

ABSTRACT

RICHARDS, EVAN THOMAS. Selecting the Correct Solution to a Physics Problem When Given Several Possibilities. (Under the direction of Ruth W. Chabay.)

Despite decades of research on what learning actions are associated with effective learners (Palincsar and Brown, 1984; Atkinson, et al., 2000), the literature has not fully addressed how to cue those actions (particularly within the realm of physics). Recent reforms that integrate incorrect solutions suggest a possible avenue to reach those actions. However, there is only a limited understanding as to what actions are invoked with such reforms (Große and Renkl, 2007).

This paper reports on a study that tasked participants with selecting the correct solution to a physics problem when given three possible solutions, where only one of the solutions was correct and the other two solutions contained errors. Think aloud protocol data (Ericsson and Simon, 1993) was analyzed per a framework adapted from Palincsar and Brown (1984). Cued actions were indeed connected to those identified in the worked example literature. Particularly satisfying is the presence of internal consistency checks (i.e., are the solutions self-consistent?), which is a behavior predicted by the Palincsar and Brown (1984) framework, but not explored in the worked example realm. Participant discussions were also found to be associated with those physics-related solution features that were varied across solutions (such as fundamental principle selection or system and surroundings selections).

Selecting the Correct Solution to a Physics Problem When Given Several Possibilities

by
Evan Thomas Richards

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Physics

Raleigh, North Carolina

2010

APPROVED BY:

Dr. Bruce Sherwood

Dr. John Risley

Dr. Robert Beichner
Co-Chair of Advisory Committee

Dr. Ruth Chabay
Co-Chair of Advisory Committee

DEDICATION

To my wife, Holly, who has kept me going for many years through thick and thin.

Without her encouragement and support throughout the years, none of this would have been possible. To many more years together, down the road less traveled...

BIOGRAPHY

I have taken the scenic route to arrive at North Carolina State University's (NCSU) Physics Education R&D Group. Since I was young, I have always enjoyed exploring the technical fields. After earning a few degrees in engineering, I worked in the semiconductor industry for a while as an engineer. However, my interest in physics never wavered, and I began to take physics courses on the side. During the same time period, I also fully realized my interest in teaching.

One fateful night changed the course of my life. As I was waiting for a professor to arrive for office hours, I pursued a nearby billboard of available physics opportunities. One such opportunity combined the sciences of physics and teaching. This notice introduced me to the world of physics education research. After returning home, I worked through the night to learn about this exciting field of research. By the next morning, I knew that my life was about to dramatically change its course.

Soon after that night, I returned to the ranks of a full time graduate student to finish my M.S. degree in physics. At last, I was ready to dive into the field of physics education research, where I found a warm welcome at NCSU.

ACKNOWLEDGMENTS

I had the great fortune and honor to have found incredible mentors throughout my time at a variety of universities. Without their help, support, and tremendous patience with my never-ending questions, my dreams would be much further away.

I thank the faculty of the North Carolina State University's Physics Education R&D Group (PERD group) for training me in the field of physics education research. I have transitioned from a naïve graduate student thinking of teaching solely through the lens of anecdotal experiences to an instructor with a more disciplined view on the science of teaching. I also want to thank the National Science Foundation for their funding and support.

I thank my researcher advisor, Dr. Ruth Chabay, for her insight and guidance generously given over the past several years. Her disciplined approach to research has been a key influence to my professional development. My thoughts continually turn to asking, "how would Ruth respond?"

I thank Dr. John Risley for providing early opportunities to gain experience with physics education research. I owe much of my early training in qualitative research methodology to him. I also owe my first teaching experiences in a studio-learning environment to his kind offer to be a teaching assistant. Much of my current teaching approach directly stems from his example. I deeply appreciate his advice and guidance over the years.

I thank Dr. Robert Beichner for his insightful comments and questions over the years. On countless occasions, his insights have caused me to reflect upon (and in many cases redefine) my initial analysis. Without his support and the North Carolina State University's STEM Education Initiative, the Qualitative Educational Research Laboratory would not have been possible. I greatly enjoyed my time building and using the laboratory, not to mention my appreciation on the knowledge that I gained through those efforts. I am continually amazed by his boundless energy. Despite his busy schedule, he always made time to talk whenever I (or any other graduate student) needed advice, and his humor always made those conversations particularly memorable (may I never forget "bogus points").

I thank Dr. Bruce Sherwood for his kind advice and equally wonderful humor over the years. He provided a never-ending stream of fascinating physics problems to ponder, many of which I think back upon within the context of my own courses. He has always been generous with his time, and I deeply appreciate the many conversations we have had. His insights helped hone my own thinking on many occasions.

I thank Dr. H. Vince Kuo now of the Colorado School of Mines. I very much appreciated his time as a postdoctoral research associate at North Carolina State University. His door was always open for the graduate students. I deeply appreciate the training and advice he gave me (not to mention showing me around my first AAPT meeting).

Before finishing their doctoral degrees, Dr. Matthew Kohlmyer and subsequently Dr. Lin Ding were the senior graduate students when I started at North Carolina State

University. They were always happy to provide guidance and answers to my many questions as a new graduate student to the research group. When I assumed the same role later on, I tried to uphold their superb legacy.

I also extend my thanks to my fellow graduate students in the PERD group. In many ways, our progress through the program is a result of a collective effort not only on the part of the faculty, but also the graduate students. I appreciate the many sanity checks and sounding board sessions. I especially want to thank Jon Gaffney for his collaboration and our many insightful conversations, Jeff Polak for his above-and-beyond the call of duty efforts in helping me with this research, Brandon Lunk for his help with the data collection, and Shawn Weatherford for his contributions to the coding scheme verification.

I extend my appreciation to Dr. Roy Chaney and Dr. Gregory Earle of the University of Texas at Dallas. They made it possible for me to return to graduate school and follow my dreams. I thank Dr. Glenn Smith of the Georgia Institute of Technology for his mentorship (his advice heavily influenced my decision to return to graduate school to pursue my dreams). I thank Dr. Kai Chang and Dr. Robert Nevels for starting me on the path to academia at Texas A&M University. Thanks to them, I gained my first research experiences. They fostered my sense of fascination and wonder along instilling a work ethic that has seen me through life's winding path.

Most importantly, I thank my family for their encouragement. In particular, I want to thank my father, Tom Richards. He introduced me to the technical fields that in large part set me on the path that has led me to North Carolina State University.

On a special note, I thank my mother, Nan Richards, for her constant support. More so than anyone else, she made it possible for me to return to graduate school to pursue my physics degrees. I extend my deepest thanks to her for always keeping my vision toward my dreams, and taking the road less traveled...

TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES	xiii
CHAPTER 1: INTRODUCTION	1
CHAPTER 2: LITERATURE REVIEW	4
2.1: Worked Examples versus Conventional Problem Solving	7
2.2: Observed Actions	15
2.2.1: Good versus Poor Performers	17
2.2.2: Action Profiles	30
2.3: Modified Worked Examples	44
2.3.1: Action Leveraged Modifications	45
2.3.2: Memory Model Inspired Modifications	51
2.3.2.1: Integrated Worked Examples	52
2.3.2.2: Modular Worked Examples.....	55
2.3.2.3: Multimode Worked Examples.....	59
2.3.3: Transition Implementations	62
CHAPTER 3: METHODOLOGY	69
3.1: Task Development.....	69
3.1.1: Problem Set Development	71
3.2: Structure of the Study	73
3.2.1: Participant Background	74

3.2.2: Study Sequence	75
3.2.3: Study Environment.....	76
3.3: Analysis Methodology	76
3.3.1: Action-Coding Pass.....	79
3.3.2: Context-Coding Pass	82
3.3.3: Focus-Coding Pass	83
3.3.4: Correct/Incorrect-Coding Pass.....	84
3.3.5: Restate Type-Coding Pass	85
CHAPTER 4: RESULTS	87
4.1: Task Actions.....	88
4.1.1: Connections to Palincsar and Brown (1984)	89
4.2: Distinctions in the Data.....	97
4.2.1: Distinctions across Participants	98
4.2.2: Distinctions across Problem Sets	103
CHAPTER 5: CONCLUSIONS	132
5.1: Future Work Directions	132
REFERENCES	137
APPENDICES	148
APPENDIX A: PROBLEM AND SOLUTION SETS.....	149
A.1: “Close the Door” problem and solution set	149
A.2: “UCM Spring” problem and solution set	152
A.3: “Fission” problem and solution set	156

A.4: “Two Pucks” problem and solution set	160
APPENDIX B: RESEARCHER SCRIPTS	164
B.1: Welcome and introduction script – lead-in to the protocol warm-up exercises.....	164
B.2: Think aloud protocol warm-up exercises.....	166
B.3: Follow-up script to the warm-up exercises – the final instructions to the participants	167
B.4: Reference Sheet.....	168
APPENDIX C: FOCUS CODING BREAKPOINTS.....	171
C.1: “Close the Door” solutions	171
C.2: “UCM Spring” solutions.....	173
C.3: “Fission” solutions.....	176
C.4: “Two Pucks” solutions	179

LIST OF TABLES

Table 2.1	Four experimental states for the proximity effect.....	9
Table 2.2	Worked example modifications.....	65
Table 3.1	Error distribution across information representations (an “X” indicates the presence of an error(s) in the associated representation)	72
Table 3.2	Coding pass overview	77
Table 3.3	Coding reliability	78
Table 4.1	Whitney-Mann U-test results for Judgers vs. Others average percent actions for all problems.....	102
Table 4.2	Planning segments from “Close the door” problem	104
Table 4.3	Whitney-Mann U-test results for Judgers vs. Others average percent actions for “UCM spring” problem	106
Table 4.4	Whitney-Mann U-test results for Judgers vs. Others average percent actions for “Fission” problem.....	107
Table 4.5	Whitney-Mann U-test results for Judgers versus Other absolute change in actions for “Close the door” with “Two Pucks” against “UCM” and “Fission”	109
Table 4.6	Math Context Judgments for the “UCM Spring” problem	116-117
Table 4.7	Physics Context Judgments for the “Fission” problem.....	118-119

Table 4.8	Example segments of participants rejecting the correct solution to “Two Pucks”	121
Table 4.9	Physics Context Judgments for the “Close the door” problem	123-124
Table 4.10	Example segments of undergraduates judging “Two Pucks” based upon a preconceived answer	131

LIST OF FIGURES

Figure 2.1	Underlying connections between Palincsar and Brown (1984) factors and worked example study actions (the factor number appears in parentheses).....	40
Figure 2.2	Links across worked example implementations (the bold text represents shared underlying ideas of the implementations)	66
Figure 4.1	Ven Diagram illustrating the connections of the participant actions exhibited in the study to the actions in the Palincsar and Brown (1984) framework	89
Figure 4.2	Average percent nature of restate segments.....	95
Figure 4.3	Diagram in the “Two pucks” problem statement	96
Figure 4.4	Diagram in the “UCM” problem statement.....	96
Figure 4.5	“Igor’s” percent actions (note the concentration upon Judging, Restating, and Other)	99
Figure 4.6	Judgers vs. Others Actions for “Close the door” problem	100
Figure 4.7	Judgers vs. Others Actions for “UCM Spring” problem.....	100
Figure 4.8	Judgers vs. Others Actions for “Fission” problem	101

Figure 4.9	Judgers vs. Others Actions for “Two Pucks” problem	101
Figure 4.10	Judgers vs. Others actions for all problems.....	102
Figure 4.11	Judgers vs. Others actions for the “Close the door problem”.....	103
Figure 4.12	Judgers vs. Others actions for the UCM spring problem.....	106
Figure 4.13	Judgers vs. Others actions for the “Fission” problem.....	107
Figure 4.14	Judgers vs. Others absolute change in average % actions for Close the door and Two Pucks minus UCM Spring and Fission	109
Figure 4.15	Average percent correct and incorrect judgments	110
Figure 4.16	Percent incorrect judgment context for “UCM”	111
Figure 4.17	Percent incorrect judgment context for “Fission”.....	111
Figure 4.18	Percent incorrect judgment context for “Close the door”	112
Figure 4.19	Percent incorrect judgment context for “Two Pucks”	112
Figure 4.20	Average Percent Context of the Judging Segments.....	113
Figure 4.21	Detailed Average Percent Context of the Judging Segments.....	114
Figure 4.22	Average number of equal signs in the given solutions for each problem.....	115
Figure 4.23	Average percent actions for each problem.....	126

Figure 4.24	Average percent focus of comparisons	127
Figure 4.25	Average percent context of judgments for “Two Pucks”	128
Figure 4.26	Detailed average percent context of judgments for “Two Pucks”	129
Figure A.1	“Close the door” problem statement	149
Figure A.2	“Close the door” given solution #1	150
Figure A.3	“Close the door” given solution #2	151
Figure A.4	“Close the door” given solution #3	152
Figure A.5	“UCM Spring” problem statement	152
Figure A.6	“UCM Spring” given solution #1	153
Figure A.7	“UCM Spring” given solution #2	154
Figure A.8	“UCM Spring” given solution #3	155
Figure A.9	“Fission” problem statement	156
Figure A.10	“Fission” given solution #1	157
Figure A.11	“Fission” given solution #2	158
Figure A.12	“Fission” given solution #3	159
Figure A.13	“Two Pucks” problem statement	160
Figure A.14	“Two Pucks” given solution #1	161

Figure A.15	“Two Pucks” given solution #2	162
Figure A.16	“Two Pucks” given solution #3	163
Figure B.1	First page of reference sheet.....	169
Figure B.2	Second page of reference sheet.....	170
Figure C.1	“Close the door” given solution #1 with breakpoint.....	171
Figure C.2	“Close the door” given solution #2 with breakpoint.....	172
Figure C.3	“Close the door” given solution #3 with breakpoint.....	173
Figure C.4	“UCM Spring” given solution #1 with breakpoint	173
Figure C.5	“UCM Spring” given solution #2 with breakpoint	174
Figure C.6	“UCM Spring” given solution #3 with breakpoint	175
Figure C.7	“Fission” given solution #1 with breakpoint.....	176
Figure C.8	“Fission” given solution #2 with breakpoint	177
Figure C.9	“Fission” given solution #3 with breakpoint	178
Figure C.10	“Two pucks” given solution #1 with breakpoint	179
Figure C.11	“Two pucks” given solution #2 with breakpoint	180
Figure C.12	“Two pucks” given solution #3 with breakpoint	181

CHAPTER 1: INTRODUCTION

Decades of research have demonstrated that worked examples can be used as effective instructional material (overviews of such research include: Sweller, et al., 1998; Atkinson, et al., 2000; Renkl, et al., 2002; van Merriënboer, et al., 2005).

Earlier work tended to focus upon comparing the use of worked examples versus problem solving practice (some examples are: Sweller and Cooper, 1985; Cooper and Sweller, 1987; Ward and Sweller, 1990; Trafton and Reiser, 1993). Such work generally argued for the use of worked examples as opposed to problem solving only.

Later work tended to focus on what processing was associated with such learning gains. Not only has the literature reported on the beneficial learning actions in the field of physics (Chi, et al., 1989 and 1991), but other areas such as computer programming (Pirolli and Recker, 1991), probability (Renkl, 1997), and accounting (Stark, et al., 2002) have been explored. Such studies have provided a profound understanding as to what actions are associated with effective learners.

Despite knowing what actions are associated with effective learning, the literature is not clear on how to get participants to do those actions within the realm of physics. Much of this work has two complications: largely untested within the realm of physics (some examples are Stark, 1999; Cantrambone, 1995, 1996, 1998; Renkl, et al., 1998) or yield mixed results with regards to learning gains such as seen in Conati and VanLehn (2000). Hence, the issue of cueing the beneficial learning actions largely remains an open question for the field of physics.

However, the ascension of reforms employing incorrect solutions provided another interesting avenue for possible cueing (some examples include Evaluation Tasks from Etkina, et al., 2006; What If Anything is Wrong Tasks with TIPERS¹; WRONG Problems from Harper, et al., 2007; Reciprocal Teaching as applied in Reif and Scott, 1999). There are unfortunately limited results available on the processing of interventions that include incorrect solutions (Große and Renkl, 2007), particularly within the realm of physics. Hence, the processing invoked by the integration of incorrect solutions has largely not been previously explored².

This study addresses the open issue of processing tasks integrating incorrect solutions. In particular, will such a task cue the beneficial learning actions within the realm of physics? The participants were given three possible solutions to a physics problem. Only one of the solutions was correct, while the other solutions had errors. The participants were asked to identify the correct solution and the errors in the other solutions. Not only is this work focused upon identifying what actions were initiated in response to this task, but is also focused upon whether the related actions from the reading comprehension literature (as summarized in Palincsar and Brown, 1984) emerge with this task.

Overall, the results are encouraging. Actions related to those from prior work do indeed emerge. Particularly satisfying is the emergence of the internal consistency check

¹ TIPERS (Tasks Inspired by Physics Education Research), see <http://tycphysics.org/tipers.htm>, downloaded on 11/8/09

² There certainly has been work focused on the assessment of reforms integrating incorrect solutions (such as Etkina, et al., 2006), however, little work is available on the processing of such reforms.

action as predicted by Palincsar and Brown (1984) but unseen in the worked example literature.

The results further suggest that participant discussion tended to include those physics-related solution features that were varied across solutions (such as fundamental principle selection or final and initial state selections).

To report on the answers to the above research questions, this paper is organized into three sections. The first section reviews the prior related research with a particular focus on the worked example literature, augmented by the reading comprehension literature. Most relevant to the questions addressed in this study, the review places an emphasis on the available beneficial learning action frameworks from both areas of the literature, and the connections spanning the worked example realm to the reading comprehension realm.

The second section encompasses both the development of the task itself as well as the research methodology employed for the data collection and subsequent analysis.

The final section discusses the results as well as suggested avenues for future work with a particular interest in implications for instructional intervention development.

CHAPTER 2: LITERATURE REVIEW

This chapter discusses the development of the research leading to the study reported in this paper. Early on, conventional wisdom considered that problem solving practice was the best way to learn new material. However, research began to question this assertion, and began to compare using worked examples against problem solving practice (for example, see Sweller and Cooper, 1985; Cooper and Sweller, 1987; Zhu and Simon, 1987). Much of such research tended to favor the use of worked examples. Not surprisingly, researchers desired to understand how worked examples aided learning. Hence, the research began to focus on what actions are invoked when learning from worked examples (for instance, see the seminal studies of Chi, et al., 1989 and 1991). After gaining a robust understanding of what actions are beneficial for learning, researchers then turned to designing worked examples (or more generally, supplementary materials) to foster those beneficial actions (some examples are Ward and Sweller, 1990; Catrambone, 1998; Renkl, et al., 1998). Unfortunately for those interested in implementing such reforms in the realm of physics, the research is largely untested for physics or yields rather mixed results. In addition, there are limited results for reforms integrating incorrect solutions (an exception is Große and Renkl, 2007), and yet such reforms are arising in the research (such as Reif and Scott, 1999 and Etkina, et al., 2006).

As discussed in the previous chapter, there are several relevant branches of research leading to the study reported in this paper.

1. Early research assessing the use of worked examples as learning material

2. Observed actions that students exhibit while examining worked examples
3. Worked example implementation reforms

This chapter discusses each of these branches of research in separate sections.

Section 2.1 examines the literature providing evidence of worked examples aiding learning. The use of worked examples have demonstrated advantages in learning not only in the realm of physics (Ward and Sweller, 1990), but also in other subjects such as mathematics (Sweller and Cooper, 1985; Cooper and Sweller, 1987; Zhu and Simon, 1987), and computer programming (Trafton and Reiser, 1993). The advantages of worked examples leads to the underlying questions covered in the subsequent sections: what activities do students engage in while using worked examples, and can worked examples be improved upon? If so, then how?

Section 2.2 delves into the actions of students while examining worked examples. In general, the actions fall into four categories: monitoring (Chi, et al., 1989; Pirolli and Recker, 1994; Renkl, 1997; Stark, et al., 2002), application of prior knowledge (Chi, et al., 1989; Pirolli and Recker, 1994; Renkl, 1997), extraction of knowledge (Chi, et al., 1989 and 1991; Pirolli and Recker, 1994; Zhu, et al., 1996; Stark, et al., 2002), and prediction (Renkl, 1997). Monitoring refers either to the acknowledgement of a comprehension failure (negative monitoring) or the affirmation of understanding (positive monitoring). Application of prior knowledge encompasses the use of relevant background information to make sense of a worked example, while extraction of knowledge refers to the inference or abstraction of information. Finally, prediction references activities related to anticipating a result and subsequently checking if the result

actually occurs. Section 2.2 concludes with a discussion of the connections between the observed actions and a framework from the reading comprehension literature (Palincsar and Brown, 1984).

Section 2.3 reviews research focused upon improving worked examples. The reformed implementations can be grouped into three classes: design elements, complementary scaffolding, and presentation strategies (see Table 2.2). Design elements refer to the use of a non-traditional layout of the worked example (such as integrating the solution directly into the problem statement as in Ward and Sweller, 1990) or the inclusion of non-traditional constituents (such as the subgoal labeling researched in Catrambone, 1995, 1996, and 1998). Complementary scaffolding provides support in addition to the worked example. For instance, Renkl, et al. (1998) provided training on active methods to examine worked examples. Presentation strategies refer to methods of configuring and delivering information supplied in the worked example (such as fading³ from Renkl, Atkinson, Maier, and Staley, 2002). As with the actions in section 2.2, much of the underlying ideas of the implementations in section 2.3 are shared with propositions listed by Palincsar and Brown (1984). Palincsar and Brown (1984) listed four training design elements: active participation, feedback on active participation, include information on how and when to perform pertinent actions, and structure scaffolding to a level appropriate for the student (i.e., ensure that the instructional material is not at a level beyond the student's ability to master or at a level that the student has already mastered).

³ Fading is a method of presenting a sequence of worked examples, where the solution for each subsequent worked example contains fewer presented steps, leaving the student to complete the missing steps.

Section 2.3 concludes with a discussion on how the worked example implementations are in part extensions of the training design propositions listed above.

2.1: Worked Examples versus Conventional Problem Solving

A perception quoted in a review of worked example research by Atkinson, Derry, Renkl, and Wortham (2000) states that problem solving should be taught by requiring students to solve multitudes of problems. Zhu and Simon (1987) called this “learning by doing” (p. 138)⁴. However, this adage came into question as several studies compared the use of worked examples as an alternative to conventional problem solving (Sweller and Cooper, 1985; Cooper and Sweller, 1987; Ward and Sweller, 1990; Trafton and Reiser, 1993). Indeed based on self-reported data, Kim and Pak (2002) present evidence of a limited correlation between conceptual understanding and solving a substantial number (hundreds or in some cases thousands) of problems.

The overarching research questions within the aforementioned studies are generally centered about whether the use of worked examples paired with related problems as instructional material yields comparable or superior performance on subsequent problem solving as compared to providing paired, practice problems as instructional material. Furthermore, questions of learning efficiency are also addressed: do students learn more efficiently from worked examples as opposed to traditional instruction?

⁴ For consistency, I will adopt the term “conventional problem solving” typically used by Sweller and his colleagues (for instance, see Tarmizi and Sweller, 1988; Ward and Sweller, 1990).

In examining the use of worked examples as instructional material, this early work encompassed only unmodified⁵ worked examples. Typically, these type of studies provide instructional material containing related worked example-problem pairs to half of the participants, while providing only corresponding pairs of problems to the remaining participants. Following a learning period, where the participants are allowed to examine the instructional material, a post-test is administered.

Sweller and Cooper (1985) recruited eighth and ninth grade mathematics students to participate in a series of studies comparing isomorphic⁶ and transfer⁷ post-test performance covering topics in algebraic manipulation. No differences in transfer performance between the worked example group and conventional problem solving group were reported. However, the worked example groups required less time to complete the isomorphic post-test problems and yielded superior isomorphic performance.

A proximity effect related to isomorphic performance emerged. When the transfer post-test problems preceded the isomorphic post-test problems (effectively increasing the “distance” between the worked examples and the isomorphic post-test problems),

⁵ For the purposes of this review, I define an unmodified worked example as a worked example that is not constructed nor have additional scaffolding developed on a research basis.

⁶ I will use the term isomorphic problem to describe the class of problems that share a common structure with the worked examples and/or problems in the instructional material.

⁷ I will use the term transfer post-test problem to describe the class of problems that have a structure that varies from the worked example and /or problems in the instructional material.

isomorphic post-test performance decreased as compared to reordering the post-test problems to provide a closer proximity of the isomorphic problems to worked examples.

Trafton and Reiser (1993) extended the proximity effect to a different domain, LISP programming. They examined the proximity effect of four different example and problem sequences (see table 2.1).

Table 2.1 - Four experimental states for the proximity effect

Alternating Example	Example (topic 1), Problem (topic 1)	Example (topic 2), Problem (topic 2)	...	Example (topic 4) Problem (topic 4)	...
Alternating Solve	Problem (topic 1), Problem (topic 1)	Problem (topic 2), Problem (topic 2)	...	Problem (topic 4), Problem (topic 4)	...
Blocked Example	Example (topic 1), Example (topic 2)	Example (topic 3), Example (topic 4)	...	Problem (topic 1), Problem (topic 2)	...
Blocked Solve	Problem (topic 1), Problem (topic 2)	Problem (topic 3), Problem (topic 4)	...	Problem (topic 1), Problem (topic 2)	...

The alternating example group received instructional material that placed the worked example and problem pairs together (the alternating solve group placed the associated problems together), whereas the blocked example group was provided with instructional material that separated the worked example from the associated problem (similarly, the blocked solve group had instructional material that separated the associated problems).

