

Are All Wrong FCI Answers Equivalent?

Helena Dedic^{1,2}, Steven Rosenfield^{1,2} & Nathaniel Lasry^{2,3,4}

1 Vanier College, Montreal Canada

2 Center for the Study of Learning and Performance, Montreal Canada

3 John Abbott College, Montreal Canada

4 School of Engineering & Applied Sciences, Harvard University, Cambridge MA, USA

Abstract. The Force Concept Inventory (FCI) has been efficiently used to assess conceptual learning in mechanics. Each FCI question has one Newtonian answer and four wrong answers (distracters). Researchers and practitioners most frequently use measures of total score to assess learning. Yet, are all wrong answers equivalent? We conducted Latent Markov Chain Modeling (LMCM) analyses of all choices (right and wrong) on a subset of four FCI questions. LMCM assesses whether there are groups of students sharing similar patterns of responses. We infer that students sharing similar patterns also share similar reasoning. Our results show seven reasoning-groups. LMCM also computes probabilities of transition from one reasoning-group to another after instruction. Examining transitions between groups, we note a clear hierarchy. Groups at the top of the hierarchy are comprised of students that use Newtonian thinking more consistently but also choose certain wrong answers more frequently; suggesting that not all wrong answers are equivalent.

Keywords: Force Concept Inventory, Latent Markov Chain model, Latent Class Analysis, knowledge in pieces, student conceptions.

PACS: 01.40.Fk, 01.40.gf, 01.30.Ha

INTRODUCTION

In introductory physics courses students bring to the classroom an understanding that is primarily derived from their perception of everyday experiences. Students may inaccurately reason from everyday experiences when attempting to solve physics problems concerning motion/interaction of physical objects. Consequently, their solutions often differ from those of experts. The Force Concept Inventory (FCI) was developed to assess students' reasoning [1]. Each FCI item has one correct Newtonian response and four incorrect responses. These incorrect answers are distracters in the sense that they are based on the most common incorrect answers provided by students during interviews.

Since its first publication researchers have been scoring FCI items as dichotomous (correct/incorrect) and reporting the score as the total number of correct answers. This score is then interpreted as a measure of the extent to which students use Newtonian reasoning. Changes in this score after instruction (*e.g.*, $\langle g \rangle$) are used to assess the extent of students' conceptual change in Newtonian reasoning, and consequently the effectiveness of instruction [2]. To further enhance our understanding of student reasoning [3,4,5], some have recently begun to investigate students' incorrect choices on the FCI.

In this study we seek to increase our understanding of student reasoning by examining all choices made in each question. We studied four FCI questions (04, 15, 16 and 28)¹ that probe students' knowledge of Newton's Third Law. We used a statistical method called Latent Markov chain modelling. This method has been used in education to examine developmental processes when the data consisted of categorical variables obtained at two time points [6],[7]. Our goal was to determine whether there are identifiable groups of students who conceive of Third Law situations in similar ways. Furthermore, we intended to show that students select different wrong answers depending upon their state of knowledge.

THEORETICAL FRAMEWORK

Students' thinking has been the focus of physics education research for many years. Much of this research initially dealt with misconceptions: coherent and robust theories that are difficult to change (*e.g.*, Vosniadou [8]). A different perspective is offered by di Sessa's [9], with knowledge seen as being

¹ We have used Latent Class Factor Analysis to identify groupings of items in the FCI. The first factor that emerged from this analysis included these four questions. We intend to report on the results of the factor analysis later.

constructed from phenomenological primitives (a.k.a. p-prims) that are derived from everyday experiences. According to this view, students do not recall a formulated theory, but rather formulate a theory on the spot from p-prims. Redish [10] expanded the notion of p-prims to reasoning primitives: statements about the physical world that cannot be explained further, and that are neither incorrect nor correct (e.g., ‘*closer implies stronger*’).

Reasoning primitives are resources that are activated when thinking about a problem, usually in association with one or more other resources. If a set of resources is frequently activated together, or as a sequence, or activated in certain contexts, then the probability of activation of this set or sequence increases. When the probability of activation of a set/sequence of resources is large, then the set/sequence can be viewed as a larger cognitive unit. In this perspective, the learning process consists of compiling primitive resources into large cognitive units. Further, as students learn, they are likely to activate these cognitive units more frequently and in more varied contexts. This is how, according to Redish [10], these cognitive units become robust. In such cases, Redish refers to them as schemas.

We adopt this terminology, and hypothesize that the probability of selecting a particular response on an FCI question depends on the probability of activation of a schema that leads to this response. The probabilistic approach of this perspective is consistent with the probabilistic approach of the statistical method we selected.

