Hence the impetus conception is very likely to effect the way students answer questions about the trajectory of an object in the absence of a net (Newtonian) force. In fact, this is precisely the scenario in which the impetus conception and the Newtonian conception are in direct conflict. In the Newtonian worldview, an object's trajectory does not change in the absence of a force. Since force is produced by interaction with another object, an isolated object cannot change its trajectory. Contrarily, in the impetus worldview an object carries the cause of change in trajectory along with itself and thus an isolated object can change its own trajectory.

Thus it is not surprising that preconceptions, relating the way that the impetus of an object changes as it moves, anticorrelate with Newtonian items relating to the first law.

The second group of preconceptions in this factor are those in group D. These items are again strongly associated with the impetus view of the world with the bigger emphasis being on the active-passive agent notion. The largest factor loading is associated with 16D supporting our statement: students picking this item think that forces are exerted by active agents on passive agents or objects.

C. Factors 3 and 4

Before addressing the remaining four factors we note that the eigenvalues of these factors are significantly smaller than the eigenvalues of the first two factors; see Fig. 3. We thus do not expect that these last four factors are as important as the first two. However, the eigenvalues of these factors are well above the noise level and it is clear that the associated factors are robust. We thus attempt to provide an interpretation of these factors but with the caveat that they probably play a lesser role in determining student responses to the FCI.

We begin our interpretation of the remaining factors by noting that all contain incorrect items from question 1, and, as discussed above, this may indicate that these factors appear in students who have a significant deficit in their general scientific background. Table V shows the members of factors 3–6.

The most striking fact that connects the items in factor 3 and 4 is that these items represent the same option choice in several questions. Thirteen items out of fifteen appearing in factor 4 are items labelled in the FCI test with the letter C. Similarly, all of the items in factor 3 involve choosing option A (except item 1C whose loading is so low that it could be excluded from this factor). When students are completely perplexed by a question they tend to select option C (factor 3) or option A (factor 4). We struggle with reading these factors in any other way. Thus, we would hypothesize that these factors indicate some kind of choice bias in students [33,34]. In other words we have captured an "urban myth" coping strategy of students which is not specific to physics, rather it is a hallmark of the multiple-choice assessment type. We also point out that these two factors are quite robust, they appear whether we use an orthogonal or an oblique rotation, and they also appear when we vary the number of factors chosen for the analysis. One could test our hypothesis of choice bias by rearranging only the incorrect items in questions corresponding to factors 3 and 4.

Consequently, these factors do not provide us with any useful information about the structure of force preconceptions held by students, and so we proceed with the analysis of factors 5 and 6.

D. Factors 5 and 6

The items in these factors along with their factor loadings are shown in Table V. These two factors are quite robust to changes in the number of factors used in the factor model. They appear in 5, 6, and 7 factor models (with the exception that naturally factor 6 does not appear in the 5 factor model). This is worth noting as these two factors exhibit an unusual effect not seen in the other factors: conceptual conflict between preconceptions.

Factor 5 contains thirteen items, all of which are incorrect options. One item, 15E, is negatively loaded onto this factor with a loading of −0.907. This is a very high loading indicating that the concept underlying this factor strongly disposes the student against selecting this option. The other items in this factor relate to a reasonably diverse collection of preconceptions, ranging from kinematic preconceptions (14A) to ones relating air pressure and gravity (29C). The item with the largest (positive) loading is 27B, which encodes the belief that an object slows down and stops due to its mass. There does not seem to be a clear interpretation that explains the presence of all of these items in the same factor and the loadings of these items onto this factor are also reasonably weak. We suggest that this indicates that interpretation of this factor should be based largely on the rejection of item 15E rather than the presence of the positively loaded items.

Item 15E is an option in a question relating to Newton's third law that indicates that the student who selects this options believes that "Obstacles exert no force" (to use the language of Halloun and Hestenes [35]). Option 15E is a particularly simple replacement for the third law, effectively avoiding the use of a force concept entirely. This item could therefore be seen as indicating a viewpoint in which the student does not see the need for an explanation. This item would therefore be in direct conceptual conflict with the attempt to construct a coherent explanation of motion. Thus, this factor may be seen as representing the rejection by the students of the viewpoint that no explanation of motion is necessary.

If this interpretation is correct, the positively loaded preconceptions may be described as a collection of ideas that represent an initial attempt at just such an explanation of motion. We would very tentatively suggest that this factor may indicate the process of theory formation among students with little or no scientific background (as indicated

TABLE V. Items with their loadings on the remaining factors are given. Their order follows that of the FCI test. No correct answer is assigned to these factors.