Undergraduate volunteers with minimal previous exposure to programming (and no prior LISP instruction) were given an introduction to LISP. As with Sweller and Cooper (1985), the subjects were subsequently given the instructional material followed by a post-test.

Consistent with the proximity effect in Sweller and Cooper (1985), the alternating worked example group outperformed the blocked worked example group. In contrast to Sweller and Cooper (1985), the blocked solve group outperformed the blocked example group. Trafton and Reiser (1993) propose that the lowered recall levels due to the blocked condition hinder the value of the examples, whereas solving problems in the blocked condition provides additional practice, thereby aiding subsequent problem solving. Given the increased complexity of the LISP domain as compared to the Sweller and Cooper study (1985), the participants would presumably need to maintain a more substantial information set in memory, possibly impacting post-test performance. More extensive domain and participant variations are required to confirm this conjecture.

In particular, Trafton and Reiser (1993) were interested in testing two premises that have been proposed to characterize the impact of worked examples on learning. Sweller and his colleagues (Sweller and Cooper, 1985; Cooper and Sweller, 1987; Tarmizi and Sweller, 1988; Sweller, et al., 1998) largely frame their discussion of worked examples in terms of schema⁸ acquisition. Worked examples are contended to foster formation of schemas, which affords the student procedural information that can be applied to subsequent problem solving. From a related viewpoint, learners generate a rule set for problem solving based upon the information presented in worked examples (VanLehn, et al., 1992). Trafton and Reiser (1993) label such a viewpoint as example generalization.

⁸ Schema are “...mental frameworks for representing knowledge...” (Sternberg, 2003, p.254) in long term memory that, “...categorizes elements of information according to the manner in which they will be used.” (Sweller, et. al, 1998, p.255).

Alternatively, the knowledge compilation view contends that pertinent knowledge application information is obtained when solving problems rather than when examining worked examples. Instead, worked examples simply provide declarative information that forms the basis of subsequently obtained problem solving knowledge (Anderson, 1987; Pirolli, 1991; Pirolli and Recker, 1994).

Trafton and Reiser (1993) proposed a series of predictions based upon both viewpoints.

The knowledge compilation theory (applied knowledge is gained through working problems) would predict that the blocked solve sequence would be preferred versus the blocked example sequence since knowledge is retained from working problems as opposed to studying examples. By extension, the notion that little knowledge is retained from examples implies that the alternating example sequence would be superior to the blocked example sequence, since the alternating sequence has a smaller separation between examples and their related problems.

In contrast, if one ascribes to the view that much knowledge is indeed retained from worked example study, then either example sequence (alternating or blocked) would be equivalent. In addition, one would predict that the example sequences would be preferred to the problem (or solve) sequence.

As noted above, the data from Trafton and Reiser (1993) does support the knowledge compilation viewpoint. However, their comparison examines a rough knowledge grain size. In other words, no allowance is made for the possibility that

elements (but not necessarily a complete framework) of problem solving knowledge are obtained during worked example study, much less from prior learning episodes. In the absence of supporting protocol data, such distinctions have not been definitively addressed.

Unfortunately, Trafton and Reiser (1993) did not pursue transfer effects. However, Cooper and Sweller (1987) revisited the question of transfer performance raised in their earlier paper (Sweller and Cooper, 1985). They hypothesized that a less complex domain (a subset of the algebraic manipulations encompassed by the earlier paper) would aid transfer performance since less material would be required to be learned within a limited amount of time. With the less complex domain, the initial studies indicated a ceiling effect (no isomorphic performance differences), while confirming a difference in transfer performance (favoring the worked example group over the conventional problem solving group).

The subsequent studies in the Cooper and Sweller (1987) paper uncovered two factors that were associated with variations in post-test performance (in addition to the aforementioned proximity effect): student background and amount of instructional material. High school participants were recruited from a more advanced mathematics course (stronger background) and a less advanced mathematics course (weaker background). Furthermore, the knowledge acquisition period was varied. The high practice groups (longer knowledge acquisition period) received twelve sets of worked example-problem pairs (for the worked example groups) or problem-problem pairs (for

the conventional problem solving groups), while the low practice groups received only four sets of pairs.

Consistent with the aforementioned studies, the worked example group outperformed the conventional problem solving group on the isomorphic post-test problems, but only for the low practice, weak background students (i.e., the students presumably requiring the most aid). When compared to the longer acquisition times, such data suggests that learning from worked examples is more efficient than conventional problem solving (i.e., the effect of the worked examples manifests earlier for the weaker background students than with conventional problem solving). The worked example impact emerged with the transfer problems for low practice groups (both strong and weak backgrounds). Yet again, the worked example groups required less instructional material to yield superior performance as compared to the conventional problem solving groups (in this case, the more challenging transfer problems presumably degraded the background impact). However, significant worked example transfer gains for the long acquisition time groups (as opposed to their conventional problem solving counterparts) only emerged for the weaker background participants. Thus, for the more difficult transfer problems, the weaker background students needed the impact of the worked examples (over a more substantial knowledge acquisition phase) to achieve significant gains beyond conventional problem solving.

The above studies sacrificed ecological validity in lieu of a controlled laboratory setting. Zhu, et al. (1987) incorporated a more ecologically valid design by comparing

traditional classroom instruction (a control group, “... who were taught by lecture in the normal way.”, Zhu and Simon, 1987, p.156) against instruction emphasizing the use of worked examples. Of course, testing at the classroom level introduces additional confounds that were damped in the laboratory controlled environments (such as variations across schools, classroom populations, instructor variations, etc.). In turn, the results from Zhu, et al. (1987) were more mixed than in the discussion so far. Initially, selected topics spanning the areas of geometry to polynomials were tested. Overall, the results suggest that worked example instruction yields comparable or better performance as well as comparable or increased learning efficiency. The studies culminated into a dramatic comparison between, “...a Beijing middle-school class judged to be of ‘average’ ability.” (Zhu and Simon, 1987, p.159), which implemented the worked example learning material, and three other control classes in session during the same time period. The experimental class completed a three year curriculum in only two years with comparable standard assessment test scores to three other classes that finished the same year on the standard three year curriculum.

In summary, several common themes arise from the above studies, which largely favor the use of worked examples, rather than conventional problem solving. Worked examples can be a more efficient use of learning time (Sweller and Cooper, 1985; Zhu, et al., 1987). With careful ordering of worked examples and problems (Sweller and Cooper, 1985; Trafton and Reiser, 1993), isomorphic problem solving performance can be enhanced (or at least not hindered) with the use of worked examples (Cooper and Sweller, 1987; Zhu, et al., 1987). Transfer effects are more apparent with less complex

domains (Cooper and Sweller, 1987). The impact of worked examples manifests itself with limited exposure (Cooper and Sweller, 1987), although the worked example effect does emerge with weaker background students given a longer exposure as well (Cooper and Sweller, 1987).

However, the above body of literature does not address the underlying questions of how and why do worked examples aid learning? What are students doing while examining worked examples? How are worked examples mentally processed? The next section sheds some light on these questions.

2.2: Observed Actions

This section discusses the research that sought to address the aforementioned question: What are learners doing when examining worked examples? In particular, what actions distinguish more skilled learners from less skilled learners while studying worked examples?

Typically, these studies trade-off a relatively small sample size (ranging from less than ten analyzed participants as with Chi, et al., 1989, to the order of fifty participants as with Stark, et al., 2002) to obtain a rich data set of verbal protocols. A rich data set allows for a detailed analysis of participant actions. Generally, the participants are introduced to relevant topics, and then given instructional material (including worked examples) to examine. Next, the participants complete a post-test. Verbal protocols are obtained as the students are examining the instructional materials and/or working on the post-test.

Transcripts of the protocols are subsequently segmented and coded. Much of the codes used in studies following the seminal research of Chi, et al. (1989) were influenced by the codes developed by Chi and her colleagues (for instance, see Pirolli and Recker, 1994; Renkl, 1997; Stark, et al., 2002). Finally, the codes can be quantitatively analyzed to augment a qualitative analysis of the protocol data.

In order to establish distinctions across participants, this class of research categorizes participants in two manners: good versus poor performers (as in Chi, et al., 1989 and 1991; Pirolli and Recker, 1994) and action profiles (as in Renkl, 1997; Stark, et al., 2002). A good or poor performer is sorted based upon post-test performance. Hence, distinctions of more productive (associated with the good performers) and less productive (associated with the poor performers) actions are gauged by the post-test. The other method of sorting participants for analysis resides in cluster analysis. Cluster analysis is a general term for a collection of methods⁹, which group together elements (in the case of this research, group together participants) sharing common characteristics. As a result, cluster analysis yields groups of participants exhibiting a similar action profile, whose post-test scores can then be compared.

In a sense, cluster analysis is the reverse of good versus poor performer comparisons. The analysis for the good versus poor performers begins with analyzing the post-test scores (to group the participants), and subsequently analyzes the actions to compare the two groups, whereas, cluster analysis begins with examining the actions (to

⁹ Renkl (1997) and Stark, et al. (2002) used the Ward method (Ward, 1963).

group the participants), and subsequently analyzing the post-test scores to compare the action profiles. In addition, the latter methodology allows for post-test performance and action comparisons prior to any grouping scheme (such as correlation measures as done in Renkl, 1997). Both methodologies have their merits. For instance, consider the most extreme case (in terms of sample size) of Chi, et al. (1989). Out of eight protocols used in the analysis, the top four post-test performers were categorized as good performers and the lower four were categorized as poor performers. Such a coarse segmenting (only two bins), is more apt to draw out larger trends, which was of interest with the early work (identifying distinctions between good and poor performers). Renkl (1997) expressed concern that participants on the lower end of the good performing group and on the upper end of the poor performing group may in reality be more comparable to each other, rather than to their other respective group members. Such concerns might be reduced by imposing a more substantial difference between the binned post-test scores of the good and poor performers (i.e., do not include the data of additional middle performers in the analysis). On the other hand, Renkl (1997) prefers the use of statistical measures allowing for analysis of fluctuations across the individual participant level, rather than a predetermined performance grouping (essentially, varying the analysis grain size).

2.2.1: Good versus Poor Performers

Chi, et al. (1989) recruited students with minimal physics background to participate in a study targeting the use of worked examples from chapter five of a physics

textbook (Halliday and Resnick, 1981). Chi, et al. (1989) used an elaborate training scheme necessitated by the amount of prerequisite background knowledge required to comprehend the target material. The participants were required to self-study the first three chapters of the textbook. At the end of each chapter, the participants were tested for relevant knowledge (and additional learning episodes were provided as needed to address deficiencies). Upon completing a self-study of the target chapter (chapter five), the participants were tested for declarative knowledge. In addition, they were given worked examples to examine (and concurrently provide verbal protocols). Finally, a post-test was administered. Such a research design has the advantage of controlling for participant background. Given the participants' limited prior exposure to physics, the researchers were better able to control prior physics domain training as opposed to recruiting participants with a more extensive background (who would in turn allow for a less extensive training routine to prepare for the material in the target chapter). The training scheme allowed the researchers to normalize (to a degree) the prerequisite knowledge basis of the participants, thereby, better isolating the impact of the target chapter's instructional material.

In particular, Chi, et al. (1989) was interested in characterizing the protocol statements of the good and poor performers. Overall, Chi, et al. (1989) characterized the content of the good performers' explanations as possessing a greater depth than the poor performers. The following responses of a poor and good learner are characterized as representative statements of their respective groups. While examining an equation ("T - $m_1g = m_1a$ ", Chi, et al., 1989, p.165) from a worked example, a poor performer

stated, "Okay, cause the acceleration is due to gravity." (Chi, et al., 1989, p.165) Aside from the questionable validity of the statement, there is minimal elaboration of the meaning underlying the given relationship. On the other hand, regarding the same step in the worked example a good performer noted, "Okay, so it's basically a way of adding them together and seeing if there is anything left over. And if there is anything left over, it equals the force: mass times acceleration." (Chi, et al., 1989, p.165) The good performer provides a fuller elaboration of the same equation including inherent connections to material covered in the Halliday and Resnick (1981) textbook.

Chi, et al. (1989) identified multiple types of statements from the protocols. Monitoring statements either acknowledge understanding (positive monitoring), "I can see now how they did it" (Chi, et al., 1989, p.161) or recognize a comprehension failure (negative monitoring), "I was having trouble with $F-mgsin\theta = 0$ " (Chi, et al., 1989, p.161). The good performers generated more negative monitoring statements than the poor performers, implying that the good performers were more apt to identify instances where they did not understand the worked example. Typically, monitoring statements would be followed by explanations. However, the poor performers only identified comprehension failures related to mathematical arguments, whereas the good performers also recognized conceptual comprehension failures. Chi and her colleagues propose that the good students are more apt to identify the issue at the heart of the failure versus the poor students. This contention is supported by the tendency of poor students posing more vague ("Well, what should you do here?" Chi, et al., 1989, p.171) or simplistic (ie, simply reiterating the example, "I was having trouble with $F-mgsin\theta = 0$." Chi, et al.,

1989, p.171) monitoring statements. In contrast, the good students tended to raise more targeted questions about their misunderstandings (“Why the force has to change?” Chi, et al., 1989, p.171). Chi and her colleagues (1989) conclude that, “... the Poor students seem oblivious to the fact that they do not understand, in part because they only have a superficial understanding of what they read.” (p.172)

Good performers also tended to apply relevant background information to construct explanations. In a prior paper, Chi, et al. (1982) delineated elements of Newton’s Laws. Chi, et al. (1989) identified and tallied the explanations following from the elements. The good performers referenced more elements than the poor performers. Chi, et al. (1989) do acknowledge that one may argue that the good learners produced more explanations based on physics principles simply due to a better understanding of the principles from the prior studying. However, they propose that the comparable results of the declarative testing following the self-study periods provide evidence against that possibility (although, such an issue cannot be fully discounted). Furthermore, good performers referenced more elements while examining worked examples not previously mentioned in prior testing. Chi, et al. (1989) suggests that the additional elements represent new knowledge obtained during worked example study. “Perhaps ...this is why the Good students feel so compelled to ‘explain’ while studying, since they are learning and encoding new knowledge.” (Chi, et al., 1989, p.167) However, beyond the prior assessment results, no further evidence is presented to refute definitively the possibility that the additional elements were encoded prior to the worked examples.

Another distinction of the good performers is the tendency of, “...relating consecutive example statements to each other.” (Chi, et al., 1989, p.166) In other words, the good performers tended to subgoal¹⁰. Indeed, Catrambone (1998) has proposed a Subgoal Learning Model, which can be distilled into three steps:

- “ 1. *A cue leads learners to group a set of steps.*
2. *After grouping the steps, learners are likely to try to self-explain why those steps go together.*
3. *The result of the self-explanation process is the formation of a goal that represents the purpose of that set of steps.*” (Catrambone, 1998, p.358)

However, why is the formation of subgoals useful? The literature details several advantages. Upon encountering a novel problem, a student learning from worked examples must search through the solution steps of the example and determine which steps must be modified to solve the novel problem. Subgoals reduce the search space (Catrambone, 1998). Rather than searching through individual steps, a student can begin a reduced search by subgoal instead. Once the student identifies the appropriate subgoal to modify, then the individual steps within the subgoal can be examined for appropriate modifications. Thus, the student avoids searching through all of the example’s steps individually. Similarly, if the student recognizes that an error has been made in the problem solution, the search space for the error can be reduced by searching through subgoals first. Once the subgoal with the error has been identified, then the subgoal’s individual steps can be examined to isolate the error (Eylon and Reif, 1984).

¹⁰ Assign a purpose or goal to step(s) within a worked example’s solution.

Subgoals promote a hierarchical knowledge structure, which has been associated with expert problem solvers (Larkin, et al., 1980). Dufrense, et al. (1992) promoted a hierarchically structured solution with a software tool called HAT (Hierarchical Analysis Tool). HAT requires the student initially to select a physical principle. Once the principle was selected, HAT would then proceed to more specific prompts further analyzing the principle in accordance with the problem under investigation until the required quantity is isolated and calculated. Dufrense, et al. (1992) report that enforcing a hierarchical solution structure promotes student recognition and application of problem deep structure as well as aiding schema formation.

Furthermore, Chi, et al. (1989) noted that the participant explanations discussed the application conditions of the subgoals. For instance, one participant responded to the selection of an axis along an inclined plane by stating, “and it is very, umm, wise to choose a reference frame that’s parallel to the incline, parallel and normal to the incline, because that way, you’ll only have to split up mg , the other forces are already, component vectors for you.” (Chi, et al., 1989, p.164)

Zhu, et al. (1996) advance a related argument on the importance of denoting both the subgoal and the corresponding conditions for applying the subgoal with a sample case from the field of algebra. “For example, in algebra, students are taught that they may add the same quantity to both sides of an equation (or subtract, multiply, or divide on both sides) without altering the values of the unknowns. This does not explain *when* to apply each action to solve an equation.” (p.1346) Zhu, et al. (1996) contend that worked examples illuminate the appropriate circumstances under which a given procedure may

be applied. Thus, not only must learners obtain procedural knowledge, knowing “how”, but they must also understand the appropriate application of the procedural knowledge, knowing “when” (cf., van Gog, et, al., 2004).

Chi, et al. (1989) also reported on how students used worked examples during problem solving. Generally, the poor performers tended to refer back to the worked examples more often than the good performers, thereby, supporting the notion that the good performers gained more from the initial study of the worked examples. In turn, the poor performers were more likely to simply reread the worked examples. Chi, et al. (1989) assert that the poor performers are seeking to map a solution to the problem (cf., Tuminaro and Redish, 2007) while the good students, who read fewer lines, have in mind a more specific piece of information to reference. Other data consistent with the distinction between good and poor performers include: starting point for the reading cases (the poor performers begin reading at the start of the example while good performers begin elsewhere in the example) and the questions posed while searching the examples (a good performer was noted as stating the specific goal of, “I’m looking at the formula here, trying to see how you solve for one (Force I) given the angle.” Chi, et al., 1989, p.175, while a poor performer declared a less focused question of, “What do they do? then proceeded to read the first line.” Chi, et al., 1989, p.175).

In a subsequent paper, Chi, et al. (1991) revisited two issues raised in the prior paper (Chi, et al., 1989). Prior to examining the worked examples, is the initial knowledge state of the good performers different from that of the poor performers? What

is the content of the good and poor performer self explanations? To address these questions, Chi, et al. (1991) reanalyzed data obtained from Chi, et al. (1989).

In Chi, et al. (1989) the researchers argued that the prior assessments in the training phase supported the claim that the good and poor performers possessed comparable knowledge states prior to studying the worked examples. Chi, et al. (1991) provides additional data supporting their claim. This data can be classified into participant background information and protocol patterns. Chi, et al. (1991) compared the participant GPA's and a spatial ability assessment. No statistical comparison treatment is provided; only the mean scores are compared (albeit, qualitatively the scores seem reasonably similar). Chi, et al. (1991) use another qualitatively based argument for the protocol patterns. Chi and her colleagues characterized the distribution of the statements referencing elements of Newton's Laws as being non-uniform during the testing in the training phase (that is, most statements reflected only a few of the elements), whereas if a complete understanding of the principles were attained prior to the worked examples, they would have expected a more uniform distribution across the elements. In addition, Chi et al. (1991) contend that the elements initially referenced in the worked example phase encompassed the conceptually based ideas (as opposed to the elements simply following from mathematical considerations). Furthermore, Chi and her colleagues supply examples of concepts noted in the protocols that are not explicitly provided in the textbook, but referenced in the worked examples. Thus, the researchers propose that the good performers attained some conceptual understanding through their self explanations while examining the worked examples.

Expanding on the explanation content analysis from Chi, et al. (1989), Chi and her colleagues (1991) analyzed knowledge constituents (“...propositions that are abstracted from each self-explanation.” Chi, et al., 1991, p.87). Overall, the good performers provided more knowledge constituents than the poor performers. The knowledge constituents were coded into four categories: systems, technical procedures, principle related, and concepts.

System related constituents embody information related to a mental model of the worked example’s situation. For instance, in reference to a worked example that involved two masses connected by a non-extendable string, a participant noted, “...since they’re connected by a string that doesn’t stretch, ummm, their accelerations are going to be opposite but equal.” (Chi, et al., 1991, p.89). Chi, et al. (1991) report that only 5% of the system constituents could be traced back to information from the textbook, thereby, leaving the bulk of the systems constituents stemming from the worked examples. Of course, the results do not preclude the possibility that prior experience plays a role in mental model elaboration as well.

Technical procedure constituents refer to procedural knowledge, “...such as choosing a reference frame and projecting force vectors onto the axes.” (Chi, et al., 1991, p.92) As with systems constituents, Chi and her colleagues (1991) contend that much of the procedural knowledge arises from the worked examples rather than the textbook. Similarly, LeFerve and Dixon (1986) present evidence that students prefer to learn procedural information from worked examples as opposed from written instructions.

Indeed, the technical procedure category encompasses the largest amount of constituents coded from the transcripts.

Principle related constituents encompass information pertaining to physics principles. Chi and her colleagues (1991) report that the principle related constituents are dominated by phrases that rearticulate Newton's Laws in various ways. For illustration, sample phrases for Newton's Second Law are provided:

"If the forces acting on a body do not sum to zero, then the body will move.

If a body is accelerating, then its net force must not be zero.

When a body has acceleration, then it must experience a net force.

If a body has a net force, then it is accelerating.

If a body is not at rest, then the net force will not equal zero (incorrect)." (Chi, et al., 1991, p.94)

By enunciating these various applications of Newton's Laws, Chi et al. (1991) propose that each variation becomes a separate rule for a specific application. Similar frameworks have been used in other contexts (cf., VanLehn, et al., 1992; Zhu, et al., 1996). Furthermore, Chi and her colleagues (1991) suggest that, "...once articulated, the knowledge no longer remains in tacit form." (p.95) Similarly, the phrases (or rules) conform to a conditional format, which may imply that the self explanations serve to

foster application of the principle in its mathematical form (for instance, $\bar{F}_{net} = \frac{d\bar{p}}{dt}$ for Newton's Second Law) to a given problem by denoting the conditions of the specific application of the principle to the problem (one might argue that such application information is parallel to the aforementioned subgoal application elaborations). However, due to the influence of the student's, "...naïve intuitions." (Chi, et al., 1991, p.95), Chi, et al. (1991) note that errors can be introduced through the integration of the knowledge constituents into the student's pre-existing mental framework (cf., Reif and Larkin, 1991). The common misconception of motion implying a non-zero net force (Clement, 1982; Halloun and Hestnes, 1985; diSessa, 1993; Reiner, et al., 2000) is provided as an example of such an error.

Concept constituents elaborate on a given quantity such as force. Chi, et al. (1991) contend that concept related constituents elucidate the properties (cases) of the respective concepts. For instance, while reading through a worked example, one participant arrived at the realization that weight is simply a type of force, and not a distinct quantity. The construction of a broader list of concept instances is suggested to increase the learner's flexibility in using the concepts. As an example, Chi, et al. (1991) referred to a transcript passage, where a participant recognizes that a point can be a body.

In a similar study, Pirolli and Recker (1994) analyzed statements recorded while students were learning concepts in LISP programming. Consistent with Chi, et al. (1989), the good performers provided significantly more explanations. In particular, the good

performers were associated with more meta-cognitive activities (monitoring, strategy¹¹, and activity¹²) while reading an introductory text to LISP¹³ along with more domain and strategy statements while studying the worked examples.

Pirolli and Recker (1994) further analyzed what type of LISP related (domain) elaborations distinguished the good performers from the poor performers. Collectively, the domain elaborations from the good performers tended to link the examples to external information. Thus, as noted in Chi, et al. (1989), the good performers are not viewing the worked example in isolation, rather they are generalizing by drawing upon a larger (more global) context. Good performers also tend to generalize through abstraction. Pirolli and Recker (1994) examined participant spontaneous statements made upon completing a post-test problem. The good performers tended to abstract the problem's solution by discussing the underlying facets, rather than simply restating the solution. Additionally, the good performers tended to compare and contrast the problem's solution against other solutions. Similar actions have been found to foster problem solving. Catrambone and Holyoak (1989) found that providing more relevant examples (across surface features), more learners were able to apply the underlying concepts to subsequent problem solving. Scaffolding explicitly promoting a comparison of the worked examples further enhanced performance. In a review of literature pertaining to acquiring cognitive abilities, VanLehn (1996) characterizes this effect as generalization. In particular, VanLehn (1996) is

¹¹ Statements that plan an action ("I'll look at an example; maybe that will help." Pirolli and Recker, 1994, p.257)

¹² Comments on the learning materials or learning task at hand.

¹³ A preliminary passage from a LISP textbook (Anderson, Corbett, and Reiser, 1987) was used to introduce the participants to LISP programming.

referring to the use of worked examples when solving problems by mapping solutions (cf., Tuminaro and Redish, 2007). If the learner is recalling the example solutions by deep structure rather than surface features, then in subsequent problem solving the student is, “...fooled less by surface features during mapping.” (VanLehn, 1996, p. 520) Thus, VanLehn (1996) argues a deep structure focus can be fostered by comparing multiple examples with the same deep structure but varying surface features.

Recall the characterization from Chi, et al. (1989) that poor performers provided more vague (less targeted) monitoring statements. Pirolli and Recker (1994) reported a similar phenomenon when characterizing self explanation prompts. Two self explanation prompts were identified: “... comprehension failures or self-imposed explanation goals.” (Pirolli and Recker, 1994, p.261) In addition to providing more overall self explanations, the good performers had more self explanations initiated in response to a self-imposed goal than the poor performers, which yet again suggests that the good performers exhibit more targeted actions in processing the instructional material.