METHOD

We begin by identifying schemas that students may have used when responding to the four FCI questions. A description of the statistical methodology follows.

Schemas

Students may select a correct answer when solving the FCI questions. We hypothesize that such students have activated a Newtonian schema (**N**)². Halloun and Hestenes [11] observed students to use the dominance principle: *Motion is determined by a compromise between competing forces* in reasoning. They compiled a ‘dominance’ schema. There are two versions of the ‘dominance’ schema: *larger (or more massive) objects exert larger forces* (**D₁**); *objects that initiate motion exert larger forces* (**D₂**). They also observed students arguing that *motion is determined by physical*

obstacles in the path of moving objects. We hypothesize that these students compiled a schema that we call ‘physical obstacles’ (**PO**). Lastly, in the context of FCI question 16, Thornton et al. [12] observed students arguing that:

$$F_{net} = F_{by\ the\ car\ on\ the\ truck} + F_{by\ the\ truck\ on\ the\ car}$$

We hypothesize that these students have compiled an incorrect schema of ‘net force’. We expect that some students in our sample may have compiled the same ‘net force’ schema (**NF**).

We assume that one of these five schemas (**N**, **D₁**, **D₂**, **PO** and **NF**) is activated when a response is selected. Conversely, when a response is selected, we identify which of the five schema(s) is likely to have been activated. For example, the response 1 (“truck exerts a larger force”) to **q04** is hypothesized to indicate an activation of the ‘dominance’ schema, **D₁**.

We use labels to facilitate the reader’s recall of the problem situation in each FCI question: **q04**, involves a collision between a truck and a car, and is labelled ‘**collision**’; **q15** involves a car pushing a truck while speeding up, and is labelled ‘**speeding up**’; **q16** involves a car pushing a truck at cruising speed, and is labelled ‘**cruising**’; and **q28** involves a student A pushing a student B, and is labelled ‘**students**’.

Description of the Statistical Method

Latent Markov Chain Models (LMCM) seek to identify groups of students who respond in a similar way. We analyze 4 FCI questions using LMCM. Let Y_{it} ($i = 1, 2, 3, 4$) be observed variables, where each variable is a response to the i -th FCI question at time t ($t = 0, 1$). We think of the responses as a vector \mathbf{Y} that has components Y_{it} . Given that there are five possible responses to each question, particular values of Y_{it} can be $y_{it} = A, B, C, D, E$, given by a student on a pre-test ($t = 0$) and on a post-test, ($t = 1$).

LMCM assumes that student responses to the four questions are not independent of each other. For instance, if we were to survey subjects about the variety of foods they prefer, the pattern of responses might depend on an underlying latent (i.e., unobserved) variable such as socio-economic status. The model assumes that responses depend on a single discrete latent (i.e., unobserved) underlying variable X_t that can take on different values, x_t ($x_t = 1, 2, \dots, C$). The x_t values are referred to as ‘latent classes’ at time t and C is the number of latent classes. In our food example, analysis might reveal that there are three groups of respondents, those who prefer: expensive exotic foods (e.g., Kobe beef); home-cooked affordable food; inexpensive prepared foods. In this case, X has three values.

² We use symbols for each of these schemas, which are shown bolded inside brackets.

The sample (N = 2275) in this study comes from three distinct institutions (N_{U1} = 213; N_{U2} = 1560; and N_{U3} = 502). Thus, student responses may also depend on another variable, Z, the respondent's institution. In our food example, the food preferences may also depend on the age of respondents. In this study, the variable Z takes on values z = 1, 2, 3.

Using the notation $Y = y$ to refer to an arbitrary response pattern, we denote the conditional probability of observing y when $Z = z$ as $P(Y = y | Z = z)$. The fundamental idea of this method is that $P(Y = y | Z = z)$ is equal to the weighted average of the conditional probabilities of observed responses y from a member of class x , $P(Y = y | X = x)$ ³, with $P(X_0 = x_0 | Z = z)$ ⁴ and $P(X_1 = x_1 | (X_0 = x_0), (Z = z))$ acting as weights.

Computation of a model begins with one latent class (*i.e.*, C=1). The expected frequencies of all response patterns are then computed and compared with the observed frequencies of response patterns. If the observed frequencies do not match the expected frequencies of response patterns, the value of C is increased by 1 and the computation is repeated. The process of adding latent classes continues incrementally until differences between observed and expected frequencies can be seen as due to chance.