Factor 3		Factor 4		Factor 5		Factor 6	
Item	Loading	Item	Loading	Item	Loading	Item	Loading
1E	−0.428	1A	0.502	1D	0.348	1E	0.328
4C	0.697	1C	−0.291	6D	0.307	3E	0.503
6C	0.474	5A	0.479	8E	0.381	5E	0.439
7C	0.301	7A	0.375	9D	0.312	10C	0.401
8C	0.296	9A	0.553	14A	0.385	12D	0.500
10C	0.347	11A	0.442	15E	−0.907	14E	−0.850
14C	0.329	12A	0.443	17E	0.294	24C	0.462
16B	0.295	13A	0.361	19A	0.310	29E	0.318
17C	0.629	18A	0.636	21D	0.331		
18C	0.344	19A	0.290	22E	0.343		
19C	0.386	25A	0.409	26B	0.354		
22C	0.537	28A	0.603	27B	0.446		
26C	0.337	29A	0.521	29C	0.365		
28C	0.318	30A	0.414				
29C	0.512						

by the presence of an incorrect question 1 item) who are perhaps even confronting the physical situations presented in FCI questions for the first time.

Factor 6 presents us with a very significant conceptual conflict between preconceptions. In this factor there are 8 items, most of these items relate to impetus and impetus dissipation. However, item 14E has a very high and negative factor loading, −0.850. Furthermore this item appears to be nearly identical to item 12D which also appears in this factor but with a large positive loading (0.500). The corresponding item is reproduced as the top illustration in Fig. 6, however it is labeled as option C due to a slight modification between the 1992 and 1995 versions of the FCI. Thus, there appears to be a significant conceptual conflict between these two items even though, to the Newtonian viewpoint, these items appear to be very nearly identical.

In question 12 students are asked to select the correct trajectory of a cannon ball, which has been fired from the top of a cliff. The question is graphical in the sense that the student is presented with a diagram and a set of possible trajectories labeled from A to E. Item 12D (in the 1995 revised FCI) identifies a trajectory in which the cannon ball travels horizontally for some distance and then curves down abruptly until it is traveling vertically downward. This is considered, by the authors of the FCI, to indicate that the student believes that the horizontal impetus of the
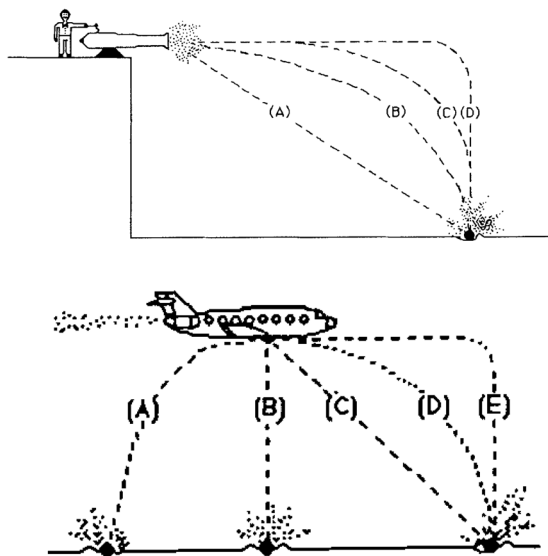

FIG. 6.   The visual cues appearing for questions 16 and 23 in the original 1992 FCI, or, alternatively, for questions 12 and 14 in the 1995 revised FCI. Reproduced from Ref. [3], with the permission of the American Association of Physics Teachers. We draw the readers' attention to the minute difference between questions 16 (1992) and question 12 (1995). In the former, only four choices were offered to students, while in the latest version five options are given. This deviation does not alter our analysis and here we used the 1995 version only.

cannon ball has been imparted by the cannon and slowly dissipates as the cannonball moves. Once the horizontal impetus has dissipated to a sufficiently small value, the vertical impetus imparted by gravity begins to have an effect. Once the horizontal impetus has completely dissipated the vertical impetus is all that remains and the trajectory is vertical.

In question 14 students are asked to select the correct trajectory of an object, in this case a bowling ball, dropped from the underside of an airliner. The question again provides a graphical representation of the trajectory options to be selected. The trajectories in this question are essentially identical to those of in question 12. Item 14E represents a trajectory in which the bowling ball travels horizontally until it turns sharply and then falls vertically to the ground. Item 12D represents a very similar trajectory in which the cannon ball travels horizontally, turns slightly less sharply and then falls vertically to the ground.