Pirolli and Recker (1994) also propose a law of diminishing returns for self explanations. They acknowledge that the available data from prior work argues simply that more explanations are desirable. However, Pirolli and Recker (1994) question whether latter, excessive amounts of self explanation would continue to provide the same, increasing amount of benefits as earlier self explanation. “After all, how much can one really elaborate on a six-page text before one begins to become repetitive or stray from important concepts relevant to problem solving?” (Pirolli and Recker, 1994, p.264) To

test their conjecture, the researchers calculated curve fits (linear, power, and exponential) for number of errors while solving problems to the number of explanations while examining the worked examples. The power curve did exhibit the best fit. However, there are contradictory results. In a subsequent study also centered on self explanations from worked example study protocols, Renkl (1997) reported that his data does not support a diminishing return hypothesis. Although, Renkl (1997) argues that the ideas motivating Pirolli and Recker's (1994) hypothesis is based upon data obtained from a single person (intraindividual), whereas, both self explanation studies (Pirolli and Recker, 1994; Renkl, 1997) used data across several individuals (interindividual). Renkl (1997) conceded that an intraindividual law may indeed exist. "Perhaps the power law relation was not confirmed on the interindividual plane, because there were no subjects that explained the worked-out examples to themselves in a redundant, "over"-thoroughly manner so that additional explanations were of diminishing use." (Renkl, 1997, p.25) Hence, more research is required to confirm or falsify the diminishing return hypothesis.

2.2.2: Action Profiles

Another paper influenced by Chi, et al. (1989) was Renkl (1997). However, Renkl (1997) questioned whether time on task was a factor impacting the prior paper's results. In the Chi, et al. (1989) study, the good performers tended to study the worked examples over a longer time period, thereby, allowing time for more self explanations. In order to control the time participants studied the examples, the worked examples were provided

by a computer interface. Each worked example was incrementally displayed (i.e., rather than initially providing a complete solution, the learner would have to click on a button to see a new solution step until all steps were displayed). At the end of twenty-five minutes, the computer would cease to display worked examples. Think aloud protocols (Ericsson and Simon, 1993) were obtained during the twenty-five minutes of worked example study. Thus, depending on the study pace of the individual participant, the amount of worked examples examined per subject varied. In essence, Renkl (1997) exchanged normalized time on task for normalized amount of worked examples (as in Chi, et al., 1989; Pirolli and Recker, 1994). Indeed, Renkl (1997) did note an association between post-test transfer performance and number of worked examples examined. Such a confound is suggested by prior literature (Catrambone and Holyoak, 1989; VanLehn, 1996) indicating that examining more worked examples with common deep structures fosters subsequent problem solving. Another distinction from the work of Chi, et al. (1989) is the knowledge domain used. Rather than physics, Renkl (1997) focused on probability.

In contrast to the results from Chi, et al. (1989), Renkl (1997) reported that lower post-test scores were associated with increased instances of negative monitoring. Renkl (1997) argues that the effectiveness of negative monitoring depends upon the student's ability to resolve the comprehension failure. Hence, Renkl suggests that the instructional material available in his study may have been insufficient to aid the student in overcoming comprehension failures, thereby impacting post-test performance. Additionally, he proposed that the material covered in the examples might be less

conductive to the “... illusion of understanding” (Renkl, 1997, p.24), where learners casually accept a given result without critical evaluation of their own understanding (i.e., taking a result at “face value”). Ferguson-Hessler and de Jong (1990) reported a similar phenomenon with less skilled readers. They found that students who exhibited a less comprehensive knowledge structure from reading reference material tended not to question a result, but rather “taking [a result] for granted.” (Ferguson-Hessler and de Jong, 1990, p.47) Effectively, the “...illusion of understanding” (Renkl, 1997, p.24) represents the misuse of positive monitoring (affirming one’s understanding). In summary, “...the number of negative monitoring statements was more or less an indicator of difficulties in the learning process and not of effective metacognitive control. Taken together, many negative monitoring statements cannot be generally viewed as characteristic of good learners. Depending on context variables, such as the quality and difficulty of the instructional materials or the availability of help devices, negative monitoring statements may also be a characteristic of poor learners.” (Renkl, 1997, p.24) However, the discrepancy in negative monitoring results remains an open question to be addressed more definitively by further research.

On the other hand, Renkl (1997) confirmed other aforementioned results. Both principle related statements and subgoal statements were positively correlated to post-test performance. Furthermore, Renkl (1997) examined the associations between actions, and principle related explanations and subgoal explanations had the strongest (positive) correlation. Hence, Renkl (1997) concludes that such a correlation indicates, “... that learners who tended to assign meaning to operators did this in two ways: relating

operators to domain principles and to goals.” (p.14) In more general terms, as found with Chi, et al. (1989) and Pirolli, et al. (1994), participants with higher post-test scores tended to exhibit actions conductive to a more global view and a broader framework.

Additionally, given the weaker correlations spanning the other actions, Renkl (1997) argues that learners are not necessarily proficient in all types of explanations. That is, learners who exhibit a tendency toward one type of action may not necessarily exhibit tendencies toward other types of actions.

Thus, the question arises: are there any patterns of action combinations associated with the good performers? As previously discussed, Renkl (1997) performs a cluster analysis to determine action profiles of the participants. Two profiles standout with higher assessment scores. One such profile is consistent with the prior work discussed, with an emphasis on principle-based explanations and goal-operator connections.

However, the second profile has an emphasis on an action not previously discussed (anticipative reasoning), which arose as a useful happenstance due to the manner in which the worked examples were delivered. Recall that Renkl (1997) provided the solutions to worked examples on a step-by-step basis in order to control time on task. However, when given a solution one step at a time, some participants were observed to predict the next step. Indeed, anticipative reasoning was positively associated with post-test performance. Renkl (1997) suggests that the benefit of anticipative reasoning resides with the inherent monitoring aspect of making a prediction and subsequently verifying if

the prediction was correct. If the prediction does not hold true, then a comprehension failure is indicated (Markman, 1981).

Stark, et al. (2002) reanalyzed data from a prior study (Renkl, et al., 1998) which provided self explanation training to half of the participants in an effort to induce the self explanation effect raised in Chi, et al. (1989). Renkl, et al. (1998) succeeded in producing significant post-test gains with the trained group versus the control group, so Stark, et al. (2002) sought to characterize the self explanations of the participants. The trained group exhibited significantly more principle related explanations, subgoal explanations, and system explanations, which have previously been associated with good performers (Chi, et al., 1989 and 1991; Pirolli and Recker, 1994; Renkl, 1997). In addition, the trained group was associated with comparing solutions from multiple examples (cf., Catrambone and Holyoak, 1989; Pirolli and Recker, 1994). However, the meta-cognitive actions (e.g., monitoring) were not significantly weighted to the trained group. Stark, et al. (2002) argue that such a result is reasonable considering that the training focused only on the cognitive actions.

Following Renkl's (1997) analysis, Stark, et al. (2002) also produced action profiles from a cluster analysis. The two highest performing profiles emphasized either the cognitive actions from the training or meta-cognitive actions (positive and negative monitoring). Stark, et al., (2002) reported that the meta-cognitive participants were not significantly weighted to either the trained group or control group.

Overall, the meta-cognitive profile had the highest post-test performance on both isomorphic and transfer problems. The cognitive group yielded significant transfer performance gains as compared to the less elaborative participants, and in turn the meta-cognitive group significantly outperformed the cognitive group on transfer performance as well. The participants were asked to gauge their mental effort after each worked example. Only the meta-cognitive group reported a significantly higher mental load than the passive group. Stark, et al. (2002) conjecture that the additional mental effort of the meta-cognitive group may be a factor in the enhanced post-test performance that emerged across the training split. At the onset of the study, participants also completed a questionnaire (Dalbert, 1999 as cited in Stark, et al., 2002) gauging their tolerance of ambiguity. The meta-cognitive group had the highest tolerance of ambiguity rating. Stark and her colleagues (2002) suggest that the association of monitoring with a high ambiguity tolerance implies that those with a high tolerance are not intimidated by the prospect of comprehension failures; rather, they identify and attempt to reconcile the failure. On the other hand, those with lower tolerance shield themselves from such inconsistencies by a more superficial analysis of the example's solution; thereby, making the "illusion of understanding" possible. They conjecture that the learner might be cognizant of deficiencies, and the superficial analysis might be a defensive mechanism.

In summary, the research in sections 2.1 and 2.2 characterize actions of participants learning from worked examples. The actions can be grouped into four overarching threads: monitoring, application of relevant background knowledge, extraction of information, and prediction.

Monitoring encompasses both the affirmation of understanding (positive monitoring) and the acknowledgement of a comprehension failure (negative monitoring). Two factors have been suggested which influence the utility of monitoring while examining worked examples: the quality of the monitoring statements and availability of resources to resolve comprehension failures. Negative monitoring statements must extend beyond a surface question and strike at the heart of the misunderstanding (Chi, et al. 1989). Once the specific issue of the comprehension failure is identified, then there must exist resources (whether internal or external) to allow for resolution of the disconnection (Chi, et al., 1989; Renkl, 1997). Furthermore, positive monitoring statements must exist at a deeper level to avoid the “illusion of understanding” effect as termed in Renkl (1997) or as Ferguson-Hessler and de Jong (1990) noted of less skilled learners, “taking for granted” (p. 47) a given result.

In conjunction with the reading comprehension literature, worked example research denotes several comprehension failure triggers¹⁴. A noted internal or external inconsistency is an indication of a comprehension failure (Markman, 1981). An internal inconsistency refers to seemingly incompatible statements within a single worked example (according to the perception of the learner). Assuming that the worked example is written correctly, any inconsistencies within the solution is due to a flawed understanding of the solution. On the other hand, external inconsistencies refer to incompatible information from the worked example versus information obtained from an

¹⁴ By comprehension trigger, I refer to a mechanism by which the presence of a comprehension failure is recognized, whereas the negative monitoring statement is the acknowledgement of a comprehension failure.

external resource such as previous lectures, textbook readings, etc. Hence, again assuming that the worked example and external material do not contain errors, the inconsistency resides within the learner's understanding of the material. If a statement allows for multiple interpretations without a clear indication of the correct meaning according to the learner's perception (i.e., the statement is seemingly ambiguous, see Markman, 1981), then the learner's understanding is incomplete (assuming that the flaw does not lie with the worked example). A third comprehension failure trigger denoted in Markman (1981) is a prediction failure. If the learner is anticipating a specific result (for instance, see Renkl's, 1997, anticipative reasoners) but a different result occurs, then the learner is alerted to a flaw in their perception of the material. The final comprehension failure trigger is an inability to understand how a given step (or steps) within a solution of a worked example follows from the prior step(s) (i.e., a derivation standstill, VanLehn, et al., 1992).

Throughout the literature, higher post-test performance is associated with learners applying relevant background knowledge to the instructional material (Chi, et al, 1989 and 1991; Pirolli and Recker, 1994; Renkl, 1997; Stark, et al., 2002). In particular, good performers tend to invoke relevant information to interpret and expand upon the information presented in worked examples as opposed to producing superficial explanations of the examples (i.e., not connecting to a more significant context). In essence, the worked examples do not exist in isolation, rather they are part of a larger “picture” (or a global context).

Extracted knowledge refers to the inference or abstraction of information from a worked example. In particular, good performers extract the following types of information:

1. Procedural knowledge (Catrambone and Holyoak, 1989; Chi, et al., 1991; Pirolli and Recker, 1994)
2. Subgoals, including when they can be used (Chi, et. al., 1989 and 1991; Zhu, et al., 1996; Renkl, 1997; Stark, et al., 2002)
3. Concept related knowledge (Chi, et al., 1991)
4. System related knowledge (Chi, et al. 1991; Stark, et al., 2002)
5. Principle related knowledge (Chi, et al., 1989 and 1991).

Such generalization provides a broader knowledge base for subsequent problem solving.

Prediction (Renkl, 1997) integrates two of the above actions: inference (extracted knowledge) and monitoring. Initially, the learner must propose a (inferred) result, and then the scaffolding present in the worked example can provide feedback informing the learner if the prediction was correct (monitoring). Similar to an argument posed in Renkl, Atkinson, Maier, and Staley (2002), one might suggest that the mechanisms underlying the creation of a prediction are related to those with problem solving (i.e., similar abilities are required in both cases in order to generate a subsequent, unknown step). Such an argument suggests that prediction provides practice in exercising relevant skills.

Resulting from a review of the literature on reading comprehension, Palincsar and Brown (1984) identified six actions that are at the root of comprehension skills and strategies. “*(1) understanding the purposes of reading, both explicit and implicit; (2) activating relevant background knowledge; (3) allocating attention so that concentration can be focused on the major content at the expense of trivia; (4) critical evaluation of content for internal consistency, and compatibility with prior knowledge and common sense; (5) monitoring ongoing activities to see if comprehension is occurring, by engaging in such activities as periodic review and self interrogation; and (6) drawing and testing inferences of many kinds, including interpretations, predictions, and conclusions.*” (Palincsar and Brown, 1984, p.120)

These factors¹⁵ are strikingly similar to the actions described above (monitoring, application of relevant background knowledge, extraction of information, and prediction).

¹⁵ The first factor can be viewed in terms of epistemology (in the case of worked examples, the inherent goals of examining the examples). Such information is largely probed indirectly since people typically do not explicitly reflect directly upon their individual epistemology (as argued in Hammer, 1994). There are studies that provide activities of students while examining examples that might potentially be used to infer goals. For instance, Tuminaro and Redish (2007) report that a problem solving technique that students were observed to use is to identify example solutions that contain quantities denoted in the problem they are solving, and apply the solution to solve the problem (i.e, directly mapping the example’s solution to the problem at hand). However, I shall concentrate upon the remaining factors that avoid such inferences and provide for direct connections.

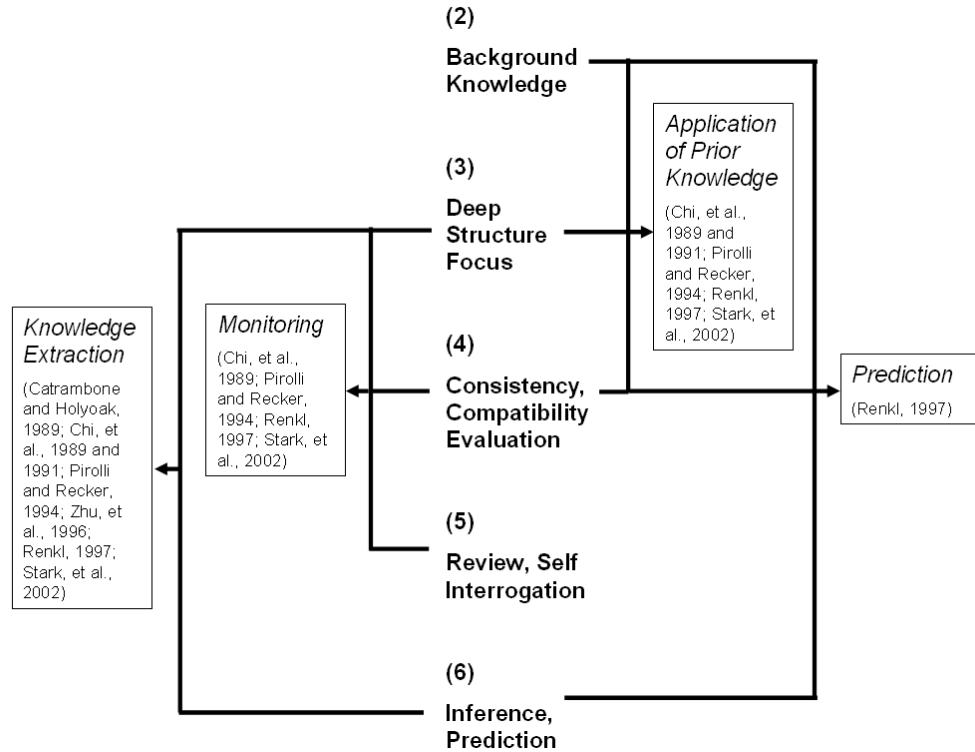


Figure 2.1 – Underlying connections between Palincsar and Brown (1984) factors and worked example study actions (the factor number appears in parentheses)

Successful monitoring requires the integration of several factors. Periodic self interrogation is at the core of monitoring (Chi, et al., 1989; Pirolli and Recker, 1994; Stark, et al., 2002). Both positive and negative monitoring consists of comparing one's knowledge against information contained in instructional material (includes checking for compatibility and consistency). Without careful (deep) periodic checks on one's understanding, then the danger of missing comprehension failures exists (Chi, et al., 1989). To be able to generate the deep monitoring statements, one must focus upon the key aspects (deep structure) and de-emphasize the superficial matters (Chi, et al., 1989).

The most direct link of prior knowledge application to the Palincsar and Brown (1984) framework is the factor of invoking pertinent background knowledge. Skilled learners tend to address worked examples in a global fashion, whereas less skilled learners have a more restricted view (Chi, et al., 1989 and 1991; Pirolli and Recker, 1994; Renkl, 1997; Stark, et al., 2002). By linking the examples to a broader context, a more general understanding can emerge (for instance, see the aforementioned sample dialogs from Chi, et al., 1989, that are noted as being typical for good and poor performers). Inherent to drawing more general connections to external information is the factor of attending the deep structure (Catrambone and Holyoak, 1989; Pirolli and Recker, 1994; Catrambone, 1998). Finally, as Markman (1981) argues, a coherent structure is required for understanding. Thus, when integrating information from worked examples into one's mental structure, the factor of verifying compatibility between existing knowledge and information from the worked example is needed to foster comprehension.

At the heart of knowledge extraction is the factor of inference. Successful knowledge extraction fosters generalization through abstraction (Pirolli and Recker, 1994), deep structure focus (Catrambone and Holyoak, 1989; Catrambone, 1998), and broadening of relevant background knowledge (Chi, et al., 1989 and 1991). As part of broadening background knowledge and subgoaling is the exploration of both the conditions of knowledge (Chi, et al., 1989 and 1991) and subgoal application (Chi, et al., 1989 and 1991; Renkl, 1997; Catrambone, 1998; Stark, et al., 2002) along with the respective consequences (i.e, generating conclusions, a factor from Palincsar and Brown, 1984).

Finally, prediction (Renkl, 1997) follows directly from Palincsar and Brown's (1984) creating and testing inferences and predictions. Inherent to creating the predictions is the use of relevant background knowledge, while verifying the predictions follows from checking one's knowledge for compatibility and consistency.

Much of the literature (for instance, see relevant literature overviews by Atkinson, Derry, Renkl, and Wortham, 2000; Renkl and Atkinson, 2002) describes the above actions through an overarching concept stemming from Chi, et al. (1989): the self explanation effect. "Self-explaining (i.e., generating new knowledge), whether prompted or spontaneous ... is a process of reflection. Students are encouraged to reflect, think about, and infer what they are reading ... the process of generating explanations is adding inferences and constructing a mental structure." (Chi, et al., 1994, p.459) Chi, et al. (1994) propose three aspects of self explanations which foster learning. These aspects follow directly from the factors denoted in Palincsar and Brown (1984): self explanation, "... is a constructive activity." (Chi, et al., 1994, p.470) (factor 6), self explanations incorporate extracted information into pre-existing knowledge (factor 4), and "... self-explaining is carried out in a continuous (sic), ongoing, and piecemeal fashion ..." (Chi, et al., 1994, p.472) (factor 5).

The above connections between reading comprehension and learning from worked examples suggest that both activities (perhaps in part) share underlying components. Essentially, one might argue the both reading comprehension and learning

from worked examples involve the recognition, interpretation, and encoding of information for subsequent use.

Despite the reasonably robust understanding of actions invoked by worked example study, there is one area that has not been as fully addressed: the integration of incorrect solutions as part of the learning material. All of the above research has been focused upon the use of correct solutions. Fundamentally, the worked examples in the above literature are comprised of a problem statement and a correct solution. However, limited results are available for learning materials that include incorrect solutions, particularly within the realm of physics. Such an issue of is of particular concern given the rise of reforms that do integrate incorrect solutions (several examples are WRONG Problems in Harper, et al., 2007; Evaluation Tasks in Etkina, et al., 2006; PALs in Reif and Scott, 1999), and yet there is a limited understanding of what processing is actually invoked in response to the inclusion of incorrect solutions. Große and Renkl (2007) is a notable exception. Reassuringly, Große and Renkl (2007) found similar actions to the worked example literature when the participants were also given incorrect solutions to study. However, the study took place outside the realm of physics, and offers only a rare instance of the processing of incorrect solutions, unlike the more proven and robust frameworks available for the processing of correct solutions only.

2.3: Modified Worked Examples

This section discusses research focused on enhancing the impact of worked examples as instructional material. The overarching research question is: will modified worked examples yield superior problem solving performance when compared to unmodified worked examples?

These type of studies typically divide participants into experimental group(s) (using the modified worked examples) and a control group (using unmodified worked examples). Subsequently, the participants complete a post-test to gauge the worked example impact on problem solving. As with much of the aforementioned work, these studies generally sacrifice ecological validity for laboratory controlled environmental conditions, which aid in damping the impact of external confounds to the utility of the worked examples.

Generally this research falls into one of three categories: action leveraged, memory model, and transition implementations. Action leveraged modifications sought to enhance worked examples through modifications intended to promote productive actions identified in section 2.2 (for example, see Catrambone, 1995, 1996, and 1998; Renkl, et al., 1998; Stark, 1999; Conati, et al., 2000; Derry, et al., 2000). Memory model implementations sought to optimize worked examples to account for the limitations of human memory (Tarmizi and Sweller, 1988; Ward and Sweller, 1990; Mousavi, et al., 1995; Gerjets, et al., 2004). Finally, transition implementations were intended to ease the

transition from studying worked examples to solving problems (Renkl, Atkinson, Maier, and Staley, 2002).

2.3.1: Action Leveraged Modifications

Catrambone (1995, 1996, and 1998) performed a series of experiments centered on developing worked examples largely based upon the action of subgoaling (such experiments lead to the aforementioned Subgoal Learning Model from Catrambone, 1998). These experiments focused on worked examples and problems using the Poisson distribution to predict the probability of getting a given number of “successes” (for example, the probability of a randomly chosen lawyer owning 4 suitcases). These types of problems typically have relatively basic subgoal structures as compared to typical physics problems, including those in college level introductory courses.

In order to foster the subgoaling action, Catrambone (1995, 1996, and 1998) modified worked example solutions to label the subgoals. Overall, Catrambone (1995) found that labeling the subgoals did enhance transfer performance on the post-test. Furthermore, he optimized the labeling effect through testing several variations. Irrespective of whether the labeled subgoal was placed on its own separate line or if the labeled subgoal was placed directly within calculations, transfer performance was enhanced (Catrambone 1996). Arising from a concern as to whether the content of the labels provided additional information to the subgoal cued groups, thereby yielding the superior performance, Catrambone (1996 and 1998) also included less meaningful

(abstract) labels (rather than the labels providing an explicit description of the subgoal, a Greek symbol was used to label the subgoal). The abstract label group exhibited superior performance when tasked with solving role reversed problems (Catrambone, 1998). Role reversed problems possessed a different surface structure than the worked examples. Rather than computing objects per person (such as computers per information technology expert), the reversed role problems involved computing people per object (such as viewers per television program). Catrambone (1998) hypothesized that the abstract labels yielded a looser surface feature connection to the deeper subgoals, allowing for superior performance with the reversed role problems.

Furthermore, Catrambone (1998) studied the effect of student (relevant) background on cued subgoal performance. Participants with two different backgrounds in mathematics were recruited for the study: either no university calculus coursework experience (categorized as a weaker background) or two to four university semesters of calculus subsequent to the introductory calculus courses (categorized as a stronger background). For the stronger background group, the label conditions (meaningful and abstract) outperformed the no label conditions. However, only the meaningful label group outperformed the no label group for the weaker background condition.

Hence, throughout the above experiments, cueing the subgoal structure tended to yield superior performance in solving transfer problems. Indeed, in a review paper of the cognitive tutor development at Carnegie Mellon University, Anderson, et al. (1995) denote the importance of imparting the subgoal structure of solutions to students as one

of the driving tutor design principles. To a degree, the Catrambone's (1998) data suggests that those with a weaker relevant background require the contextual information in the superficial cues to construct transferable subgoal structures. However, the superficial label results are tempered by the abstract labels yielding improved performance with transfer problems (at least for students with an appropriate background) involving varied surface feature structures (implying a broader generalization capability with the abstract labels). Catrambone (1998) conjectures, "Labels tied to surface features might encourage the learner to take a short-cut to explain the steps' purpose (i.e., the label almost becomes the explanation)..." (p. 364)

Atkinson and Derry (2000) combined ideas from Catrambone's work (1995, 1996, and 1998) on subgoaling with Renkl's (1997) anticipative reasoning (by presenting the worked example one step at a time). Rather than written cues (labels) as used in Catrambone's work, Atkinson and Derry (2000) used temporal cueing (depicting one subgoal at a time in sequential order). Such worked examples were noted as: "...sequential/subgoal-oriented (SE/SO)..." (Atkinson and Derry, 2000, p.132), while the conventional worked examples (solution shown in entirety without subgoal cues) were called, "...simultaneous/non-subgoal-oriented (SI/NS)..." (Atkinson and Derry, 2000, p.132). Furthermore, Atkinson and Derry (2000) included a split to study the use of worked examples prior to the post-test (the SE/SO-absent group and the SI/NS-absent group) versus allowing access to the worked examples during the post-test (the SE/SO-present group and the SI/NS-present group). Atkinson and Derry (2000) reported that the SE/SO groups' protocols indicated a stronger conceptual understanding than the SI/NS

groups. Unfortunately, no discussion was provided in the specific measures to gauge conceptual understanding. When the subjects were polled on whether they understood the worked examples, more SE/SO participants noted that they did understand the worked examples than the SI/NS participants. Within the SE/SO group, the present group provided more conceptually correct solutions to the problems than the absent group. However, when polled on understanding the examples, the absent group exhibited a higher confidence in understanding the worked examples.