RESULTS

Latent Markov Chain Models C=1 to C=8 were tested. Two tests that were used to select the model C=7 indicate that in this model the differences between expected and observed frequencies are likely to be due to chance. This means that we have identified seven groups of students in the sample. We label these classes A to G. To make sense of these reasoning-classes we examined the transition probabilities from one reasoning-class on the pre-test to another class on the post-test. These results are illustrated in Table 1.

Many cells in the matrix above have a transition probability of zero. For instance, class D has zero probability to move after instruction into any other class beside itself (0.88) and class C (0.11). It is therefore possible to re-label the classes and obtain the transition probability matrix shown in Table 2.

The astounding result is that reasoning-classes organize hierarchically. That is, transitions move in a

specific direction with little or no transition 'backwards'. This hierarchy allows us to label these groups C1 through C7 in such a way so as to show that students are likely to "move" from a latent class C_m, where $m = 2, \dots, 7$, to any latent class C_n where $n < m$, after instruction. Hence, the observed hierarchy.

Table 1: Transition Probabilities between 7 Reasoning-Classes

| Pre \ Post | A | B | C | D | E | F | G |
|------------|------|------|------|------|------|------|------|
| A | 0.23 | 0.25 | 0.00 | 0.01 | 0.17 | 0.15 | 0.30 |
| B | 0.00 | 0.58 | 0.00 | 0.00 | 0.00 | 0.00 | 0.12 |
| C | 0.10 | 0.01 | 0.62 | 0.11 | 0.14 | 0.21 | 0.03 |
| D | 0.65 | 0.05 | 0.37 | 0.88 | 0.16 | 0.17 | 0.04 |
| E | 0.00 | 0.03 | 0.00 | 0.00 | 0.28 | 0.02 | 0.09 |
| F | 0.00 | 0.04 | 0.00 | 0.00 | 0.25 | 0.45 | 0.05 |
| G | 0.02 | 0.03 | 0.00 | 0.00 | 0.01 | 0.01 | 0.37 |

Table 2: Hierarchy from Transition Probabilities

| Pre \ Post | C1 | C2 | C3 | C4 | C5 | C6 | C7 |
|------------|------|------|------|------|------|------|------|
| C1 | 0.88 | 0.37 | 0.65 | 0.17 | 0.16 | 0.05 | 0.04 |
| C2 | 0.11 | 0.62 | 0.10 | 0.21 | 0.14 | 0.01 | 0.03 |
| C3 | | | 0.23 | 0.15 | 0.17 | 0.25 | 0.30 |
| C4 | | | | 0.45 | 0.25 | 0.04 | 0.05 |
| C5 | | | | | 0.28 | 0.03 | 0.09 |
| C6 | | | | | | 0.58 | 0.12 |
| C7 | | | | | | | 0.37 |

To identify how students in a latent class might be reasoning we examined all response patterns in this latent class and paid attention to those with high frequency. We also constructed response patterns with high conditional probability of being observed in this latent class from the conditional probabilities of each response to each question. We note that a very large number of response patterns, each with low frequency, are classified as belonging to the two latent classes C6 and C7. Furthermore, responses that indicate the use of the 'physical obstacle' schema, the conditional probabilities are significant amongst students in these two latent classes. This implies that the population of these two classes (17.4% of the sample) consists of outliers, probably novices, and/or students who are thinking inconsistently. However, response patterns allow us to describe more clearly the thinking of students in each of latent classes C1 through C5. Table 3 below lists information concerning the reasoning of students in each of these classes.

³ Assuming that observed responses are independent of each other within each class and when they were measured, we obtain

$$P(Y = y | X = x) = \prod_{t=0}^1 \prod_{t=1}^4 P(Y_{it} = y_{it} | X_t = x_t).$$

⁴ The probability of belonging to a class x_0 on pre-test when $Z = z$, $P(X_0 = x_0 | Z = z)$; the transition probability of belonging to a class x_1 on post-test when $X_0 = x_0$ and $Z = z$, $P(X_1 = x_1 | (X_0 = x_0), (Z = z))$.

Table 3: Patterns with highest conditional probabilities in each class⁵

| | q4 (collision) | q15 (speeding-up) | q16 (cruising) | q28 (students) |
|-----------|----------------------------------|--|--|----------------------------------|
| C1 | E (N; 0.98) | A (N; 0.97) | A (N; 0.94) | E (N; 0.99) |
| C2 | E (N; 0.76) | C (D₂ , NF ; 0.95) | A (N, NF ; 0.94) | E (N; 0.83) |
| C3 | E (N; 0.64) | A (N; 0.77) | A (N, NF ; 0.76) | E (N; 0.66) |
| C4 | A (D₁ ; 0.91) | C (D₂ , NF ; 0.87) | A (N, NF ; 0.88) | D (D₁ ; 0.76) |
| C5 | A (D₁ ; 0.88) | C (D₂ , NF ; 0.90) | C (D₂ , NF ; 0.90) | D (D₁ ; 0.70) |

DISCUSSION

The transition probabilities indicate that there is an observable hierarchy of reasoning-classes, with C1 at the top of the hierarchy, and C6 and C7 at the bottom.