The differences between items 12D and 14E appear to be very slight when viewed from the Newtonian perspective. These students may be distinguishing between objects that are fired and objects that are carried and then dropped. In the first case, the object has received an amount of horizontal impetus from the cannon during the "hit." However, the bowling ball in the airliner has not been hit, it was simply dropped. Therefore the bowling ball has not received horizontal impetus. Thus students may perceive option 14E as being in significant conflict with their impetus based theory. This factor could also represent the process of theory formation (again due to the presence of an incorrect question 1 item), but in this factor the process is more advanced as the students are perhaps considering the process by which impetus is imparted.

We note that it is also possible that students see some significance in the sharpness of the turn from horizontal to vertical trajectories. In general, this would reflect some subtle difference in interaction between horizontal and vertical motive properties. However, it seems more likely that the division is between an object which is impelled in the horizontal direction and an object which is just allowed to drop.

We should again emphasize that our interpretation of these factors is tentative and that a more focused research program is required to confirm this hypothesis.

V. CONCLUSION

In this paper we have considered the conceptual coherence of students' responses to FCI questions. We have retained preconception information contained in the FCI data and have observed conceptual coherence in the preconception data. This indicates that students hold coherent non-Newtonian worldviews. We have also observed that these worldviews are in direct conflict with the Newtonian conception of mechanics. We have noted that the primary alternative worldview is based on the idea

of impetus. The conflict between these worldviews and the coherence of the impetus view should be taken into account when designing courses of instruction in this topic. It is worth emphasizing that the current study cannot comment on the coexistence or on the competition of these worldviews within the same student, since neither our data nor its analysis is able to capture such subtle effects. Thus we could only hypothesize that such competition occurs within a student, since this seems to be the natural progression in concept learning. A detailed interview study might be able to capture this effect.

We have also seen interesting structure in some of the factors associated only with incorrect items. We have interpreted this structure as representing the process of theory formation among students with very little scientific background. This last suggestion should be taken as tentative and indicative of the need for further research.

ACKNOWLEDGMENTS

APPENDIX: SINGULARITY OF THE CORRELATION MATRIX

We can easily demonstrate how the singularity of the correlation matrix appears when groups of variables are linearly dependent. For the sake of transparency we use the standard Pearson correlation. The same argument applies for polychoric or tetrachoric correlation functions, although the calculation is more involved and would make the discussion much longer and less clear.

We have explained how splitting a multiple-choice item into five dichotomous variables automatically creates a linear relationship between these variables. In other words these variables are linearly dependent on each other, i.e., any four of these variables determine the fifth.

We examine the consequence of such linear dependence in a simpler example, between two variables. Let X be a random variable with standard deviation σ_X, while another random variable Y linearly depends on it as $Y = 1 - X$; thus, the standard deviation of Y is $\sigma_Y = \sigma_X$. It is straightforward to calculate the covariance of X and Y: $\mathbf{cov}(X,Y) = \mathbf{cov}(X, 1-X) = \mathbf{cov}(X,1) - \mathbf{cov}(X,X) = 0 - \mathbf{cov}(X,X) = -\sigma_X^2$. Consequently, their correlation is $\mathbf{cor}(X,Y) = \mathbf{cov}(X,Y)/(\sigma_X \sigma_Y) = -1$, as one would expect. By definition the correlation of a variable with itself is unity, therefore the entire correlation matrix associated with the variable set $\{X, Y\}$ is

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

The determinant of this matrix is zero.

The same conclusion holds if one expands the set of variables with a third random variable Z, which is linearly independent of X. Indeed, let c_{XZ} denote the arbitrary value of the Pearson correlation between X and Z. The correlation between Y and Z is then $\mathbf{cor}(Y,Z) = \mathbf{cor}(1-X, Z) = -\mathbf{cor}(X,Z) = -c_{XZ}$, thus the correlation matrix is

$$\begin{pmatrix} 1 & -1 & c_{XZ} \\ -1 & 1 & -c_{XZ} \\ c_{XZ} & -c_{XZ} & 1 \end{pmatrix}.$$

The determinant of this matrix is again zero. In an iterative manner one can prove that further expansion of the set of random variables would not change the determinant of the correlation matrix.

We have thus shown that reformatting the answers to a five option multiple-choice question into five dichotomous variables automatically leads to a singular correlation matrix.