Following from Renkl's (1997) anticipative reasoning results, Stark (1999) tested the use of incomplete worked examples. Stark (1999) recruited education students to learn from worked examples and work problems on probability topics. Much like Renkl's (1997) study, the worked examples were presented in a step by step manner. However, the incomplete worked examples initially replaced the result of each displayed step with a question mark (the learner was asked to state what the result was), whereas the complete worked example group received the result at the onset (no question mark). The post-test consisted of three types of transfer problems. Close transfer problems had the same deep structure but a different surface structure as the worked examples, while the problems noted as middle transfer had a different deep structure with the same surface structure. Finally, far transfer changed both the deep and surface structures. The incomplete worked example group outperformed the complete worked example group for both the close and middle transfer problems.

Chi, et al. (1994) reported that self explanation training for participants reading instructional passages on the circulatory system yielded higher learning gains versus

untrained participants. Renkl, et al. (1998) extended self explanation training to the examination of worked examples across surface feature variations (cf., Catrambone and Holyoak, 1989). Fifty six banking apprentices were divided across two splits (into four groups total): worked examples with constant surface features (securities only) versus varied surface features (shares, securities, and loans), and self explanation training versus no self explanation training.

For learners with a weaker relevant background, the self explanation training yielded superior isomorphic performance. For transfer performance, Renkl, et al. (1998) predicted and confirmed the following conditions from highest to lowest performance: “... multiple/elicited, uniform/elicited, uniform/spontaneous, multiple/spontaneous” (Renkl, et al., 1998, p.102), where multiple refers to varying the worked examples’ surface features (conversely, uniform refers to constant surface features) and elicited refers to self explanation training (while spontaneous refers to no training).

Self explanation elicitation has not been limited to worked examples. Conati and VanLehn (2000) prompted self explanations with the addition of a module called Self Explanation Coach (or SE-Coach) to the computer-based tutoring system, ANDES (Gertner and VanLehn, 2000). As the student uses the tutor, SE Coach generates a model of the student’s state of understanding. Based upon the model, SE Coach prompts the student to submit self explanations.

Conati and VanLehn (2000) reported mixed results. Participants were recruited from four colleges. For two of the colleges, the full SE Coach module yielded higher performance, while the other two colleges had no such gains. Indeed, the control group of

the no gain colleges outperformed the control group of the gain colleges and performed nearly as well as the experimental groups from the gain colleges. However, the no gain colleges began their semester a week earlier than the gain colleges. Conati and VanLehn (2000) suggest that the head start of the no gain groups implied that they were at a later learning stage, thereby impacting their dependence on the worked examples.

“• Rich scaffolding for self-explanation, like the one provided by the complete SE-COACH in the experimental condition, can improve students’ performance at an early learning stage. At this stage, students are still unfamiliar with the subject matter. Hence, they benefit more from structured help in using domain knowledge to generate effective self explanations and are more motivated to put substantial effort in exploiting this help at best.

• As students become more proficient in the subject matter, even minimal prompting, like that provided by the masking interface in the control condition, can help improve their self-explanations. At this stage, more elaborate scaffolding can actually be less effective, if it requires students to put too much effort in studying examples, because they may lack the motivation to do so.” (Conati and VanLehn, 2000, p.411)

Conati and VanLehn (2000) do stress that the above inferences require additional work to verify. Such inferences may represent an analog to Pirolli and Recker’s (1994) suggested law of diminishing returns: as a learner’s experience increases, the utility of scaffolding decreases.

2.3.2: Memory Model Inspired Modifications

Working memory (for example, see Baddeley, 2000 and 2003) is a model used in psychology of active, mental processing of information in a finite capacity (earlier work characterized such a capacity in terms of information chunks¹⁶). “Humans are conscious of and can monitor only the contents of working memory. All other cognitive functioning is hidden from view unless and until it can be brought into working memory.” (Sweller, et. al., 1998, p.252)

Sweller, et al. (1998) have attributed several effects to memory limitations. Sweller and his colleagues label their characterization of such effects as Cognitive Load Theory (see Sweller, et al., 1998 and van Merriënboer, et al., 2005, for overviews). Several studies on worked example design and implementation have referenced their characterization (for instance, see Tarmizi and Sweller, 1988; Ward and Sweller, 1990; Mousavi, et al., 1995; Gerjets, et al., 2004).

¹⁶ See Miller (1956) for an introduction to the concept of information chunks, or as described by Miller, a single familiar element comprised of a set of information. Hence, chunks can vary in size (i.e., the amount of information within a chunk). Miller (1956) illustrated such size variations through an example of someone learning to translate Morse code from dots and dashes to phrases. Initially, the person’s chunks are small (consisting of individual dots and dashes). With practice, the person is able to resolve the dots and dashes into letters; thereby expanding the chunks to groups of dots and dashes representing a single letter. Next, the person expands the chunks from individual letters to words. Finally, the person connects the words to form phrases (yet again extending the chunks to encompass multiple words).

2.3.2.1: Integrated Worked Examples

Integrated worked examples seek to mitigate the split attention effect (Tarmizi and Sweller, 1988; Ward and Sweller, 1990; Sweller, et al., 1998, van Merriënboer, et al., 2005). Below is a worked example that follows the format of the integrated worked example design.

A motorcycle traveling in one direction changes its speed from 15 m/s (u) to 20 m/s (v) with an acceleration of 0.5 m/s^2 (a): $[v = u + at, t = (v - u)/a = (20 - 15)/0.5 = 10 \text{ s}]$. How long does it take for the motorcycle to change its speed?

Rather than spreading a learner's attention to multiple regions of information, integrated worked examples allow the learner to focus attention in a more concise manner; thereby, "... reducing the need for mental integration." (Sweller, et al., 1998, p.279) In order to focus the learner's attention to a smaller region, the integrated examples included the mathematical representation interlaced within the prose of the problem along with the solution, as opposed using a more typical format of separating the prose of the problem statement from the solution.

Hence, Sweller, et al. (1998) contend that more mental capacity is freed for learning. Or as stated by Ward and Sweller (1990), "This process of integration will require cognitive resources to be devoted to activities that, at best, are marginal to schema acquisition and rule automation." (p. 18)

Tarmizi and Sweller (1988) tested integrated geometry worked examples against conventional problem solving and non-integrated worked examples. Reminiscent of

Renkl (1997), participants had a fixed time study time. The participants were divided into three groups: integrated worked examples, non-integrated worked examples, and conventional problem solving. During the study time, the worked example groups were given pairs consisting of a worked example and a related practice problem, while the conventional problem solving group was provided with pairs of related problems.

The integrated worked example group examined the most worked examples and problems during the fixed study time. Tarmizi and Sweller (1988) argue that the increased amount of processed problems for the integrated worked example group implies a lower cognitive loading state for the integrated worked examples. Furthermore, the integrated example group was generally in the minority for use of a backward reasoning strategy (rather the integrated example group generally opted for a forward reasoning strategy). Tarmizi and Sweller (1988) contend that the integrated worked example group's bias toward forward reasoning supports the view that the integrated worked example format fosters schema development, whereas others were forced to fall back upon mean-ends style analyses.

The integrated example group also generally required less time to arrive at their solutions, leading to a significant reduction on total post-test time for the integrated worked example group. Indeed, the reduction in time and use of forward strategy was correlated.

Tarmizi and Sweller (1988) summarize their findings by noting that in order for scaffolding (from worked examples) to be maximally effective, cognitive loading must be

minimized. “We can conclude that the critical factor is not the surface format of the problem and its presentation but rather the deeper, cognitive implications of its presentation format.” (Tarmizi and Sweller, 1988, p.435).

Building off the above integrated worked example results for the subject of geometry, Ward and Sweller (1990) extend integrated worked examples to pre-college level physics, which is suggested as another area susceptible to the split attention effect.

Ward and Sweller (1990) began with a set of experiments that tested the use of non-integrated worked examples against conventional problem solving (similar to the research covered in section 2.1). However, the experiments were performed as part of a high school science course. The general research design consisted of classroom instruction, a homework assignment (which was varied in the experiments), and a subsequent test during the following class meeting.

For the initial experiments, the participants were divided into two groups. One group (worked example group) received homework consisting of non-integrated worked examples paired with practice problems. The other group (conventional problem solving) was provided with homework only consisting of paired problems. The studies covered topics in optics and kinematics (along with a transfer problem in collisions). For the optics topics, the worked example groups exhibited superior isomorphic performance. However, similar gains were not noted with the kinematics topics. Ward and Sweller (1990) suggest that the kinematics worked examples typically divide attention between

text and mathematical relationships. Hence, Ward and Sweller (1990) propose that such worked examples are worthy candidates for an integrated format.

The next stage of research divided the participants into three groups: integrated worked example, non-integrated worked example, and conventional problem solving. As with the prior studies, the worked example groups received homework consisting of pairs of worked examples and related problems, while the conventional problem solving group had pairs of problems to solve. The integrated worked example group solved more problems with less errors than the other two groups.

Ward and Sweller (1990) argue that the integrated worked examples freed cognitive resources for learning resulting in improved performance beyond the conventional problem solving group and the non-integrated worked example group. Thus, as with the conclusions from Tarmizi and Sweller (1988), when worked examples are used, their design should be optimized for cognitive load in order to maximize instructional benefits. In particular, such studies suggest leveraging the split attention effect by reducing the spread of a learner's attention (i.e., reducing the need for mental integration).

2.3.2.2: Modular Worked Examples

Intrinsic cognitive loading is a characterization from Sweller, et al. (1998) that encompasses the complexity of the situation to be processed. This degree of complexity is dependent upon the level of experience a person possesses. For example, when a

novice chess player examines the location of the chess pieces in the middle of a match, the novice may consider each piece as a separate element, while an expert may see the pieces as part of a connected whole. As described in a paper by de Groot (1966), both novice and expert chess players were shown a chess board in the middle of a game and were subsequently asked to recall the locations of the pieces. The experts exhibited superior performance in recalling the chess piece locations.¹⁷ Due to the expert's additional knowledge and experience, the chess board appears less complex versus a novice's perspective of the pieces being separate, non-related elements¹⁸.

In contrast to intrinsic loading, Sweller, et al. (1998) characterize two other loading types: germane and extraneous. Extraneous loading refers to the processing of information not relevant to learning, and so such loading is taking up space in working memory which might be used instead to process information pertinent to learning (germane loading).

Gerjets, et. al. (2004) describe instructional techniques that have been proposed to reduce intrinsic loading.

“Part-whole sequencing...” (Gerjets, et al., 2004, p.41) leverages the subgoal action described in section 2.2. Learning begins with training on performing the subgoals. Once the students have gained experience with the subgoals, then the subgoals can be

¹⁷ When the pieces were randomly placed, the expert exhibited comparable recall performance to the novices (Chase and Simon, 1973).

¹⁸ Expanded expert chunking is not isolated to chess. For instance, Abernethy, et al. (1994) describe similar results with expert and novice snooker players. For more discussion, see Simon and Gobet (2000) for an overview of related work.

pieced together for more complex tasks. Thus, when the complex tasks are introduced to the students, then they will have already developed relevant background knowledge (much like the aforementioned chess experts). The complex task now merely becomes a matter of connecting previously learned subgoals together. Hence, intrinsic loading lowered as compared to those who skipped the initial subgoal training and proceeded directly to the complex tasks.

“Simplified whole tasks...” (Gerjets, et al., 2004, p.41) in contrast seeks to begin training with a simplification of the complex task, and gradually remove the simplifications until the full complexity of the task is reached. Hence, the task complexity manipulation, again allowing for relevant background knowledge generation, seeks to reduce intrinsic loading.

The above techniques spawned an application to worked examples, which Gerjets, et al. (2004) named modular worked examples. A modular worked example presents a solution path highlighting, “...smaller meaningful solution elements and their relation to individual structural task features” (Gerjets, et al., 2004, p.42), rather than a focus on, “...problem categories, clusters of structural task features, and category specific solution procedures.” (Gerjets, et al., 2004, p.42).

The modular group outperformed the molar group for both isomorphic and transfer problems, which had a modified deep structure requiring the multiplication of, “... two complex event probabilities ...” (Gerjets, et al., 2004, p. 47)

Gerjets, et al. (2004) propose that molar worked examples focus upon providing a solution to a specific class of problems, thereby, forcing learners to classify first a problem to determine the correct solution path to take. Therefore, the question arises: will the modular format degrade the ability of learners to recognize types of problems?

Gerjets, et al. (2004) tasked participants with matching worked examples to problems that share, "... structural features." (Gerjets, et al., 2004, p.49). In addition, a comparison task that required the subjects to identify, "... structural similarities and differences" (Gerjets, et al., 2004, p.49) of problems. No differences on the classification results were reported. Only prior knowledge was identified as a factor affecting the comparison task. In line with the self explanation effect (specifically with the knowledge extraction action), Gerjets, et al. (2004) suggest that students learning from modular worked examples obtain the classification information through inference unlike the molar worked examples which directly provide the classification information. From the perspective of Cognitive Load Theory, the molar worked examples encourage a higher loading state by including, "...problem categories, clusters of structural task features, and category specific solution procedures." (Gerjets, et al., 2004, p. 43) Gerjets, et al. (2004) claim that such a classification approach forces a student to load working memory with numerous structural aspects in order to determine which solution classification is appropriate for use in solving a given problem. Thus, less working memory resources are available for additional germane loading. Hence, Gerjets, et al. (2004) argue that the modular examples free additional cognitive resources for additional germane loading such as generating self explanations. In contrast, one might argue that in both worked

example formats, the structural information is in memory (whether inferred or directly obtained). Hence, an alternate hypothesis suggests that, given the aforementioned benefits of the inference processes, subsequent problem solving benefits. However, more research is required to address definitively the hypotheses.

Another phase of the research by Gerjets, et al. (2004) suggests that the modular format does provide a lower loading state. To gauge cognitive loading, the subjects were also required to complete a questionnaire based on the NASA-TLX (Hart, et al., 1988). In addition, the researchers varied the amount of elaboration in the worked examples, which was considered possibly to impact cognitive loading. However, Gerjets, et al. (2004) report no significant influence due to elaboration levels, but cognitive loading generally favored the modular worked examples. Also, the modular worked example group outperformed the molar worked example group on the post-test.

2.3.2.3: Multimode Worked Examples

The model of working memory contains separate processors for visual-spatial (visuospatial sketchpad) information and audio (phonological loop) information (Baddeley, 2000 and 2003). There is evidence that these processors can perform their functions in parallel (Penny, 1989). Thus, multimode worked examples seek to increase the information throughput by distributing information across both processors rather than a single processor (Mousavi, et al., 1995).

Targeting geometry concepts, Mousavi, et al. (1995) performed a series of experiments to confirm and optimize the multimode effect. Initially, two methods of multimode information conveyance were tested against a control group (visual information only): concurrent (providing information visually and aurally simultaneously) and complementary (providing some information visually and other information aurally). As with studies from Sweller and his colleagues (Sweller and Cooper, 1985; Cooper and Sweller, 1987; Tarmizi and Swelller, 1988; Ward and Sweller, 1990), the instructional material consisted of worked example-problem pairs.

The multimode groups required less time to complete the practice problems paired with the worked examples as well as the isomorphic post-test problems versus the control group (no transfer effects were noted).

However, when the time to study the worked examples was fixed, only the complementary group required less time to complete the isomorphic post-test problems. There were no differences in solution times of the practice problems. Thus, when the worked example times were equalized, the post-test differences remained while the acquisition stage differences disappeared. One might question if there is a comparable issue reminiscent to the proximity effect as observed in the studies on worked example and problem ordering (Sweller and Cooper, 1985; Trafton and Reiser, 1993).

Furthermore, there were no effects due to presentation ordering (whether information was fully presented at the onset or if information was incrementally presented, similar to Renkl, 1997). However, the use of diagrams versus written passages

to convey the worked example's situation was found to reduce solution times for both isomorphic and transfer problems.

Mousavi, et al. (1995) proposed a possible confound to the above results. What if the aural information is inherently faster to process than visual information (recall that the control group was visual information only)? Thus, the mixed mode results simply may be due to reduced processing time for aural information. To address this concern, Mousavi, et al. (1995) tested a visual only format against an aural only format, and the only difference they found was that the visual group required less time to solve the practice problems (arguing against the confound).

In summary, Mousavi, et al. (1995) provide evidence that mixed mode (visual and auditory) presentations can yield an advantage over single mode presentations. Such an effect is consistent with Baddeley's (2000 and 2003) model of working memory that includes separate processors for each information mode (visual and auditory). Rather than routing all information through a single processor, a mixed mode presentation presumes to spread the information across multiple processors, thereby, engaging additional resources for learning. Such an approach is in contrast to the integrated or modular worked example implementations, which instead focus on optimization of information routed to a single processor.

2.3.3: Transition Implementations

Finally, transition implementations take on the form of fading, which seeks to transition a learner from studying instructional material (such as worked examples) to problem solving. Borrowing VanLehn's (1996) framework, initially the learner is simply working on attaining relevant knowledge without attempting to apply it. In the next stage, the learner begins to gain practice in applying the knowledge (problem solving). Typically, scaffolding such as worked examples is referenced at this stage. Within this intermediate stage, the learner grapples with misconceptions or knowledge gaps. Additionally, the learner gains, "...heuristic, experiential knowledge that expedites problem solving." (VanLehn, 1996, p.516) The final stage marks the resolution of misconceptions and knowledge gaps. The learner gains proficiency and speed in solving problems with an occasional error. At this stage, the learner does not add to their relevant domain knowledge or their fundamental procedural knowledge. VanLehn (1996) cautions that these stages are not necessarily distinct in practical application. Indeed, a single student may be at different stages for varying areas of knowledge within a single domain.

Fading is intended to smooth the transition from the use of worked examples to the final problem solving stage by gradually removing scaffolding provided by worked examples until only a problem remains. Renkl, Atkinson, Maier, and Staley (2002) tested fading against the use of example, problem pairs with two classes of ninth grade students learning topics on electricity. One group received complete worked examples (control group), while the other group received worked examples with solution steps gradually removed (a backwards fading scheme was used, where the final step was initially

removed, then the last two steps were removed, etc.). The fading group outperformed the control group on the isomorphic post-test problems (no differences between the classes or for the transfer problems).

A forward fading scheme (using a modified version of Renkl's, 1997, software that incrementally presents probability worked examples) was tested with introductory psychology students. Forward fading initially removes the first step, rather than the last step (next, the first and second steps are removed, etc.). As with backward fading, forward fading also provides enhanced isomorphic performance.

Finally, both forward and backward fading were compared against complete worked examples (and against each other) with Renkl's (1997) program (modified appropriately for both fading schemes). As before, the fading groups exhibited superior isomorphic performance, but the backward fading group also outperformed the control group on the transfer problems. In addition, the backward fading group required less learning time than the forward fading group. However, there were insignificant differences with respect to post-test performance. Hence, the backward fading group attained a comparable level of competence (as measured by the post-test) within a shorter time frame.

In conclusion, there is some evidence that backward fading may better foster transfer performance than forward fading (Renkl, Atkinson, Maier, and Staley, 2002, do acknowledge that more research is necessary to obtain more definitive results). At a minimum, backward fading was found to be more efficient than forward fading. Renkl, et al. (2002) propose three possible aspects that may account for the fading scheme

differences. Renkl, Atkinson, Maier, and Staley (2002) suggest that the removal of the last steps impose a lower cognitive load versus the initial steps. The removal of the final steps was hypothesized to postpone the need to engage in problem solving to complete the example as opposed to removing the initial steps. In the same vein, one might argue that by removing the initial steps, one removes the critical set-up of the solution that forms the foundation upon which the remaining steps are derived. As a consequence, the learner obtains feedback in their ability to solve problems more quickly in the forward fading scheme. Reminiscent of the contextual advantages of worked examples argued in Zhu, et al. (1996), in the backward fading scheme, the prior solution steps are available to provide "...contextual information..." (Renkl, et al., 2002, p.312) Finally, as compared to worked example and problem pairs, fading does provide benefits for isomorphic problems; although, more work is required for transfer performance.

Overall, section 2.3 covers worked example modifications that span specific design elements, complementary scaffolding and presentation strategies across multiple knowledge domains (see table 2.2).

Table 2.2 - Worked example modifications

	Modification	Knowledge Domain Tested
Design Elements	Cued Subgoals (Catrambone, 1995, 1996, and 1998; Atkinson and Derry, 2000 ^a)	Probability
	Incomplete Worked Examples (Stark, 1999)	Probability
	Integrated Format (Tarmizi and Sweller, 1988; Ward and Sweller, 1990)	Geometry, Physics
Complementary Scaffolding	Self Explanation Training (Chi, et al., 1994; Renkl, et al., 1998)	Biology, Finance
	Interactive Self Explanation Elicitation and Feedback (Conati and VanLehn, 2000)	Physics
Presentation Strategies	Fading (Renkl, Atkinson, Maier, and Staley, 2002)	Electricity, Probability
	Modular Format (Gerjets, et al., 2004)	Probability
	Multimode Information Conveyance (Mousavi, et al., 1995)	Geometry

^aAtkinson and Derry (2000) do not specifically mention what domain was tested, although math education was listed as a key word in the paper.

Much like the actions discussed in section 2.2, the above worked example modifications are not independent. Instead, several worked example modifications share common underlying foundations (see figure 2.2).

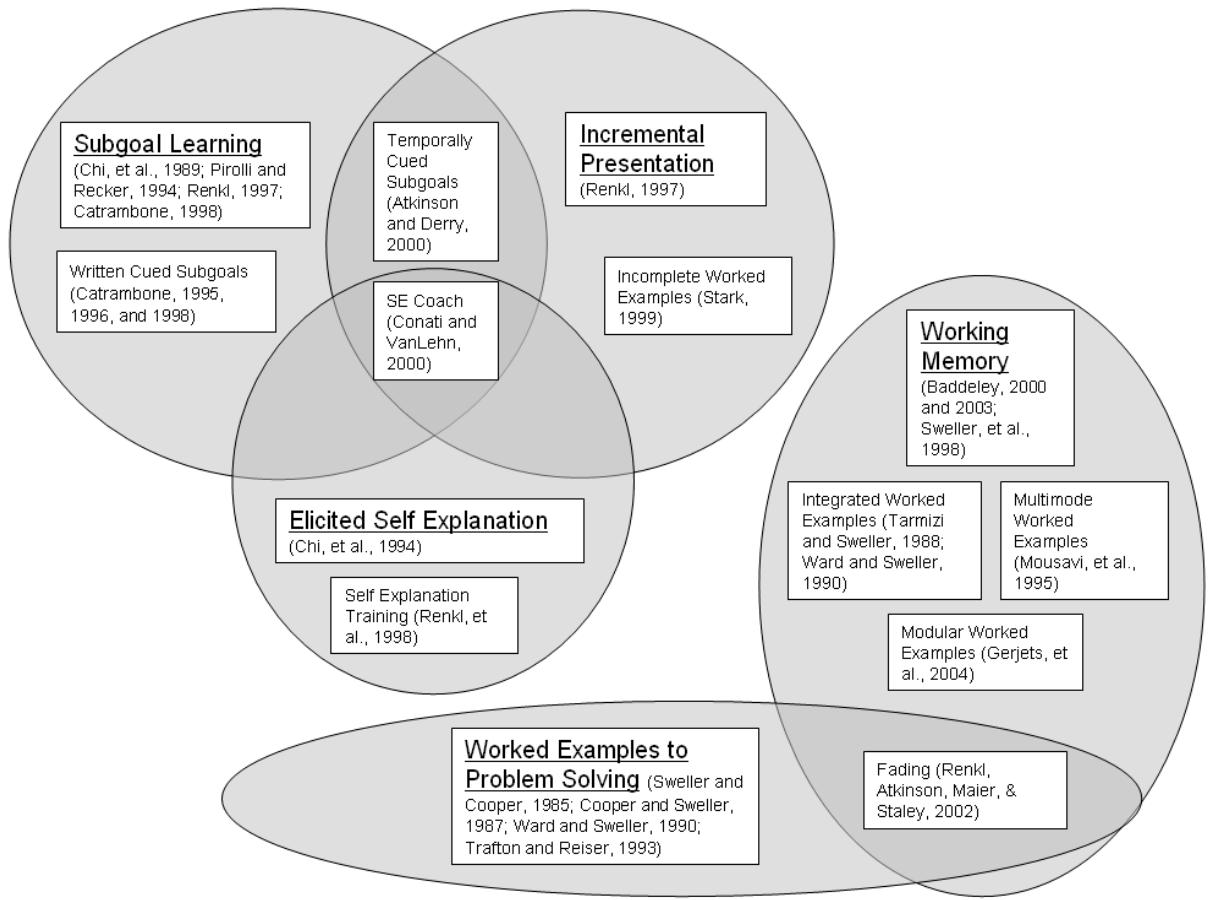


Figure 2.2 – Links across worked example implementations (the bold text represents shared underlying ideas of the implementations)

Another common theme (present throughout this review) arises yet again for the worked example implementations. In part, the underlying mechanisms of the worked example implementations can also be traced back to Palincsar and Brown (1984). Palincsar and Brown (1984) denote four key elements that guide the design of their training scheme: active participation (on the part of the learner), provide feedback on active participation, include information beyond how to perform a task (also provide the

conditions of application), and structure scaffolding, “... in anticipation of competence...” (Palincsar and Brown, 1984, p.123)

The last element has its foundation in Vygotsky’s (1978) work (Zone of Proximal Development). In essence, there is a division between material that a learner can tackle without help and material that a learner will require assistance with. Scaffolding can be used to foster participation at the levels requiring assistance. As the learner gains proficiency with the task, the scaffolding can be gradually removed. Fading (Renkl, Atkinson, Maier, and Staley, 2002) is a direct extension of such a training scheme to “push” a student from worked examples to solving problems, while the related implementations directly inspired from working memory (Tarmizi and Sweller, 1988; Ward and Sweller, 1990; Mousavi, et al., 1995; Gerjets, et al., 2004) can optimize the stages of such a “push.”