In this study the ordering arose from an analysis of empirical data from a very large sample of students. The ordering was not postulated theoretically. While the ordering of latent classes is a result of data analysis, a posteriori this ordering provides a justification for assumptions that we have made concerning schema(s) used in answering FCI questions. It is on the basis of these assumptions that we have described reasoning by members of latent classes C1 through C5.

Students who are likely to select wrong responses indicating the use of the ‘physical obstacle’ schema, **PO**, belong in classes at the “bottom” of the hierarchy (C6 and C7). Students who consistently select wrong responses indicating the use of the ‘dominance’ schema, **D₁** or **D₂**, belong in latent class C5, next up from the bottom of the hierarchy. We observe that the majority of C5 class members (72% of them) use **D₁** or **D₂** schemas less consistently and are more likely to use **N** more often after instruction. Even though only a few of them became consistent users of **N**, we observe the evolution of knowledge in terms of decreasing reliance on **D₁** or **D₂** schemas. One possibility is that some students are not correctly learning to use net forces and resort to the **NF** schema, prevalent in C2 and C4, during their mechanics courses. Indeed, students starting in C5 have a combined probability of 0.39 of ending up in C2 or C4 where **NF** is prevalent.

It is important to note that latent classes should not be viewed as a ladder of developmental stages. For example, students who initially belong in latent class C3 have a probability of remaining in this latent class equal to 0.23. Their probability of moving into latent class C2 is equal to 0.10, but their probability of moving into latent class C1, consisting of Newtonian thinkers, is 0.65. Hence latent classes are not developmental stages, because students in C3 need not go through C2 in order to arrive at C1.

CONCLUSION

To answer the question posed in our title, all wrong answers are not equivalent. Our latent Markov model, examining all answers, shows that students’ knowledge evolves and that there is a hierarchy of wrong answers. Understanding moves “upwards”, both in complexity and consistency of use. At the bottom, students use the most “wrong” schema (**PO**), then “less wrong” schemas (**D₁** or **D₂**), and then the “least wrong” schema (**NF**). While students likely compile schemas, **PO**, **D₁** and **D₂**, on their own prior to entering our classrooms, perhaps instructors can modify their instruction so as to prevent students from ever using a ‘net force’ schema.

REFERENCES

- 1 D. Hestenes, M. Wells and G. Swackhamer, *Physics Teacher* **30**, 141-158 (1992).
- 2 R. Hake, *American Journal of Physics* **66**, 64-74, (1998).
- 3 L. Bao and E. Redish, *Physical Review Special Topics – Physics Education Research* **2**, 010103 (2006).
- 4 L. Bao, K. Hogg and D. Zollman, *American Journal of Physics* **70**, 766-778 (2002).
- 5 G. A. Morris, L. Branum-Martin and N. Harshman, *American Journal of Physics* **74**, 449-453 (2006).
- 6 J. K. Vermunt and J. Magidson, (in S. Menard (ed.) *Handbook of Longitudinal Research: Design, Measurement and Analysis*, 373-385, Burlington, MA: Elsevier 2008).
- 7 D. Kaplan, *Developmental Psychology* **44**, 457-467 (2008).
- 8 S. Vosniadou and W. F. Brewer, *Cognitive Psychology* **24**, 535-585 (1992).
- 9 A. DiSessa, *Cognition and Instruction*, **10**, 105-225 (1993).
- 10 E. Redish, *Proceedings of the International School of Physics “Enrico Fermi”*, Varenna, Italy (2003).
- 11 I. Halloun and D. Hestenes, *American Journal of Physics* **53** (11), 1043-1055 (1985).
- 12 R. K. Thornton, D. Kuhl, K. Cummings and J. Marx, *Physical Review Special Topics – Physics Education Research* **5**, 010105 (2009).

⁵ In each cell we show the single response (A to E) with highest probability for the question of that column by a member of the class for that row, followed by the schema (N, **D₁**, **D₂**, **NF**) associated with that response, and then the conditional probability of that response from a member of that class, e.g., the probability of selecting response E to **q4**, which corresponds to schema N, by a member of a class C2, is 0.76.