[1] L. Viennot, Eur. J. Sci. Educ. **1,** 205 (1979).
[2] J. Clement, Am. J. Phys. **50,** 66 (1982).
[3] D. Hestenes, M. Wells, and G. Swackhamer, Phys. Teach. **30,** 141 (1992).
[4] R. R. Hake, Am. J. Phys. **66,** 64 (1998).
[5] T. Martín-Blas, L. Seidel, and A. Serrano-Fernndez, Eur. J. Eng. Educ. **35,** 597 (2010).
[6] M. McCloskey, Sci. Am. **248,** 122 (1983).
[7] T. F. Scott, D. Schumayer, and A. R. Gray, Phys. Rev. ST Phys. Educ. Res. **8,** 020105 (2012).
[8] T. F. Scott and D. Schumayer, Phys. Rev. ST Phys. Educ. Res. **11,** 020134 (2015).
[9] G. A. Morris, L. Branum-Martin, N. Harshman, S. D. Baker, E. Mazur, S. Dutta, T. Mzoughi, and V. McCauley, Am. J. Phys. **74,** 449 (2006).
[10] G. A. Morris, N. Harshman, L. Branum-Martin, E. Mazur, T. Mzoughi, and S. D. Baker, Am. J. Phys. **80,** 825 (2012).
[11] J. Wang and L. Bao, Am. J. Phys. **78,** 1064 (2010).
[12] H. Dedic, S. Rosenfield, N. Lasry, C. Singh, M. Sabella, and S. Rebello, AIP Conf. Proc. **1289,** 125 (2010).
[13] A. Maries and C. Singh, Phys. Rev. Phys. Educ. Res. **12,** 010131 (2016).
[14] Copyright of Blackboard Inc. (1997–2017).

[15] E. Brewe, J. Bruun, and I. G. Bearden, Phys. Rev. Phys. Educ. Res. **12,** 020131 (2016).

[16] R. J. Rummel, *Applied Factor Analysis* (Northwestern University Press, Evanston, 1970).

[17] In multivariate statistics the fundamental quantity is the joint probability distribution of the random variables. However, if one focuses on the distribution of a subset of the random variables, then this distribution is called the marginal distribution. In mathematical terms, if the joint probability distribution of two random variables X and Y is denoted by $f(x, y)$, then the marginal distribution of, say, variable X is

$$f_X(x) = \int_Y f(x, y) dy.$$

The marginal distribution of X can be interpreted as the probability distribution of X irrespective of Y. For discrete random variables the integral becomes summation over the possible values of Y, and the marginal distributions can be read off from the "margins" of the contingency table of X and Y.

[18] G. M. Tallis, Biometrics **18,** 342 (1962).

[19] J. H. Edwards and A. W. F. Edwards, Biometrics **40,** 563 (1984).

[20] R. D. Bock and M. Lieberman, Psychometrika **35,** 179 (1970).

[21] S.-Y. Lee, W.-Y. Poon, and P. Bentler, Stat. Probab. Lett. **9,** 91 (1990).

[22] N. J. Higham, IMA J. Numer. Anal. **22,** 329 (2002).

[23] B. L. Welch, Biometrika **34,** 28 (1947).

[24] R. B. Cattell, Multivariate Behav. Res. **1,** 245 (1966).

[25] M. S. Sabella and E. F. Redish, Am. J. Phys. **75,** 1017 (2007).

[26] A. Franklin, Phys. Teach. **16,** 201 (1978).

[27] E. J. Dijksterhuis, *The Mechanization of the World Picture: Pythagoras to Newton* (Princeton University Press, Princeton, NJ, 1986), c. Dikshoorn (Translator).

[28] E. Fischbein, R. Stavy, and H. Ma-Naim, Int. J. Sci. Educ. **11,** 71 (1989).

[29] S. P. Marshall, *Schemas in Problem Solving* (Cambridge University Press, New York, NY, 1995).

[30] D. J. Low and K. F. Wilson, Am. J. Phys. **85,** 54 (2017).

[31] J.-i. Yasuda and M.-a. Taniguchi, Phys. Rev. ST Phys. Educ. Res. **9,** 010113 (2013).

[32] K. F. Wilson and D. J. Low, Am. J. Phys. **83,** 802 (2015).

[33] B. P. Kurz, When in doubt, just pick "C", (2016), retrieved on 26 January 2017, 09:02 AM.

[34] C. Montgomery, Most Common Answer on the ACT: Guess C?, (2016), retrieved on 26 January 2017, 09:01 AM.

[35] I. A. Halloun and D. Hestenes, Am. J. Phys. **53,** 1043 (1985).