Active participation is explicitly required in several implementations such as incomplete worked examples (Stark, 1999) and self explanation elicitation (Chi, et al., 1994; Renkl, et al., 1998; Conati and VanLehn, 2000). Active participation is further inherent in subgoal learning (for instance, see Catrambone’s, 1998, Subgoal Learning Model). Although one might argue that inherent active participation extends further into all implementations (for instance, consider the fundamental notion of the working memory applications seeking to optimize use of available, active mental processing capacity).

In addition, subgoal learning not only encompasses how to solve a problem (Pirolli and Recker, 1994; Catrambone, 1998), but also the conditions and consequences of subgoal application (Chi, et al., 1991; Renkl, 1997).

Finally, there are two classes of feedback with the implementations. Self explanation training (Chi, et al., 1994; Renkl, et al., 1998) provides initial training and feedback from an expert prior to application, while a computer based system (Conati and VanLehn, 2000) can provide feedback concurrent with worked example study.

In summary, much research on worked example design and implementation has been influenced by the results from earlier research on the processing of worked examples (see section 2.2). Unfortunately, the worked example design and implementation work has either yielded mixed results or is untested within the realm of physics. Hence, there remain many possible avenues still unexplored for future work in the design and implementation of worked examples, particularly those avenues which are effective in prompting those beneficial learning actions identified in section 2.2. Additionally, the worked example literature has largely focused upon correct solutions as opposed to also including the use of incorrect solutions. Reassuringly, the limited available results (Große and Renkl, 2007), while untested within the area of physics, are reasonably consistent with the aforementioned results from the worked example and reading comprehension literature.

CHAPTER 3: METHODOLOGY

This chapter describes the study's approach to addressing the core research question: what actions are invoked when selecting the correct solution to a physics problem when given three possible solutions? The first section discusses the ideas behind the development of the study's task and materials. The second section describes how the study was performed, while the third section explains how the data analysis was done.

3.1: Task Development

This task largely emerged from two issues: the absence of reforms definitively triggering beneficial learning actions in the realm of physics, and the rise of reforms integrating incorrect solutions despite the limited results on what actions arise with the inclusion of incorrect solutions (in sharp contrast to the abundance of results on learning from correct solutions). Research from the areas of reading comprehension (Palincsar and Brown, 1984) and worked examples (Atkinson, et al., 2000) provide an understanding of beneficial learning actions and processes. However, subsequent work intended to initiate such processing has largely not been extended to a physics-specific realm. Instead, such work has focused on what can reasonably be considered as less complex domains such as basic probability or introductory finance (some examples are Catrambone, 1995, 1996, and 1998; Stark, 1999; Renkl, et al., 1998). Other research in the realm of physics has yielded mixed results (Conati, et al., 2000), where only subsets of participants exhibited learning gains associated with instructional modifications designed to encourage

beneficial processing. Such mixed results have also arisen in my own early pilot work exploring the use of spatial subgoal cueing (as opposed to Catrambone's, 1998, written subgoal cueing). My work indicated that if the material in the worked examples was overly familiar to the participants, then the worked examples had questionable impact on subsequent problem solving performance. Whereas, more complex worked examples began to overwhelm participants or otherwise suggest shallow levels of processing.

Overall, prior research has not definitively leveraged our understanding of beneficial processing into research-based instructional reforms intent upon triggering such processing.

However, the emergence of promising reforms integrating incorrect solutions (several examples are Reif and Scott, 1999; Harper, et al., 2007; Etkina, et al., 2006) suggests the possibility of leveraging the inclusion of incorrect solutions to foster beneficial learning actions. Yet there is little research on what processing actually occurs when integrating incorrect solutions into instructional material (Große and Renkl, 2007). Furthermore, the results from Große and Renkl (2007) have not been verified in the domain of physics. Indeed, inspiration for the use of incorrect solutions can be partially traced back to Peters (1982), who was not actually targeting instructional reform. Rather, Peters was probing the conceptual understanding of introductory physics students.

This task also integrated multiple possible solutions for each problem (one correct and two incorrect solutions). Catrambone and Holyoak (1989) found the comparison of multiple solutions helped participants recognize the deep structure. In essence, such a

design allows for the explicit accessing of the comparison action denoted in Palincsar and Brown (1984).

3.1.1: Problem Set Development

The researchers initially generated a pool of possible problem sets. These problem sets were tested in a separate pilot study with undergraduate participants (who did not participate in the subsequent study described in this paper). Based upon the results of the pilot study, the problem sets were revised. In addition, the pilot study provided data used in the development of the coding schemes (see section 3.3).

Based upon the recommendations of Yerushalmi and Polingher (2006), the incorrect solutions were created by the researchers and inspired by common student errors (as opposed to providing incorrect solutions directly copied from student responses).

Chi, et al. (1989) found that both “good” and “poor” performers self explained math issues, but only the “good” performers also focused upon physics issues. Hence, only physics related errors were included in this study. While the participants were not specifically informed to target only physics errors, in some instances participants inferred that only physics errors were present.

The problem sets (see Appendix A) covered mechanics topics spanning the first semester of the calculus-based, introductory physics sequence. “Close the Door” and “Two Pucks” included concepts from linear motion, energy, and rotational motion. The

uniform circular motion (or “UCM Spring”) problem set focused upon rotational motion topics. “Fission” was based upon energy concepts.

The problem sets included three types of information representations: prose, pictorial, and mathematical. Errors were introduced into all three representation types (see Table 3.1). Errors included inconsistencies between information presented in the various representations. For instance, one “Fission” solution included a potential energy (in a mathematical representation) for the interaction between a nucleus in the initial state and another nucleus in the final state (where the state selections were given pictorially).

Table 3.1 indicates such inconsistencies as errors in the mathematical and pictorial representation.

Table 3.1 - Error distribution across information representations (an “X” indicates the presence of an error(s) in the associated representation).

Problem Set	Prose Representation	Mathematical Representation	Pictorial Representation
Close the Door	X	X	
UCM Spring	X	X	
Fission		X	X
Two Pucks		X	

The problem sets also contain design features that are either unique or shared with one other problem set. “UCM Spring” and “Fission” share design features typical of common end-of-chapter style problems: use of a single fundamental principle throughout

their solutions (the other problem sets had a different fundamental principle for each solution) and a numerical final answer (the other problem sets asked for more atypical non-numerical answers). Undergraduate participants were recruited from the calculus-based introductory physics course sequence that uses the Matter and Interactions curriculum. “Fission” directly stemmed from an end-of-chapter problem from the Matter and Interactions textbook (Chabay and Sherwood, 2002), while “UCM Spring” was largely inspired from a final exam question of the same course. In contrast to the more typical solution formats employed for “UCM Spring” and “Fission”, “Close the Door” was more conceptually based with solutions further biased toward prose. “Fission” included the pictorial information representation in its solutions, while “Two Pucks” and “UCM Spring” had diagrams in their problem statements. Finally, “Fission” represented the only microscopic situation, whereas, the other problems were in the macroscopic realm.

3.2: Structure of the Study

Two researchers designed this study in order to address separate research questions. Hence, the structure of the study extended beyond the scope of the aforementioned research questions in this paper. The study took place in three stages: individual data collection and two sessions of group data collections. The scope of this research is restricted to the analysis of data collected in the first stage (see Section 4.3.1 for a discussion on future work related to the subsequent stages).

3.2.1: Participant Background

The solutions in the problem sets were designed using terminology and notation from the Matter and Interactions (M&I) curriculum. Additionally, the problem sets included concepts spanning the first semester of the introductory, calculus-based physics course. Hence, the undergraduate participants were recruited at the start of the semester from the M&I sections of the second semester, introductory, calculus-based physics course. As a requirement to participate, the undergraduate volunteers had to have completed the associated first semester of the introductory sequence.

Six graduate participants were recruited to participate prior to the undergraduate participant data collection sessions. Stemming from these initial sessions, the “Two Pucks” problem’s diagram was slightly revised (the time stamps of the pucks were labeled), and the introductory script was slightly modified (see Appendix B) in response to the participants’ questions. The graduate participants were not screened for M&I background since most graduate students would not have had previously taken courses using the M&I curriculum. Furthermore, given the higher level of physics background with the graduate participants, notational and terminology issues were less of a concern. Indeed, despite the differences in the backgrounds of the two samples of participants, few differences arose in the analysis (see Chapter 4 for a discussion of the results).

All participants were volunteers, who received monetary compensation for participation.

3.2.2: Study Sequence

Three researchers administered the individual data collection sessions. All three researchers read the same scripts (see Appendix B) and were trained to run the sessions in the same manner. In order to record the participants' mental processing during the task, think aloud protocols (Ericsson and Simon, 1993) were employed. Following an introduction to the study and the task, the participants were instructed on think aloud protocols. Subsequently, they worked through warm-up exercises to gain practice with the requirements of the think aloud protocols (see Appendix B).

In order to assure that the participants would be finishing the data collection in time for the subsequent phases, a time limit of 15 minutes was given for each problem set. After completing all four problem sets, the participants were given a break before starting the remaining phases of the study. The subsequent phases included a participant discussion on what they decided in the individual session. Participants were fully informed of the subsequent phases prior to the start of the individual sessions. Hence, one cannot discount possible effects due to the foreknowledge that their decisions would be discussed with others. To help protect participant identities, all participants selected pseudonyms, and they were called as such for the full duration of the study. In order to address possible undue anxiety, the participants were allowed to opt out in discussing their decisions for one problem set of their choice.

The phases following the individual sessions are otherwise outside the scope of the research questions addressed in this study. Hence all further discussion will focus solely on the individual data collection sessions.

3.2.3: Study Environment

The data collection took place in the Qualitative Education Research Laboratory¹⁹ at North Carolina State University. The laboratory consists of three interview rooms that surround a central observation room.

Each interview room has ceiling mounted, pan-tilt-zoom cameras that can be controlled either by the researcher administering the data collection (as was done in this work) or by another researcher in the observation room. Each room is also equipped with a hard-wired, table-top microphone to capture audio data. The video and audio data streams are routed to a computer that records the data streams directly into its hard-drive.

3.3: Analysis Methodology

All data collection sessions were transcribed. The transcriptions were segmented in a comparable fashion to Chi, et al. (1989), where each segment represents a statement of a single action. Hence, segments can encompass one or many utterances. For example, one type of action is planning (making a statement of one's own actions). "Fay" simply stated, "*Okay. I'm starting with the shortest one.*" Yet, "Marco" provided a more verbose account of a similar type of approach, "*I'm going to sort the solutions by length. In fact I'm doing that right now. I'm going to read the shortest solution first, then the longer solutions later, and then -- that's the way we're going to do it.*" Hence, while each

¹⁹ See <http://www.ncsu.edu/PER/facilities.html> for a virtual tour (downloaded on 9/20/09).

segment contains varying amounts of utterances, the underlying action (planning) is the same.

Using earlier pilot study data that is not included in the analysis reported in this paper, the analysis coding schemes were developed. The first and second coding passes were developed together (that is, each segment was coded for both passes at once, and coding agreement was only reached if both coding passes matched for individual segments). The third, fourth, and fifth coding passes were developed separately.

Table 3.2 – Coding pass overview

Coding Pass	Description
Action	Action that underlies the segment.
Context	Does the segment pertain to a physics-related context or other context?
Focus	Is the focus on the approach or on the answer?
Correct/Incorrect	Is the participant making a correct or incorrect statement?
Restate Type	What type of information representation is being restated?

Initially, a single researcher proposed a preliminary scheme. Working with an independent coder, the coding passes were tested for reliability (see Table 3.3). Per the recommendations of Ericsson and Simon (1993), data from a separate, earlier pilot study, was used to test for reliability. Much of the research in the worked example realm reports inter-rater reliability as percent agreement between two independent coders. However, such a reliability measure does not account for the possibility that the two coders might

agree simply due to chance. On the other hand, another reliability measure, Cohen's Kappa (Cohen, 1960), removes the chance agreements. A Cohen's Kappa of 0.8 or higher is considered reasonable reliability (Geisler, 2003). Table 3.3 reports the Cohen's Kappa for the coding passes.

Table 3.3 – Coding reliability

Coding Pass	Cohen's Kappa
Action + Context	0.93
Focus	0.84
Correct/Incorrect	0.88
Restate Type	0.85

The first and second coding passes required substantial development and iterations to achieve acceptable reliability. Stemming from concern of possible bias building in the second coder due to the need for repeated iterations during the development of these passes, a third independent coder was given a subset of the segments for verification (Cohen's Kappa = 0.80).

All coding passes were also tested for validity. A fourth independent researcher was provided with the code definitions and sample segments for each code. The independent researcher was asked to validate whether the code definitions accurately reflected the sample segments. All coding schemes passed validation without any required changes.

3.3.1: Action-Coding Pass

The action-coding pass identified the underlying action to each segment. This pass was the most complex, which contained seven codes (six actions and one “catch-all” other code): Planning, Predicting, Judging, Expressing Uncertainty, Comparing, Restating, and Other.

Planning refers to a participant stating what actions that he or she will undertake or is undertaking. For example, “Marco” stated an overall method to begin the task, *“I'm going to sort the solutions by length. In fact I'm doing that right now. I'm going to read the shorter solution first, then the longer solutions later, and then -- that's how we're going to do it.”* Similarly, “Fay” stated, *“Okay. I'm starting with the shortest one.”* After indicating his choice for the correct solution, “Walter” subsequently planned, *“And I'm going to try to figure out what error might be in the other problems.”*

A prediction action occurs when a participant states his or her own solution to the physics problem. These actions can encompass stating a single step or proposing several steps to a solution. For instance before examining the given solutions to the “Fission” problem, “Eduard” provides much of his approach to solving the problem. *“Okay, so I'm going to have to use, electric force (picks up and starts looking at equation sheet), equation, wherever the hell that is. Okay, U electric equals 9 x 10 to the ninth Q1 Q2 over R (Participant writes this out on the white board). Okay, they both have a charge of 46 e, if I had something to set U electric equal to, then I could solve for R.”*

A judging action is the determination of whether a piece of information (stemming from provided materials or from the participant’s own statements) is correct or

not. In a typical judgment, “Christobal” declared, “*Well, that's not quite true.*” Other judgments are more specific. In reference to a solution with “Close the door”, “Dolly” disagrees, “*Solution number one, I don't think they can say that you know that the rubber ball will bounce back with the same speed as before the collision. I think that's an assumption. I think that depends on the door.*” A common judgment is the acceptance or rejection of a solution as a whole. For example, “Ike” said, “*So, solution number one I believe is also wrong (participant draws a big red X on the top of solution number one).*”

A participant expresses uncertainty when he or she denotes doubt or poses a question. While reading the “Close the door” problem, “Ana” declared, “*I don't remember what that means.*” Similarly, “Otto” stated, “*I don't remember the formula for springs.*” While looking over a solution for the “Fission” problem, “Hanna” questioned, “*why would the initial state be two of these?*” “Gaston” questioned the physics of the “Close the door” problem. “*Just looking through them, there is, I am not quite sure what the angular momentum of the door really has to do with what the question is asking.*”

Comparing refers to a participant comparing element(s) of two or more items (such as comparing pieces of information from two of the given solutions or comparing a given solution to the participant’s predicted solution). For example, “Claudette” compared the fundamental principles used on each of the given solutions for the “Two pucks” problem. “*Well, they are using different principles for each one.*” “Teddy” provides another example of comparing two of the given solutions, “*Let's see, so is that one (points to solution number three). It looks like these formulas are pretty similar (pointing to solution number one and solution number three).*” “Marco” compares the

result of a given solution to the “Two pucks” problem to his predicted solution. “Sorry, and their conclusion is the same as mine.”

At the core of the restating action is interpretation. In essence, the participant is stating given information in his or her own words. This action includes summarizing information, such as “Marco” condensing several algebra steps in a solution for the “Two pucks” problem. “Ow, for the second puck blah blah blah ... they plug stuff in...” Other examples of restating largely resemble reading given information, however, the participants are not reading verbatim (they are using their own terminology, and hence invoking an interpretative processes aside from basic verbatim reading). For example, “Ike” substitutes the phrase “*palladium nucleus*” for “*Pd-118 nucleus*” when reading the “Fission” problem. Similarly, the participants also assign meaning to information presented in a mathematical representation (generally providing meaning to variables). For instance, “Dolly” restated equations in a solution for the “Fission” problem.

“Solution number two says one energy is equal to -- good Lord, solution two says one energy is equal to uranium energy minus the Palladium energy. This must be, okay, the change in energy. They are saying that, that energy is equal to uranium energy minus the Palladium energy.” The participants also assign meaning to information in prose form.

Upon reading that the spring has a negligible mass for the “UCM spring” problem, “Earl” noted, “*So, it is safe to assume that the masses probably like .001 or something like that, probably lighter than a feather even.*”

3.3.2: Context-Coding Pass

The context codes denote whether a segment regards the physics at hand or something else. Both Judgments and Comparing segments were coded for context. Some judgments included a justification or argument. Hence, two physics related context codes were used: one assigned to a physics argument, while the other code simply focused on judgments only referring to physics ideas, concepts, or principles without an accompanying argument (called physics-related discussion).

Essentially, the physics-related discussion code denoted that the participant was considering the physics in some fashion, but without giving an explicit indication of processing a physics-related chain of reasoning. The key distinction for a physics argument code (as opposed to the more general physics-related discussion context code) is the presence of a physics argument justifying **why** the participant concluded that a piece of information was correct or not. For example, *"This one also I just realized it doesn't include the earth, the system is incorrectly including potential energy and the equations at all, that should be part of the work."* The participant is denoting an error due to an inconsistency between the choice of system versus the inclusion of potential energy terms accounting for an external interaction. In contrast, another participant is discussing the inclusion and exclusion of energies in a “Fission” solution **without** denoting why the energies are incorrect: *"See I would say that this one included, the rest mass, the rest energy of the uranium atom, uh, and not the palladium atom. That's definitely wrong."* Hence, the participant is processing physics-related ideas but not providing a coherent justification to back the assertion that the solution is incorrect.

Two more context-codes were used for the judgment segments: Math-Related and Other. The other code is the “catch-all” code (generally used for those segments where the context was unclear). The Math-Related code included judging math-related steps (checking computations, algebraic manipulations, etc.) as well as checking equations (without an explicit regard for the physics itself, such as checking an equation against those provided in the given reference sheet). For instance, “Fay” is checking the derivative computations in a “UCM Spring” solution, *“I think the derivative of a cosine is a minus omega squared. That math is nice.”* “Arthur” provides an example of checking an equation, *“Yet that's, that's the right formula.”*

The comparing segments were coded for either physics-related or other context. Similar to the judgment segments, physics-related codes indicate where the participant was comparing physics ideas, concepts, etc. For example, “Claudette” compared which fundamental principle the “Fission” solutions used, *“okay (moves problem statement off to the side). OK, Well, they are all using the energy principle.”* The non-physics-related comparisons were coded as other context, such as Walter comparing a numerical result across “UCM Spring” solutions, *“Okay, so both of them got 4.19 radians per second part.”*

3.3.3: Focus-Coding Pass

The focus-coding pass denotes whether the participant was focused upon the approach or the answer of the solution when predicting, judging, comparing, or expressing uncertainty. To ensure coding reliability, specific break points separating the

approach from the answer were defined for each solution (see Appendix C). For the purposes of this research, the answer does not necessarily refer to the final result. For the two problem sets with a non-numerical result (“Close the Door” and “Two Pucks”), the answer is considered to be the final result. In the case of “Two Pucks”, the final result is the comparison of the magnitudes of the forces acting on each puck. For “Close the Door”, the final result is the determination of whether the clay or rubber ball is the better choice. Both “UCM Spring” and “Fission” had numerical final results. Their solutions tend to employ physics to set-up the solution. Once a relationship between the final answer and the givens has been determined, algebraic manipulations and numerical substitution are done to compute the end result. This coding scheme considered the approach completed once a relationship between the givens and the final results emerged from a physics analysis. The subsequent non-physics pertinent steps, algebraic manipulations and substitutions, were not considered part of the approach.

3.3.4: Correct/Incorrect-Coding Pass

The correct/incorrect-coding pass identifies whether the participant made a correct or incorrect judgment or prediction. When looking at the Two Pucks problem set, “Hugo” made the “intuitive” (incorrect) prediction that the spinning puck would have a larger force, “*Well, I didn't think, I thought that the second puck would go, would require a larger force.*” “Winfred” correctly predicted that the final answer to the Close the Door problem was the rubber ball, “*I think that the rubber ball is the better choice, I think that this is.*” As with all coding passes, there is a “catch-all” other code, primarily for those

segments where correct or incorrect cannot be determined. “Omar” judges something as being correct, but he does not specify exactly what he is judging, “*That's okay.*” On the other hand, “Ike” is more specific and judges correctly that there are no math errors, “*I agree with the math*” (as aforementioned, no math related errors were introduced in the problem sets). “Sally” correctly judges that the first solution in the Close the Door problem set is correct, “*I think the solution is one, one is correct, um, because I think you do have to consider that the collision is elastic.*” In contrast, “Otto” incorrectly judges that rotational motion does not apply to the Close the Door problem, “*...But there is, there is no torque involved with the door. Because the door isn't spinning in any way, it's just moving back and forth. So this can't be right (marking number two with a “x”).*”

3.3.5: Restate Type-Coding Pass

The restate type-coding pass gauged what type of information representation the participant was restating. This coding pass has two stages: whether the segment was restating the problem statement or a given solution and the information representation. For example, “Fay” restated the prose in the Fission problem statement, “*Uranium! (Inaudible) undergoes fission process resulting in two palladium nuclei. A palladium nucleus has a charge of 46 e. and the rest mass of blah, where e is this (censored), and U is that. U has a charge of 92 e and the rest mass of this. there will also be some free neutrons after the fission process, but ignore it. What is the distance between the nuclei centers just after the fission process when they are initially at rest?*” Similarly, “Sally” is restating the diagram given in the Two Pucks problem statement, “*Okay. Let's see here.*

So this is the two different pictures (gesturing to diagrams in problem statement). Of all of puck one, and all of -- this is final and initial puck one and final and initial puck two.

Looking down, okay." The Fission problem set was the only one that had diagrams in the solutions. "Ana" restated the diagrams in the second solution as, "*And they're saying the initial state is already split into palladium nuclei.*" The prose representation was also used in the solutions. "Walter" restated the third solution to the Close the Door problem set as, "*hmmm, This one (solution three) says that since the masses are equal and they have equal velocities, and the momentum, since both the clay and the rubber ball will provide the same amount of momentum to the door, they're thrown with the same initial speed, it doesn't matter which one you pick. hmm, hmm,*" The final representation, mathematical, was only used in the solutions. For example, "Teddy" restated an equation in a Fission solution, "*um, This formula (looking at number three) it is saying that there is a change in potential and a change in rest.*"

CHAPTER 4: RESULTS

This chapter covers the results from the analysis in Chapter 3. Overall, the actions stemming from the tasks of selecting a correct solutions and finding errors are largely consistent with those found in the worked example and reading comprehension literature. Few differences were found between the graduate participants versus the undergraduate participants. However, several differences arose across the problem sets, which share intriguing connections to the underlying designs of the problems and associated solutions. The data further suggests implications for directing participant attention.

The discussion of the results has two sections. The first section addresses the core research question of the study: what actions emerge when the participants were tasked with identifying the correct solution to a physics problem when provided with three possibilities (and identify the errors in the other possible solutions). Additionally, this section links such actions with prior research.

The second section of the results chapter discusses distinctions that arise in the analysis. These distinctions lie along two (dependent) dimensions: groupings of participants and problem sets. In particular, there is a group of participants who tend to approach the task much as a “grading” style exercise. In essence, these participants are largely focused on interpreting the given solutions and judging whether the solutions are correct or incorrect. Through comparing the group of these “judgers” against the other participants, distinctions across the problem sets arise. These distinctions generally divide

along the following pairs of problems: “Close the door” paired with “Two pucks” and “UCM spring” paired with “Fission.” Reassuringly, the problems in each pairing also share design characteristics, perhaps linking their results.

The results sections are comprised of both qualitative analyses and quantitative analyses. For the quantitative analyses, non-parametric statistical measures are employed since there are no action-centric constraints necessarily leading to a normal distribution of the data for the population as a whole (particularly of concern regarding an assumption of a normal distribution is the limited data available on the processing of incorrect solutions, especially for incorrect solutions in the realm of physics). Similar concerns also exist for the other coding passes. Thus, parametric measures could potentially lead to inappropriate conclusions.

4.1: Task Actions

This section addresses the central research focus of identifying what actions emerge when the participants were asked to determine the correct solution to a physics problem when provided with three possible solutions as well as note the errors in the incorrect solutions. As discussed in Section 3.3.1, six distinctive actions emerged when the participants were performing the study’s task: Planning, Predicting, Judging, Expressing Uncertainty, Comparing, and Restating.

4.1.1: Connections to Palincsar and Brown (1984)

As with actions from the worked example literature, the actions found in this study have connections to those in the Palincsar and Brown (1984) framework (see figure 4.1).

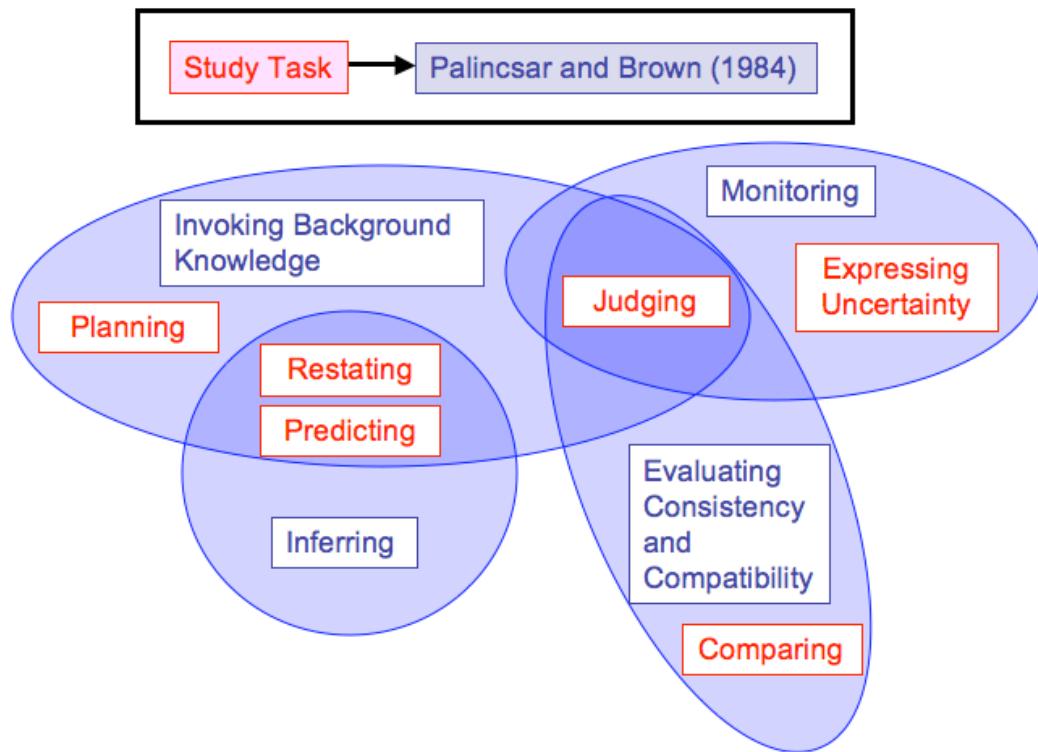


Figure 4.1 - Ven Diagram illustrating the connections of the participant actions exhibited in the study to the actions in the Palincsar and Brown (1984) framework.

Invoking background knowledge²⁰ underlies multiple actions observed in the study. Planning statements can include concepts or ideas culled from prior instruction

²⁰ I am not asserting that the participants were necessarily correctly applying background knowledge (indeed, there are several examples of incorrect application), rather I am

(invoking background knowledge), such as, “*And, ah, I'll have to determine whether or not rest energy is involved.*” or “*All I have to see if this one (points to solution number one) if the energy principle consideration that has been taken in solution one is right or not.*”

Each restating segment includes a degree of interpretation. Much of the interpretations directly reference relevant physics knowledge. For example while examining the solutions to “Two pucks”, “Fay” noted, “*Here we have, we account for a gravitational force.*” and “*So, we have now the first puck requires a larger force, so they're saying that by pulling directly on the center of the puck takes more force than adding a rotational and translational energy.*” Similarly, “Arthur” interpreted an equation in a solution associated with “UCM spring”: “*umm, Well, mv, mass times velocity is equal to. So they're setting speed of the block moving around equal to the spring's stiffness. That's the length times the time it takes to go around.*”

Predictions ranged from proposing a full solution to proposing a single step of a solution. In some cases, the predictions directly referenced physics related quantities. For instance while examining “Close the door”, “Teddy” predicted the solution as, “*Something with mass. I'm not seeing anything. All right. Force is mass times velocity ... Delta p for the momentum equals f net Delta t. The time is going to be the same for both of them I think. Force, force equals ma. So the mass would be, for the clay, cause the clay would stick to it, so it would be the mass clay plus the mass of the door (continues to*

simply asserting that background knowledge was invoked in some fashion for a range of actions.

write on white board) times acceleration times delta t.” and subsequently, “because it will have more mass with the same like acceleration and force and everything. So I would say the clay would make the, uh, make the door shut.”

Judgments factor into invoking background knowledge through physics based justifications and judgments pertaining to physics issues. While working on “Fission”, “Earl” noted, *“It was just one charge, there is no U_i , because potential energies between two objects, and since the initial object was just one, one nucleus, or one atom, there is no U_i , so that makes sense.”* While considering the same problem and solution set, “Claudette” judged, *“Just the rest energy, yeah I am going to go with two just because I feel like they have more of a potential energy at the beginning than once they’re apart, then the potential energy there isn’t, there should not be any final potential energy.”*

Both judging and expressing uncertainty have elements of monitoring. In particular, judging is associated with positive monitoring, where a participant indicates comprehension. While reading a solution for “Close the door”, “Walter” judged, *“Oh I see what it’s saying. Yeah, okay, that makes sense.”* Expressing uncertainty is linked to negative monitoring, where a participant notes a comprehension failure. While reading through “UCM spring”, “Christobal” stated, *“No, I don’t know what that means.”* When reading through a solution to the same problem, “Ana” said, *“I don’t remember what A cosine w t, what is that?”*

Restating and predicting have elements of inference. For the purposes of this research, the nature of restating is an inference of meaning for prose, mathematical, or pictorial information. For instance, “Omar” restated a portion of the problem statement

for “Two pucks” as, “*The time through which the forces are acting are same.*” The problem statement actually stated, “The threads are pulled by a constant force (of magnitudes F_1 and F_2 respectively) for 4 seconds at the free end of the thread ...” Another example is “Walter’s” restatement of a given solution to “Fission”, “*The initial state, after it explodes, they are trying to figure out the difference in mass, and how much energy that would cause and how much that would push them if I'm correct on this because some of the mass like explodes or something.*” “Arthur” provides an example of interpreting a diagram in the “Two pucks” problem statement: “*Puck 1, Time zero. Okay. So it's a time lapse (gesturing with hand to the right along the diagram, as of signifying a time axis pointing to the right). Puck 1 at zero, puck one at four. This is d. So it moves the same distance.*”

As acknowledged in Palincsar and Brown (1984), inference and prediction are tied together into a single component of their framework. Broadly speaking, a prediction is an inference stemming from the problem statement to reach the problem’s specified goal. Directly after reading the second problem statement, “Christobal” decides, “*Well, I think the easiest way to do this problem would be to look at, we need to conserve energy, and so the energy of just after the fission process, so the fission process must conserve energy and so the difference in the energy of the two nuclei is going to be compensated by the fact that they will be, in the potential well.*”

Two observed actions are associated with compatibility and consistency evaluation: judging and comparing. Inherently, judging involves the evaluation of information (from a given solution for example) against one’s own understanding. For

instance, “Winfred” recognizes an inconsistency between a solution to his understanding of “Fission’s” problem statement: “*Hmm, I think that once again that this diagram (gesturing to solution number one) is wrong because, well, because these arrows indicate that they are moving when the problem says the, right after they're split and they're at rest, because if the arrows were moving then you would have to use the kinetic energy initial here.*” In another sense, “Walter” denotes an incompatibility between his understanding of the physics underlying the situation of “Close the door” versus a given solution: “*Okay. So like the, the problem with solution three is that they, they did not assume the extra momentum that the door would gain by having the rubber ball bounce off of it...*”

At the heart of the comparing action is a consistency evaluation between two or more items (i.e., answering the question: “Are these items the same or different?”). For example, participants commonly made comparisons across the given solutions. “Earl” notes consistencies across solutions for “Fission”: “*Okay, both the solution number two and solution number three look like they have the same work.*” while subsequently noting, “*The only thing I see different is the pictures...*”

Deep Structure focus, a component of the Palincsar and Brown (1984) framework, is not included in Figure 4.1, however, it is relevant in a more global sense. Within the context of this study, deep structure focus resides within the Physics context codes. Such codes represent segments across those actions coded for context where the participant is focused upon the physics, as opposed to the math, for instance.

Overall, we have arrived at a reassuring convergence to prior work (Atkinson, et al., 2000; Palincsar and Brown, 1984). The actions observed in this study are indeed consistent with those recorded from related branches of research (the task under consideration certainly involves elements of reading comprehension and learning from worked examples).

As anticipated, the comparison does indeed emerge with the inclusion of multiple possible solutions. Indeed, the data suggests a method to utilize variations in consistency to focus participant attention. Physics-context judgments devoted to the selection of a fundamental principle tend to be associated with those problems whose solutions used different fundamental principles. Similar trends are evident with variations of system selections and final/initial state selections. Hence, the data suggests that participant processing might be focused upon a specific aspect of a solution (such as fundamental principle selection or system selection) by varying that aspect across possible solutions.

A comparable effect arises when the restate segments were examined. In particular, when the information representation is varied (i.e., conveying information through prose, mathematical representations, or pictorial representations), the sample as a whole tends to exhibit some degree of effort in interpreting these representations (see Figure 4.2).

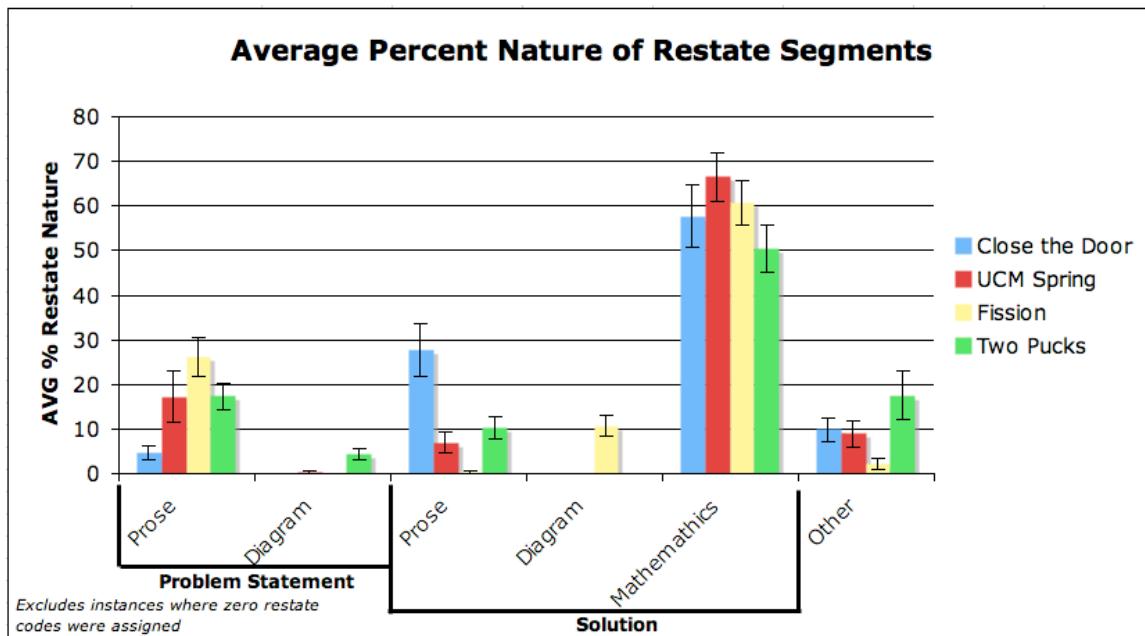


Figure 4.2 - Average percent nature of restate segments²¹

All problems had prose in their problem statements, and all problems had restate segments on the problem statement prose. The first problem ("Close the door") stands out with a lower average percent of problem statement prose restates (Kruskal-Wallis, $H=15.41$, $p=0.0015$). Given that "Close the door" had the lowest average expressing uncertainty (Kruskal-Wallis, $H=9.14$, $p=0.0275$) coupled with having the shortest problem statement, the data suggests that the lowered average of restating problem statement prose is connected with "Close the door" having the least demanding situation to process.

Only the "UCM Spring" and "Two Pucks" problems had diagrams in their problem statements; hence, they were the only two problems with instances of restating pictorial information in the problem statement. One could reasonably argue that there is

²¹ Error bars in all plots represent +/- one standard error.

insignificant activity associated with the problem statement diagram in “UCM spring.” However, such a result is not overly unexpected given the difference in complexity of the diagrams in the “Two pucks” problem (see figure 4.3) as opposed to the “UCM Spring” problem (see figure 4.4), implying that this effect may have a complexity threshold in order to emerge. Indeed, the “Two pucks” problem stands out with the most restate segments devoted to the problem statement’s pictorial information (Kruskal-Wallis, $H=12.09$, $p=0.0071$).

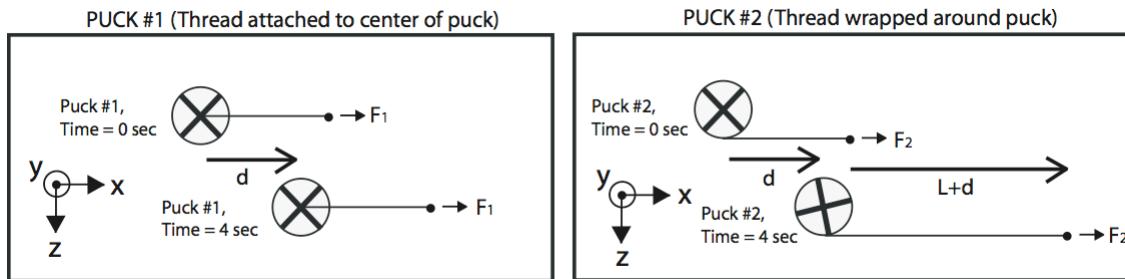


Figure 4.3 - Diagram in the “Two pucks” problem statement

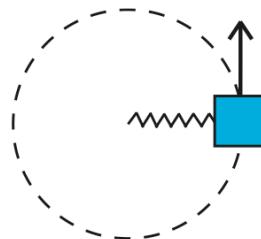


Figure 4.4 - Diagram in the “UCM” problem statement

All problems have some degree of restate segments associated with the prose in the solutions. Indeed, the “Close the door” problem, whose solutions tend toward more prose, is associated with a higher average of solution prose restate segments; while on the

other extreme, the “Fission” problem, whose solutions have minimal prose, has the lowest average solution prose restates (Kruskal-Wallis, $H=22.69$, $p=0.00005$). Given the minimal prose in the “Fission” solutions coupled with the minimal restates of prose in the solutions, there may also be a comparable complexity effect as seen with the pictorial representation.

Only the “Fission” problem has diagrams in its solutions, and the “Fission” problem has the only restates on solution diagrams (Kruskal-Wallis, $H=17.67$, $p=0.0005$).

All problems have restate segments devoted to the processing of information in a mathematical representation. As expected from the above discussion, all problems have solutions that use a mathematical representation to convey information.

Overall, as information representation was varied across the problems and solutions, restate segments were generally associated with those representations. No representation (of those included in this study) was necessarily discounted. The implications of these results suggest future work. Will other modes (such as auditory for example) also be associated with participant processing? Can participant attention be focused upon a specific representation? For example, can this lend itself to an instructional invention designed to integrate practice interpreting data plots?

4.2: Distinctions in the Data

This section covers the distinctions that arose in the analysis. A group of participants (called “judgers”) with distinctive profiles were found. These “judgers” tended to interpret given information and judge whether it was correct or not (much as a

“grading” exercise) as opposed to the other participants who exhibited more varied profiles, including prediction, comparison, expressing uncertainty and planning actions. When the actions of the “judgers” were compared against those of the remaining participants, the analysis revealed distinct profiles for the problems. Much of these distinctive features are split along two sets of problems. “Close the door” and “Two Pucks” tend to share features (as well as sharing underlying design characteristics), while “UCM Spring” and “Fission” tended to be similarly paired. This section will also discuss distinctions that arise from comparing the overall profiles of the problems.

4.2.1: Distinctions across Participants

Of the 21 participants, 5 participants stand out with distinctive profiles concentrated upon the Judging, Restating, and Other actions for all problem sets in the study. In essence, these participants tend to approach the task much as an evaluation exercise (i.e., interpreting given information and gauging whether the information is correct or not). The most prominent example of such a participant is “Igor” (see figure 4.5).

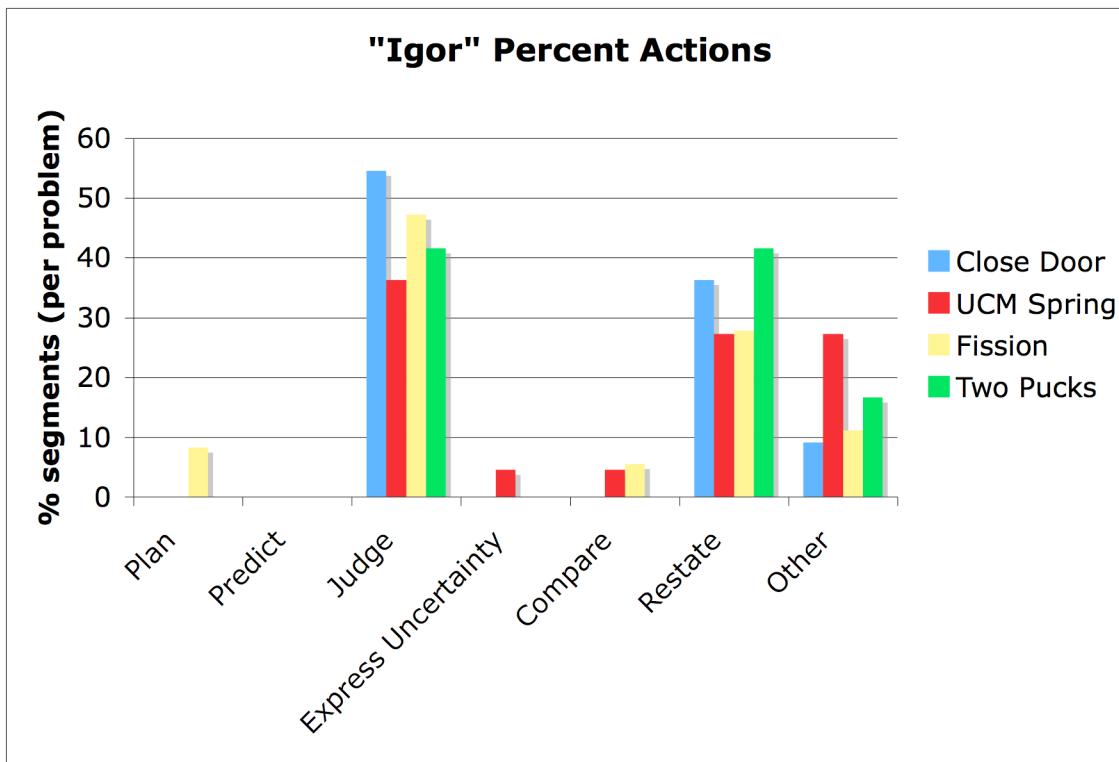


Figure 4.5 - “Igor’s” percent actions (note the concentration upon Judging, Restating, and Other)

This group of “judgers”, consists of “Igor”, “Omar”, “Christobal”, “Earl”, and “Winfred” (see figures 4.6 – 4.9).

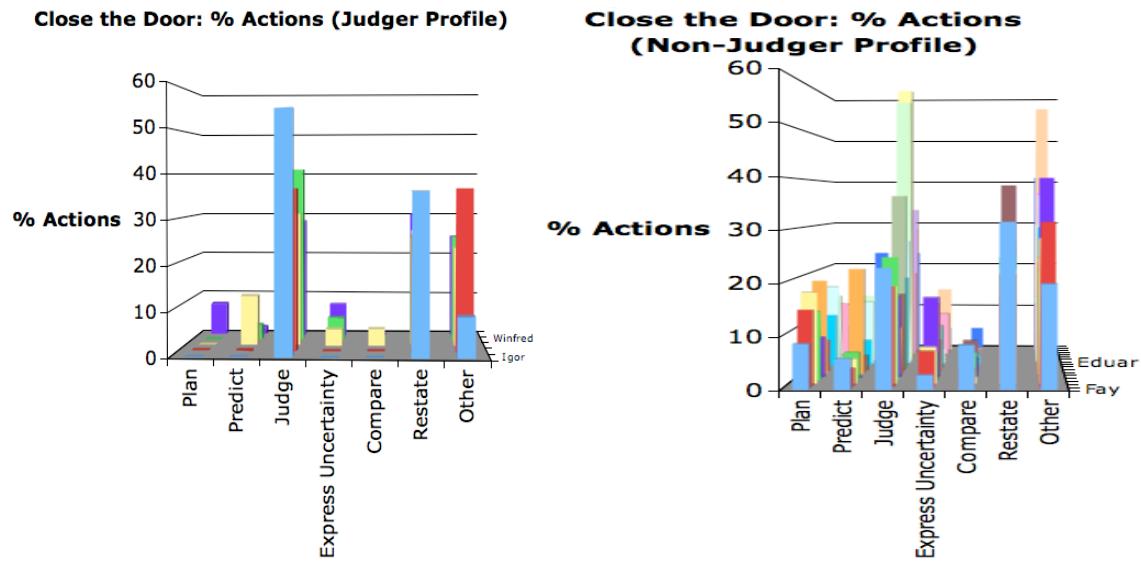


Figure 4.6 - Judgers vs. Others Actions for “Close the door” problem

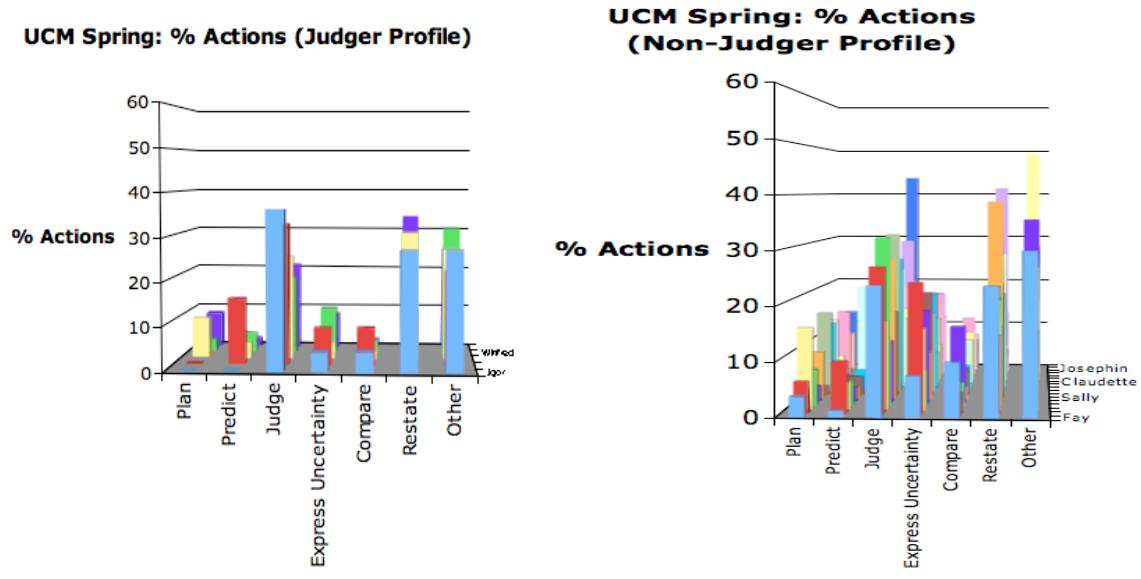


Figure 4.7 - Judgers vs. Others Actions for “UCM Spring” problem

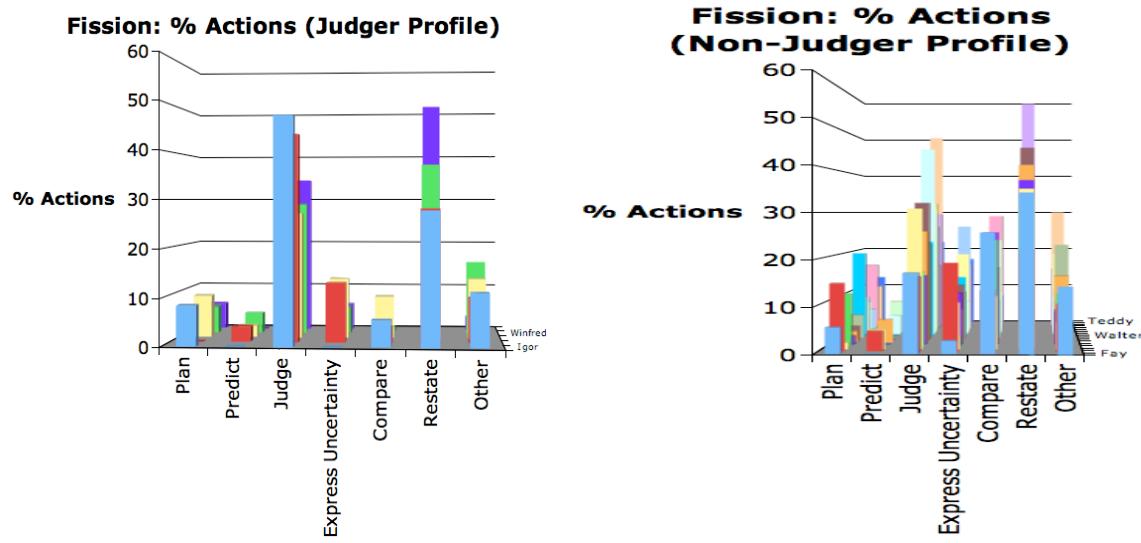


Figure 4.8 - Judgers vs. Others Actions for “Fission” problem

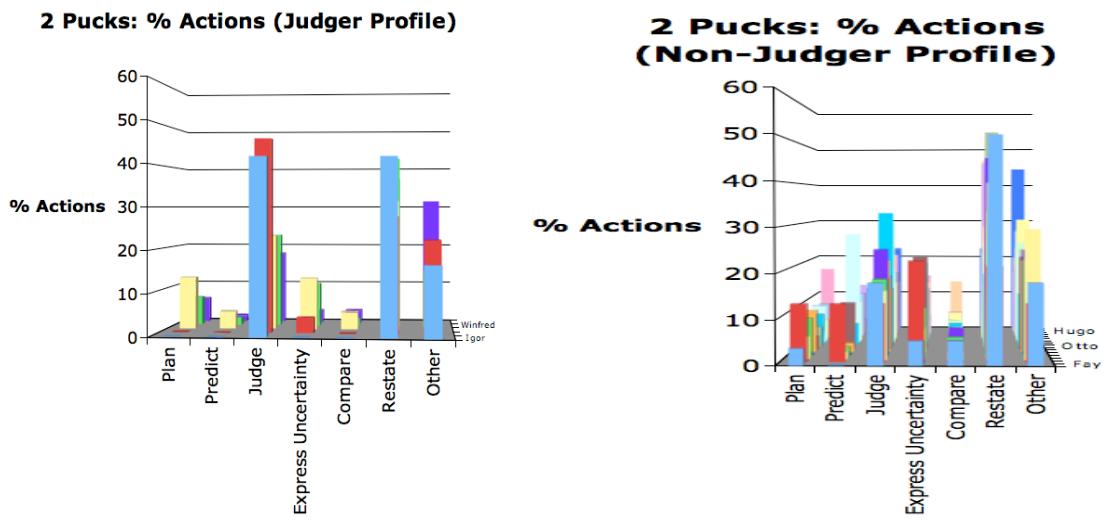


Figure 4.9 - Judgers vs. Others Actions for “Two Pucks” problem

In order to be placed into the “judger” group, the participant’s profile must have the Judging percent action at least 7% or higher than the Planning, Predicting, Expressing Uncertainty, and Comparing actions for all of the problem sets (i.e., tend toward approaching the task through interpreting the given information and judging whether that

information is correct or not). By comparison, the non-judger profiles have more prominent levels of Planning, Prediction, Expressing Uncertainty, and Comparing (i.e., the “noiser” plots in figures 4.6 – 4.9, representing a more varied approach to the task). Such a distinction identifying the judges yields significant differences from the other participants when comparing the average actions across all problems (see table 4.1 for Whitney-Mann U-test results).

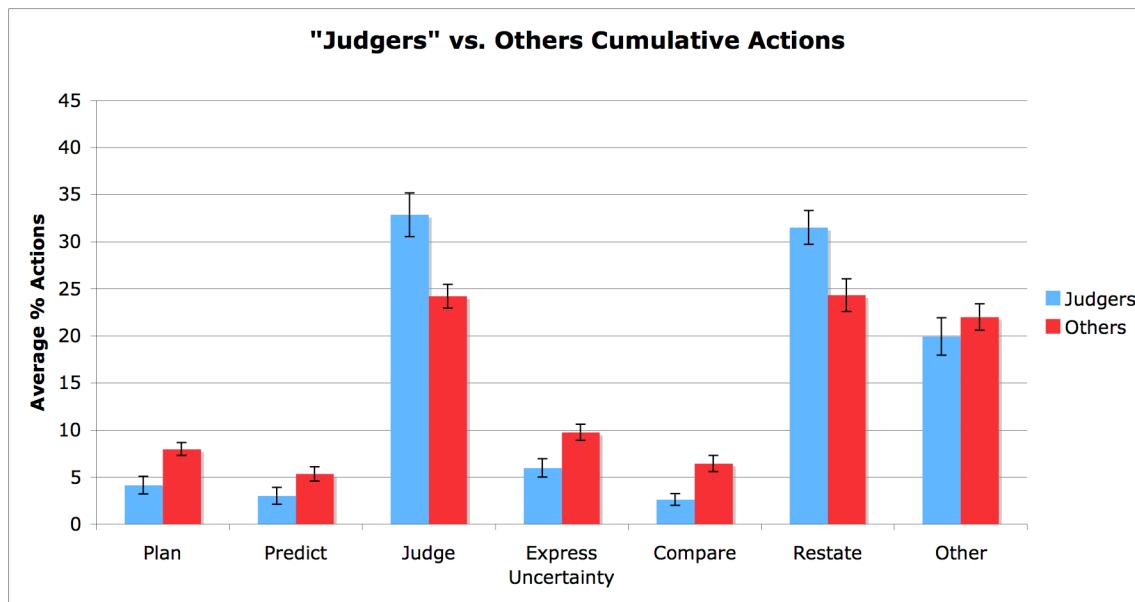


Figure 4.10 - Judgers vs. Others actions for all problems

Table 4.1 - Whitney-Mann U-test results for Judgers vs. Others average percent actions for all problems

Action	Plan	Predict	Judge	Express Uncertainty	Compare	Restate	Other
U	391	487	322.5	432	405	379	616.5
p	0.008	0.109	0.001	0.029	0.013	0.006	0.807

4.2.2: Distinctions across Problem Sets

Yet, when comparing the judges against the other participants for each problem separately, distinctions unique to each problem arise.

For the “Close the door” problem (see figure 4.11), Planning (Whitney-Mann, $U=12$, $p=0.019$) and Judging (Whitney-Mann, $U=15$, $p=0.040$) are significantly different.

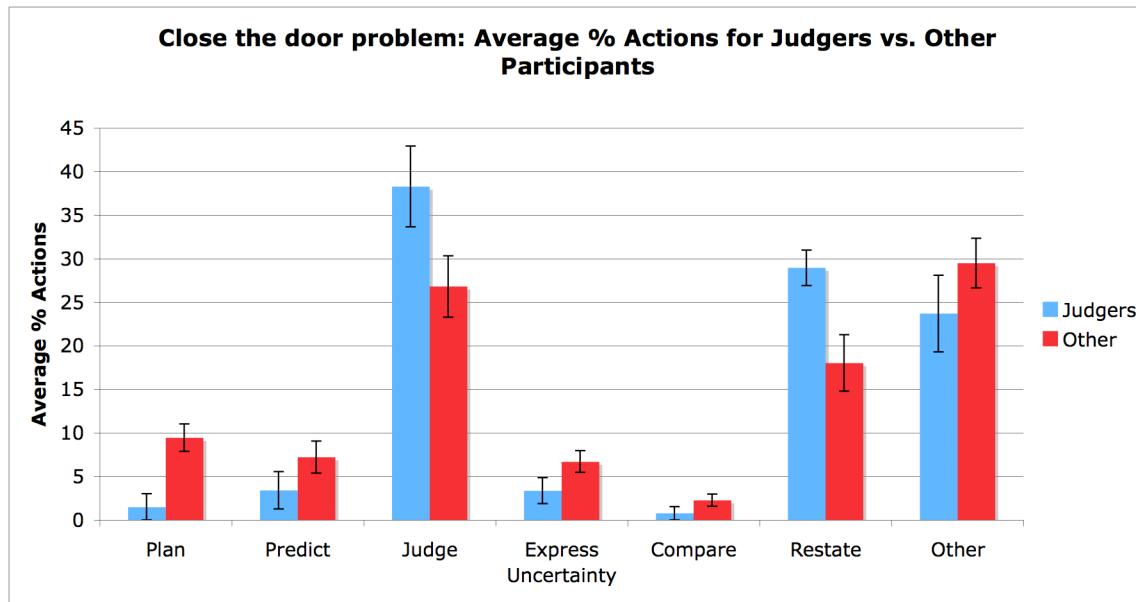


Figure 4.11 - Judgers vs. Others actions for the “Close the door” problem

As expected, the Judging action is more prominent for the judges. But Planning stands out for the other participants. Generally, there are three main trends to the Planning segments (see table 4.2).

Table 4.2 - Planning segments from “Close the door” problem.

Planning Trend	Example Segments of the Planning type
Overall strategy	<p>“Okay. I'm starting with the shortest one.”</p> <p>“I'm going to sort the solutions by length. In fact I'm doing that right now. I'm going to read the shortest solution first, then the longer solutions later, and then -- that's the way we're going to do it.”</p> <p>“Um, I am just reading through, looking for which one will basically say what I just said. I'm reading them”</p>
Denoting a single step	<p>“So I guess what I'll do first is draw a picture.”</p> <p>“but let's see if I can find something on the constants sheet.”</p>
Stating Mental Processing	<p>“I'm just reading these in my head. I can understand them better.”</p> <p>“Okay. I'm trying to remember 205 [course number for the introductory mechanics course]. Inelastic and elastic.”</p> <p>“So I'm going to say that, just thinking about real-world situations”</p>

One might argue that the prominent Planning action follows in response to prompting by the researcher to keep talking. However, 63 out of 69 such planning segments were unprompted. Another hypothesis stemmed from the actions of “Sally.” Initially, she tried working the problem on her own (never referring to the given solutions). After being reminded of the task, she expressed surprise that she could have looked at the given solutions (she then proceeded to complete the task), but this confusion only arose with 2 out of the 14 non-judger participants who planned while working in the “Close the door” problem. The remaining hypothesis revolves around the fact that the “Close the door” problem was the first problem used in the study, and so the prior instruction and/or warm-up exercises may have biased the participants’ actions, where such an effect faded as the participants worked through the subsequent problems (or perhaps as the participants gained more familiarity with the task as they gained practice in working through the initial problem and solution set). Future work can test this hypothesis by varying the order of the problem sets to disentangle the possibility of problem specific effects connecting the “Close the door” problem specifically to the Planning action as opposed to possible effects due to task familiarity or task instruction/warm-up exercises.

“Fission” and “UCM Spring” are the only two problem sets that yield statistically comparable actions for the Judgers as compared to the other participants (see figures 4.12 – 4.13, see tables 4.3 – 4.4 for Whitney-Mann U-test results).

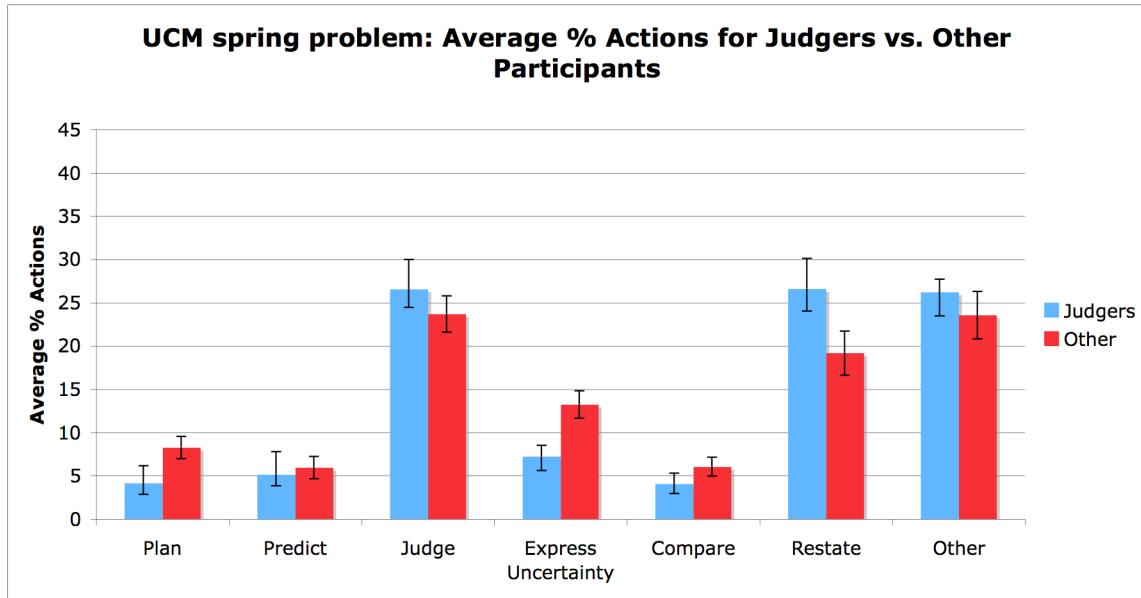


Figure 4.12 - Judgers vs. Others actions for the UCM spring problem.

Table 4.3 - Whitney-Mann U-test results for Judgers vs. Others average percent actions for “UCM spring” problem.

Action	Plan	Predict	Judge	Express Uncertainty	Compare	Restate	Other
U	28	37.5	29	23	32.5	19.5	35.5
p	0.179	0.548	0.400	0.075	0.313	0.091	0.719

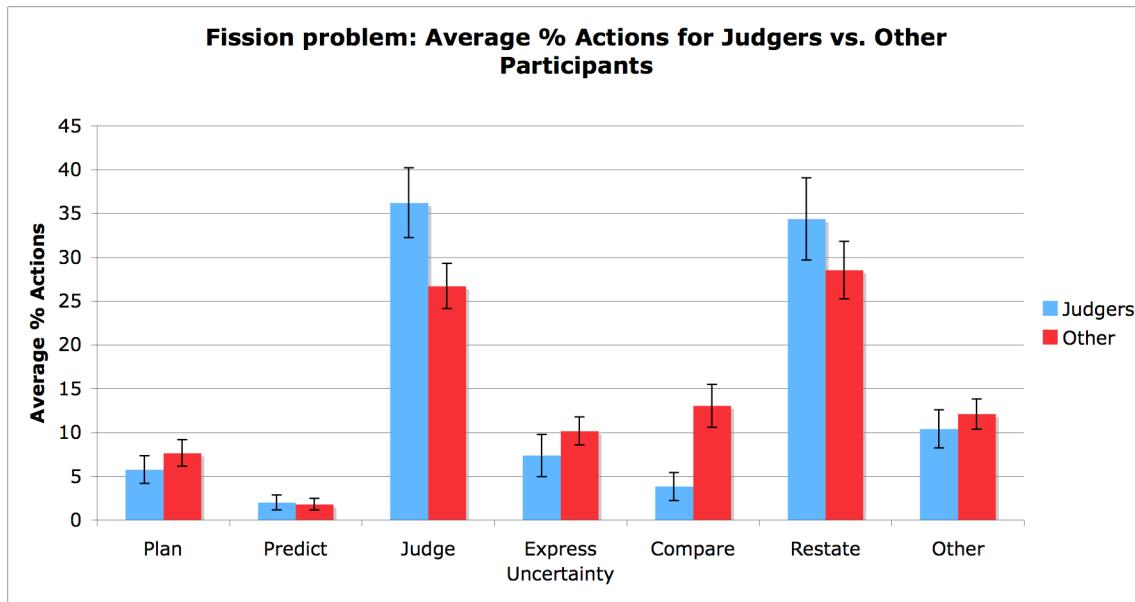


Figure 4.13 - Judgers vs. Others actions for the “Fission” problem

Table 4.4 - Whitney-Mann U-test results for Judgers vs. Others average percent actions for “Fission” problem

Action	Plan	Predict	Judge	Express Uncertainty	Compare	Restate	Other
U	39.5	37	20	34	22	26	41
p	0.660	0.842	0.109	0.400	0.062	0.275	1.000

“Fission” and “UCM spring” problems share several characteristics. Both problems are in the same “mold” as typical end-of-chapter textbook problems common to the Matter and Interactions introductory mechanics course. Indeed, the “UCM Spring” problem largely stemmed from a question in a final exam of the introductory mechanics course. Similarly, the “Fission” problem is essentially an energy problem from the

introductory course's textbook (Chabay and Sherwood, 2002). Both problems utilized solutions formatted in a fashion consistent with that used in the introductory course's textbook. In addition, only "Fission" and "UCM Spring" called for numerical final answers (more typical for end-of-chapter problems), whereas "Close the Door" and "Two Pucks" required more atypical, non-numerical results. Furthermore, "UCM Spring" and "Fission" utilize only one fundamental principle (Newton's second law or the momentum principle and the energy principle respectively), whereas the given solutions for "Two pucks" and "Close the door" varied the fundamental principle used in each solution.

While such results are consistent with the possibility that the participants might be activating a common schema for working with these typical end-of-chapter style problems, the analysis indicates that the convergence of the judges and the other participants is not sufficiently above the statistical "noise" to clearly indicate whether the judges' actions tended to shift toward the other participants, and/or whether the other participants tended to shift toward a typical judge profile. For each participant, the percent actions for the "UCM Spring" and "Fission" problems were averaged and the percent actions for the "Close the door" and "Two Pucks" were averaged. The "UCM Spring" and "Fission" average was subtracted from the "Close the door" and "Two Pucks" average on an individual participant basis, and the absolute value of the difference was computed (to determine the shift in actions). Figure 4.14 depicts the average shift for each group: judges and the other participants. A Whitney-Mann U test was used to determine if the shifts of the judge group versus the other participants was significantly different. No significant differences were found (see table 4.5).

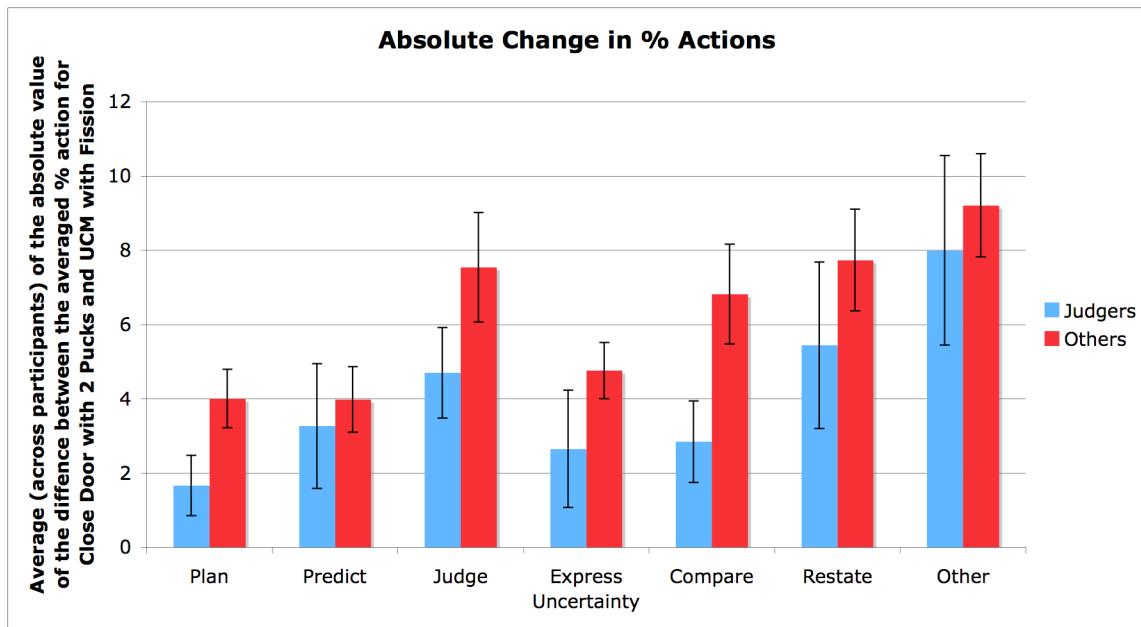


Figure 4.14 - Judgers vs. Others absolute change in average % actions for Close the door and Two Pucks minus UCM Spring and Fission

Table 4.5 – Whitney-Mann U-test results for Judgers versus Other absolute change in actions for “Close the door” with “Two Pucks” against “UCM” and “Fission”

Action	Plan	Predict	Judge	Express Uncertainty	Compare	Restate	Other
U	25	39.5	37	27	25	35	41
p	0.109	0.660	0.548	0.153	0.109	0.445	0.780

“UCM Spring” and “Fission” also yield lower instances of incorrect judgments (Kruskal-Wallis, H=13.88, p=0.0031).

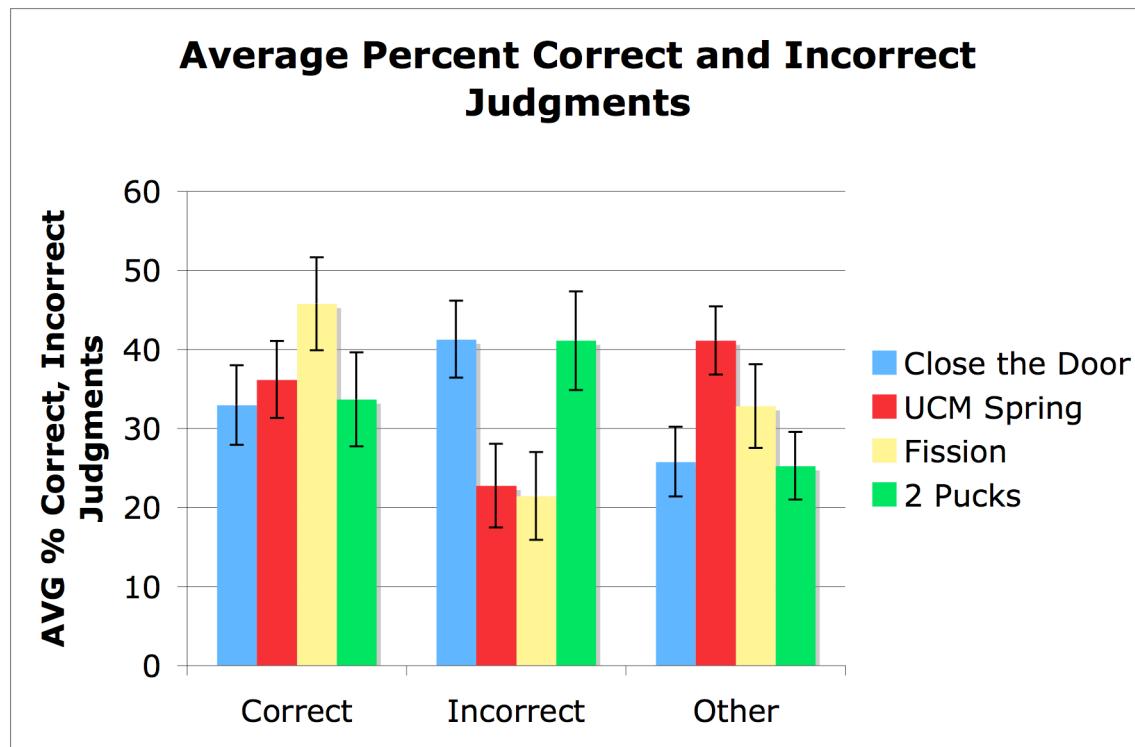


Figure 4.15 - Average percent correct and incorrect judgments

However, the context results suggest that “UCM Spring” and “Fission” do not necessarily standout for the same reasons. When analyzing the incorrect judgment segments, “UCM Spring” exhibits the highest bias to a non-physics context.

Percent Incorrect Judgment Context for UCM Spring

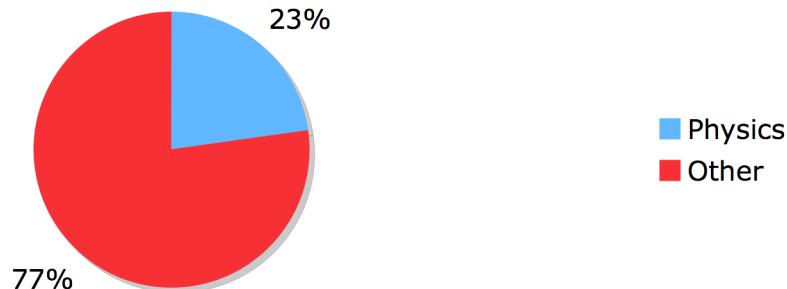


Figure 4.16 - Percent incorrect judgment context for “UCM”

Percent Incorrect Judgment Context for Fission

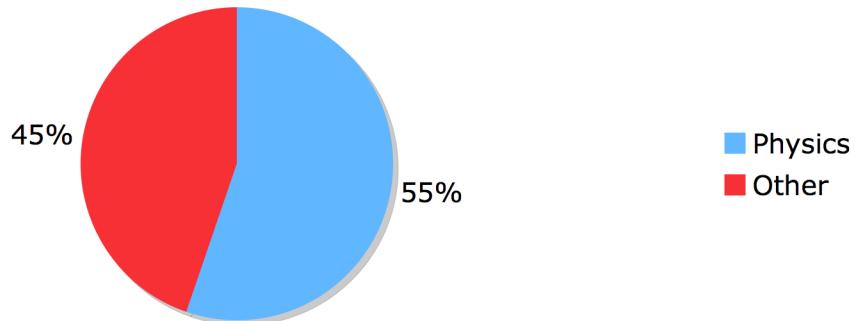


Figure 4.17 - Percent incorrect judgment context for “Fission”

Percent Incorrect Judgment Context for Close the Door

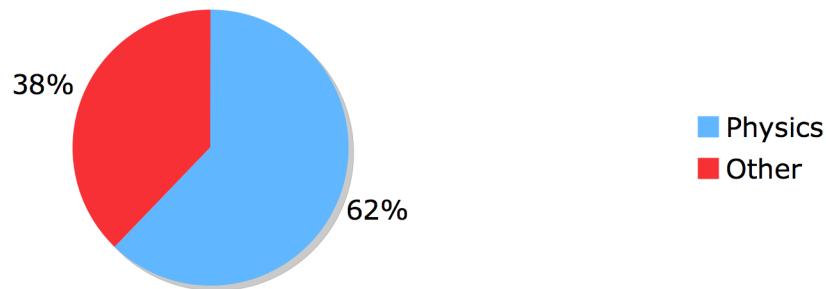


Figure 4.18 - Percent incorrect judgment context for “Close the door”

Percent Incorrect Judgment Context for 2 Pucks

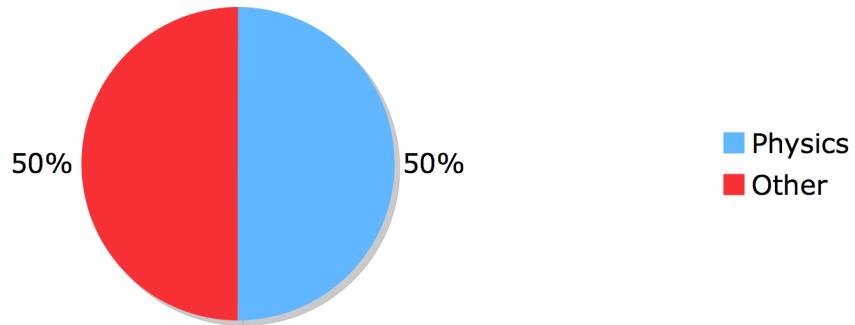


Figure 4.19 - Percent incorrect judgment context for “Two Pucks”

Indeed, the overall analysis of the judgment segments yield consistent results with “UCM Spring” having the lowest levels of physics context (Kruskal-Wallis, $H=20.5$, $p=0.0001$), while “Fission” is more comparable to the other problems.

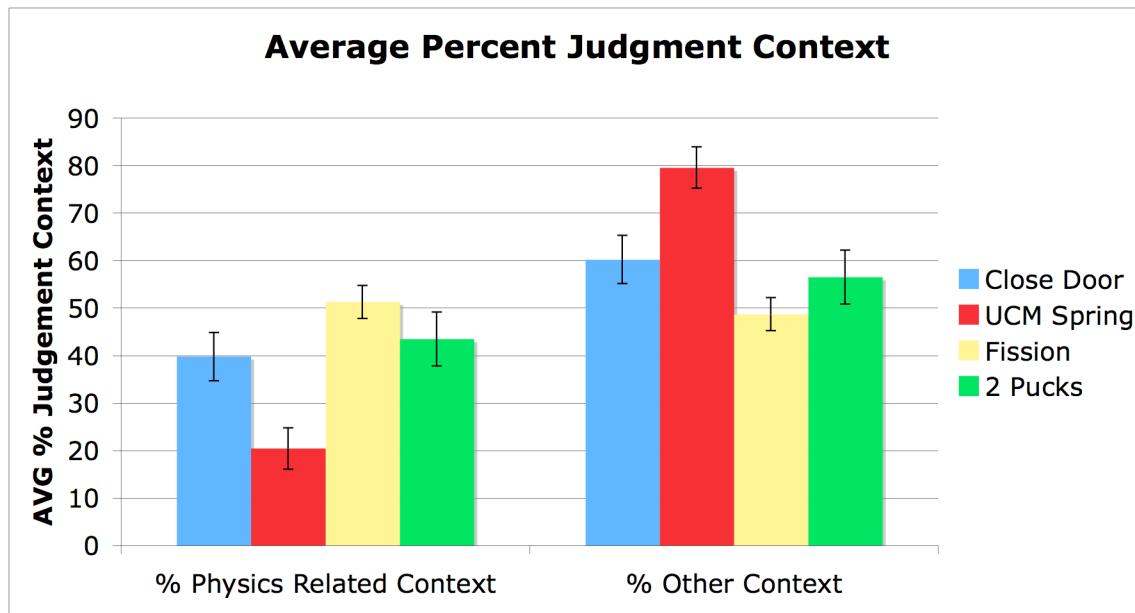


Figure 4.20 - Average Percent Context of the Judging Segments

At a more detailed level, “UCM Spring” exhibits a distinctive spike with judging mathematics (Kruskal-Wallis, $H=15.49$, $p=0.0014$).

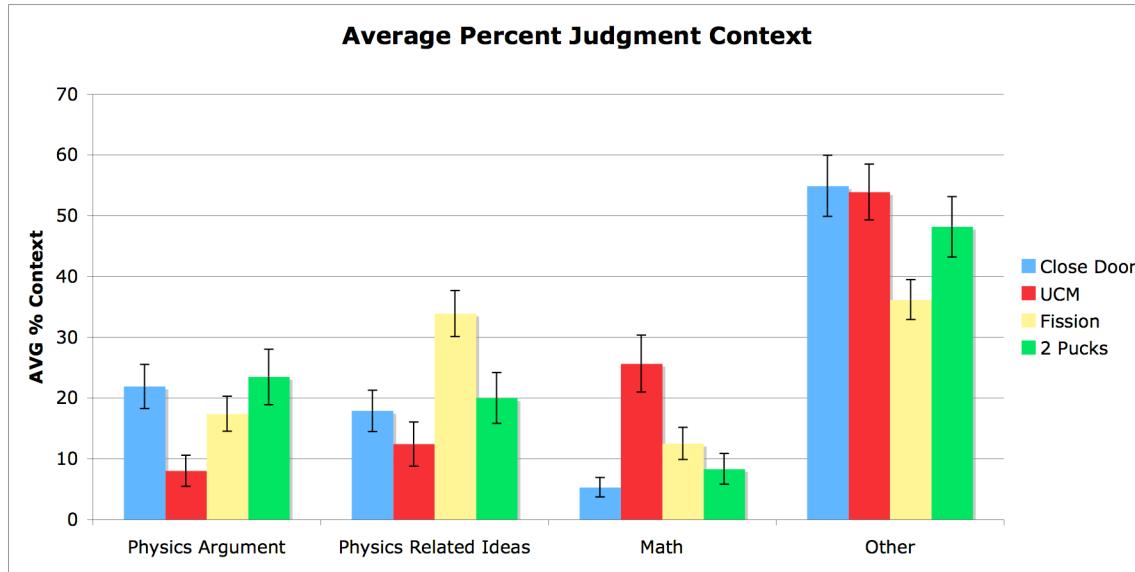


Figure 4.21 - Detailed Average Percent Context of the Judging Segments

One might argue that perhaps the “UCM Spring” problem simply possesses more math “moves” to consider, causing the increased bias to math focused judgments. Using the number of equal signs (“=”) in each solution as a rough measure of the number of math moves, problem is not a significant predictor for the average number of equal signs across each problem’s solutions (Kruskal-Wallis, $H=7.24$, $p=0.0646$).

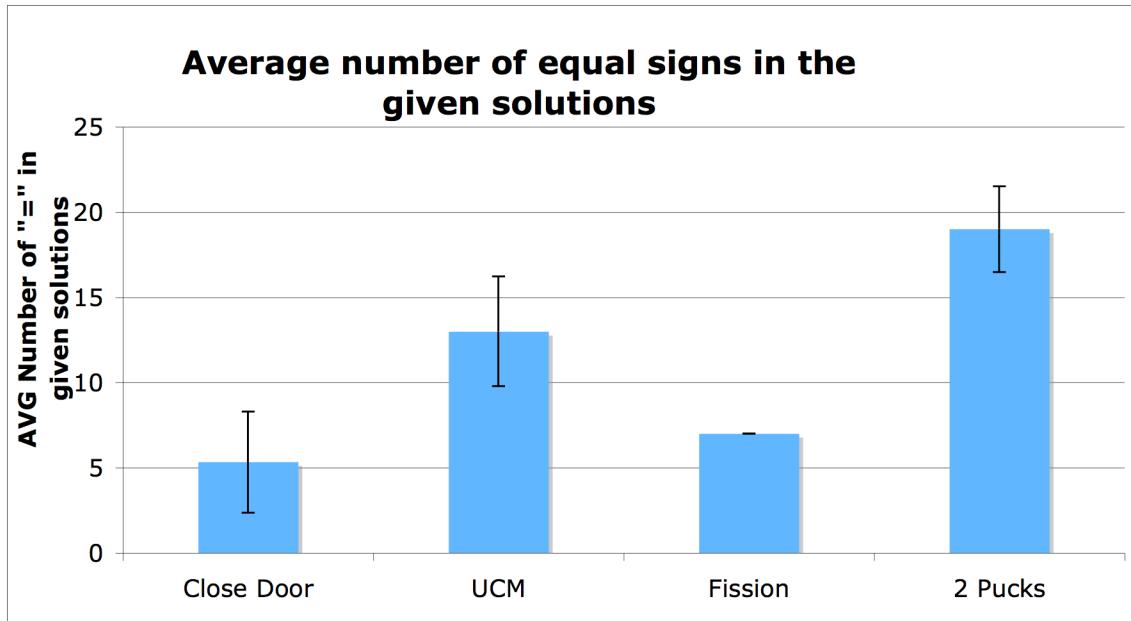


Figure 4.22 - Average number of equal signs in the given solutions for each problem

The math-focused judgments were found to have three main trends: judging the relation of pv/r as the radial component (for uniform circular motion) of the time rate change of momentum, judging a line of mathematical argument, or judging the variable “ s ” (only found with the graduate participants). Much of the pv/r discussion is tied with another, approximate form of this equation, mv^2/r , $v \ll c$ (both forms of this equation are provided in the reference sheet). While the Matter and Interactions (M&I) curriculum generally uses pv/r as the underlying relationship, it is not uncommon to see the mv^2/r approximation arise in other curricula. It is interesting that only non-M&I trained participants (the graduate participants) debated the meaning of “ s ”, which the M&I curriculum typically uses as the variable for stretch (perhaps taken for granted as such by the M&I trained participants, hence they did not bother debating “ s ”). However, possible

influences due to the undergraduate vs. graduate split cannot be discounted (perhaps the more experienced graduates were better positioned to uncover the incompatibilities of defining “s” to be stretch vs. the manner in which “s” was used in the incorrect solutions). Of course, future work can readily disentangle such effects by including splits across both graduate vs. undergraduate participants and M&I vs. non-M&I training.

Table 4.6 - Math Context Judgments for the “UCM Spring” problem

Math Context Trend	Example Segments of the Math Context
Judging pv/r	<p>“...and the MV squared over R ties in with my plebeian knowledge of non-M & I mechanics, so I feel that they start off in a good place.”</p> <p>“No, they didn't. PV over R is correctly MV squared over R, “</p> <p>“...I don't think this one is going to be right because it should be MV squared over R, not MV. So that's MV squared over R, yeah, that shouldn't be there.”</p>

Table 4.6 (continued)

Judging a line of mathematical argument	<p>"I think the derivative of a cosine is a minus omega squared. That math is nice."</p> <p>"Let me double check that, 2 pi divided by 1.5, yeah equals 4.19, KS equals W squared M, M is 0.5, so this equation is looking -- this one is looking like it is the right answer (using the calculator to double check the math) 0.19 x 0.5, 2.45, oh, 4.19 squared, ... all the numbers and work are looking right."</p> <p>"I just want to say it's, the derivative rules seem right, but at the same time there seems like there's a lot of room for error."</p>
Judging "s" (graduate participants only)	<p>"Oh, yeah. Oh! The mistake is here in the last step, where they say A cosine Omega T. goes like S..."</p> <p>"R .6 m. Yeah, oh, that's the, that's the problem, they are not making a distinction. They are defining -- this is wrong, they're defining S both is the difference and as the absolute length, so solution number three is wrong. Ah, Success. They poorly defined S in trying to trick me, but, I think I figured them out."</p>

In contrast, the incorrect judgments for “Fission” are biased toward a physics focus. In particular, the judgments have two primary trends: the selection of the final and initial states, and the energy principle expansion (see table 4.7).

Table 4.7 - Physics Context Judgments for the “Fission” problem

Physics Context Trend	Example Segments of the Physics Context
Final and Initial State Selection	<p>"Solution three the final state is wrong because, because if, because if we get positive charges on each of them, then they should push away from each other, they should not just, oh wait, they could just sit there."</p> <p>"So I would guess that the initial state is because it's totally, it's not number three, I'm pretty sure."</p> <p>“And I will say that this one is wrong because two positively charged particles aren't going to, hang on that close to each other without any opposing force,”</p>

Table 4.7 (continued)

Energy Principle Expansion	"Though I think that there is potential energy initial actually, there is initial potential energy but no final potential energy. So that is one of the errors here." "So, charge and rest mass. After the fission process, hmm, and they are breaking apart so, I would have to say, I don't think they would have a final potential just because since they are moving apart before hand, like before they would have more potential energy initial and here they do have the potential initial still in the problem (pointing to solution number two). So, I'm going to go solution number two so far based on that." "Just the rest energy, yeah I am going to go with two just because I feel like they have more of a potential energy at the beginning than once they're apart, then the potential energy there isn't -- there should not be any final potential energy,"
----------------------------	--

Both the final and initial state selection and the energy principle expansion were varied for the “Fission” problem’s solution. As noted in section 4.1.1, such variations tended to be associated with participant discussions of the varied features (suggesting that feature variation can be used as a tool to direct participant attention).

However, “UCM Spring” and “Fission” do share an intriguing characteristic that may be related to the lowered levels of incorrect judgments. “UCM Spring” and “Fission” lacked tempting distracter (incorrect) solutions of the type evident in the other problem sets (“Close the door” and “Two pucks”). A common perception associated with the “Two Pucks” problem set is the apparent need to include rotational motion quantities in the solution. In reality, to answer the question posed in the problem, only the translational motion need be analyzed. Hence, the correct solution was commonly rejected due to the “intuitive” notion that rotational motion “was not accounted for.” (see table 4.8) Two of the three given solutions for the “Two Pucks” problem set integrated (incorrectly) rotational motion.

Table 4.8 - Example segments of participants rejecting the correct solution to “Two Pucks”

Example Segments of the “intuitive” argument for “Two Pucks” problem set
"So they don't really factor in the rotation. I don't think that this one is going to be right."
"I don't feel like solution three is it just because they are not using -- they are not taking into account the rotation of the second puck, so I definitely do not think it is three."
"I don't think the forces would be equal, because in one case here we are, here we are not taking the angular momentum into consideration, and in the puck two there is rotational motion as the puck, as the thread unwinds around the puck, and therefore we have to take the rotational kinetic, rotational kinetic energy, and the angular momentum into consideration, so I think this is, this solution is simplistic, this is going to be incorrect"
"I would like to go with solution two, because the fundamental principle, the fundamental physical quantity that is involved in the second puck has been used here in the angular momentum principle."
"Well, yes, but there should be another term in there. What about the angular momentum? And you are not even bothering with that, are you? Well, this equation is wrong because it doesn't include angular momentum. So, that makes sense. That's where you went wrong."
"This one I feel very comfortable saying is wrong because ... solution number three didn't include rotation."

The second solution in the “Two Pucks” problem set used arguments based on the angular momentum principle to analyze the motions of both (rotating and non-rotating)

pucks, while the third solution used the linear momentum principle for both pucks. Interestingly, the participants tended to express concern about using the third solution's argument based upon linear momentum to analyze the motion of the second (rotating) puck **without** expressing the inverse concern of using the second solution's angular momentum approach to analyze the motion of the first (non-rotating) puck. "Claudette" does provide an illuminating comment that is perhaps at the heart of this issue. She notes that translational angular momentum may be applied to the non-rotating puck ("*I like solution two just because I felt like you have to, you are going to have to use the angular, the angular momentum principle since one is turning and the other one, since you are just pulling it from the center it is just sliding so it would be L trans.*"). Perhaps the concept of translational angular momentum negated concerns of using angular momentum to analyze the non-rotating puck.

Similarly, each solution for "Close the door" used a different fundamental principle for their arguments. The first solution in the set used an energy principle based argument. The participants tended to lean toward this energy-based argument (over half of the participants chose the energy-based solution as their final answer), or conversely tended to discount a solution based upon its fundamental principle (see table 4.9).

Table 4.9 - Physics Context Judgments for the “Close the door” problem

Physics Context Type	Example Physics Context Segments
Energy Principle focused arguments	<p>"Clay works, because it transfers the kinetic energy into the door, because it just stays there,"</p> <p>"Hmm, I don't think, because, yeah, if the clay does hit the, the door and it sticks, it should, all of the energy should be transferred to the door, which moves it back, but since like the ball is elas-- um, it's hollow and elastic, it should bounce off the door and come back, and not really transfer the energy to the door. So that might be right. "</p> <p>"I mean, that's what I would have chosen, just, if I remember, I don't believe they are equal, just because the tennis ball will, like it hits and then it bounces back, whereas the clay is going to apply more of -- put more energy on it."</p>

Table 4.9 (continued)

Discounting a solution based upon Fundamental Principle	"I don't think it has anything to do with angular momentum." "I don't think it's angular momentum just because it just doesn't make sense to me why it would be that, because that thing is going, like, so I don't think it's -- ugh -- let's see, (participant takes up the red pen). Well I definitely don't think its angular momentum, just because it doesn't seem relevant to the situation, because it's like a collision and I don't believe, if I remember correctly, that it would have anything to do with it." "Angular momentum? That doesn't really have anything to do with it."
---	--

The data further suggests that aside from the energy principle argument possibly representing an “attractive” solution, prior instruction may represent an interfering factor. In particular, “Teddy” noted, “*Umm, That's not it, I don't think because when I did this problem before we did not need to know, like, angular momentum. So I don't think that's it, that's my reason for that.*” “Teddy” claims that he has worked this problem before; however, he recalls seeing it **prior** to learning about angular momentum (in reality, the correct solution in this study did indeed use an angular momentum-based argument).

Hence, “Teddy” believes that the correct solution would not use angular momentum (and rejects the correct solution).

Overall, the results suggest that the presence of a tempting incorrect solution (an “intuitive” or otherwise appealing solution) is associated with more incorrect judgments. However, both “Close the door” and “Two pucks” also had variations in the fundamental principle. Future work is required to disentangle both effects and their corresponding influence upon the level of incorrect judgments. While the lowered incorrect judgment level for “Fission” further suggests that variations (and errors) in the final and initial state selections as well as errors in the principle expansion are associated with less incorrect judgments (in comparison to the aforementioned interference of a tempting incorrect solution). In addition, the lowered incorrect judgments with “UCM Spring” coupled with the higher level of math focus suggests that the participants are making less errors on the mathematics as opposed to the physics (as some might expect). Of course, future work is required to fully disentangle all problem specific effects, which is not the intent of this study. For example, “Fission” was also the only problem dealing within the microscopic realm (a participant from earlier pilot work had commented that such problems are difficult to mentally picture).

“Fission” also stands out when comparing the problem sets across overall average actions (see figure 4.23).

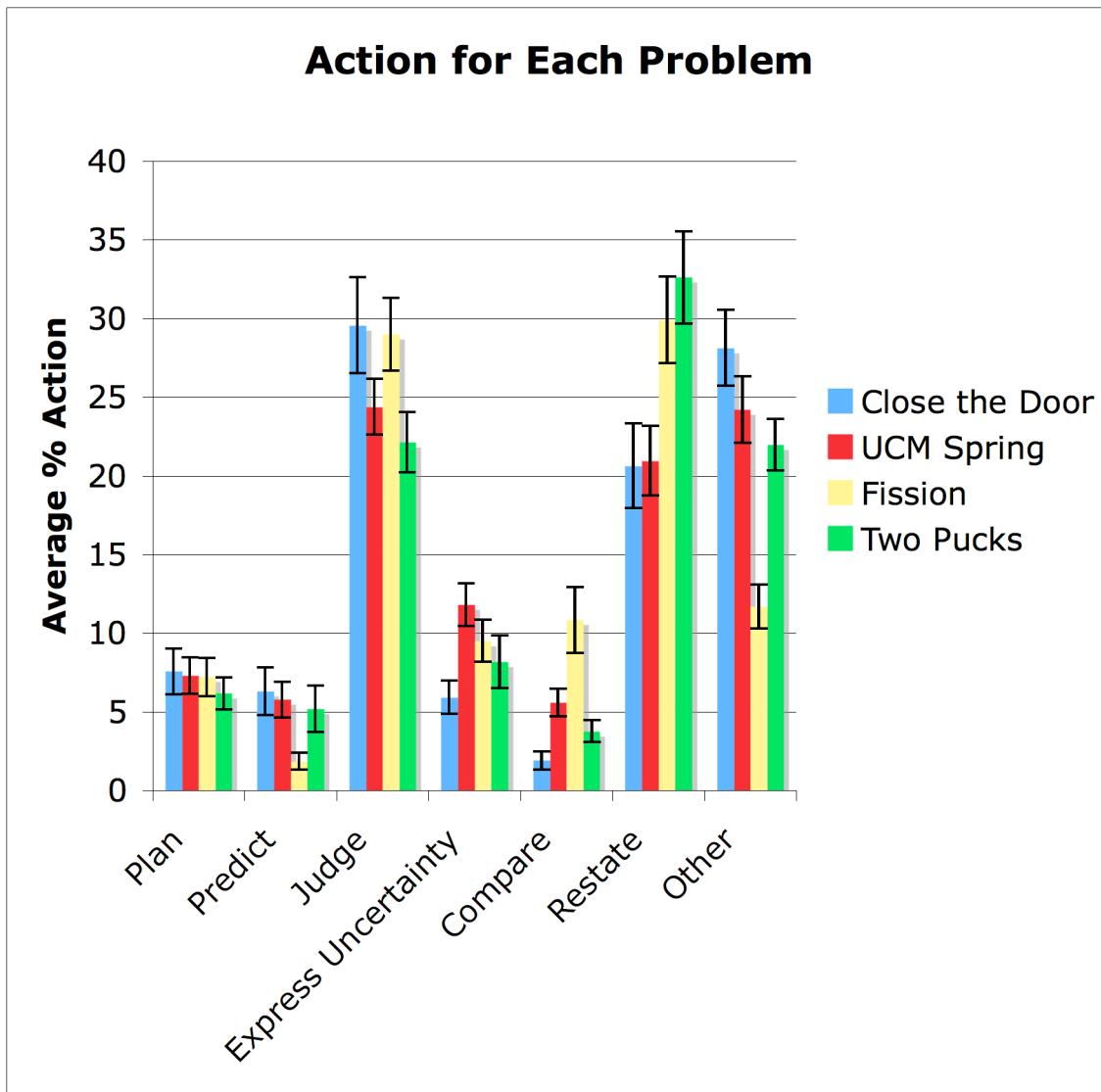
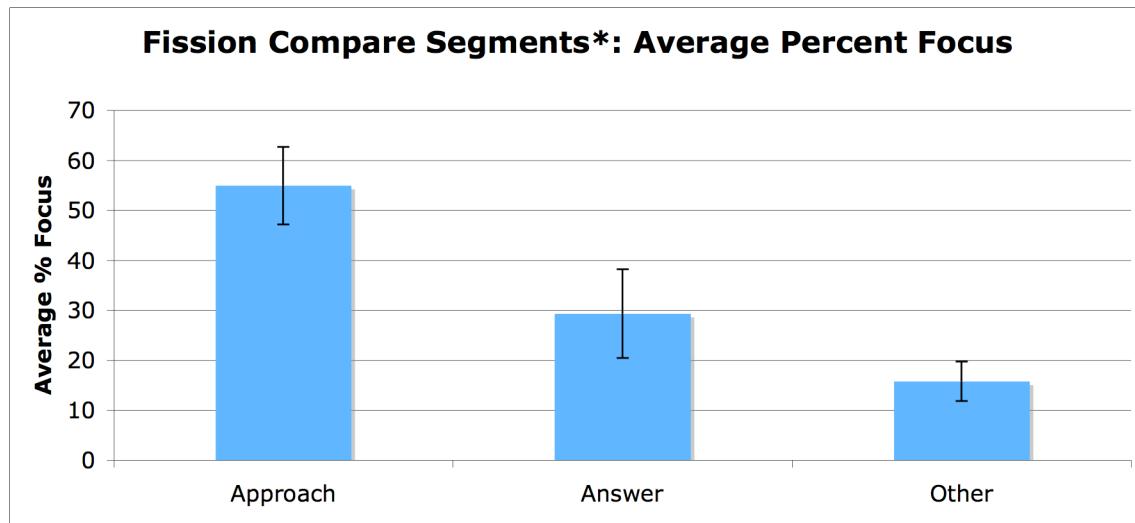


Figure 4.23 - Average percent actions for each problem

Coupled with lowered prediction (Kruskal-Wallis, $H=10.05$, $p=0.018$), “Fission” is associated with higher levels of Comparison (Kruskal-Wallis, $H=18.87$, $p=0.0003$). The majority of the Comparison segments are focused upon the approach leading to the answer as opposed to comparing the answers themselves (Whitney-Mann, $U=97.5$, $p=0.040$).



*Excluding those participants who had no Compare statements for the Fission problem

Figure 4.24 - Average percent focus of comparisons

“Fission’s” approach is unique in that all of its solutions explicitly denote the final and initial state selections using diagrams. However, the data does not suggest that the state selections are necessarily driving the comparisons of the approach. 12 of the 18 participants, who had comparison segments with the “Fission” problem set, did not explicitly compare the state selections. In hopes of finding what feature might be instigating the comparisons, the initial comparison segments were characterized. Unfortunately, no discernible pattern emerged. Roughly two-thirds of the initial comparisons occurred prior to the principle expansion. These initial comparisons spanned a large spectrum of steps, including: noting the use of the energy principle, system and surrounding selections, final and initial state selections, and principle expansions. The remaining one-third of the participants initially compared the mathematical development

following the principle expansion. Overall, the lowered prediction coupled with higher comparison suggests that the “Fission” problem set is activating a different “mode”, but the data does not clearly indicate what might be activating this “mode”, which is not the focus of this research.

The final problem set, “Two pucks”, yields an interesting distinction between the graduate and undergraduate participants (consequently, also between the M&I and non-M&I trained participants). The undergraduate participants tended to be more biased toward a physics-based context when judging (Whitney-Mann, $U=15.5$, $p=0.018$).

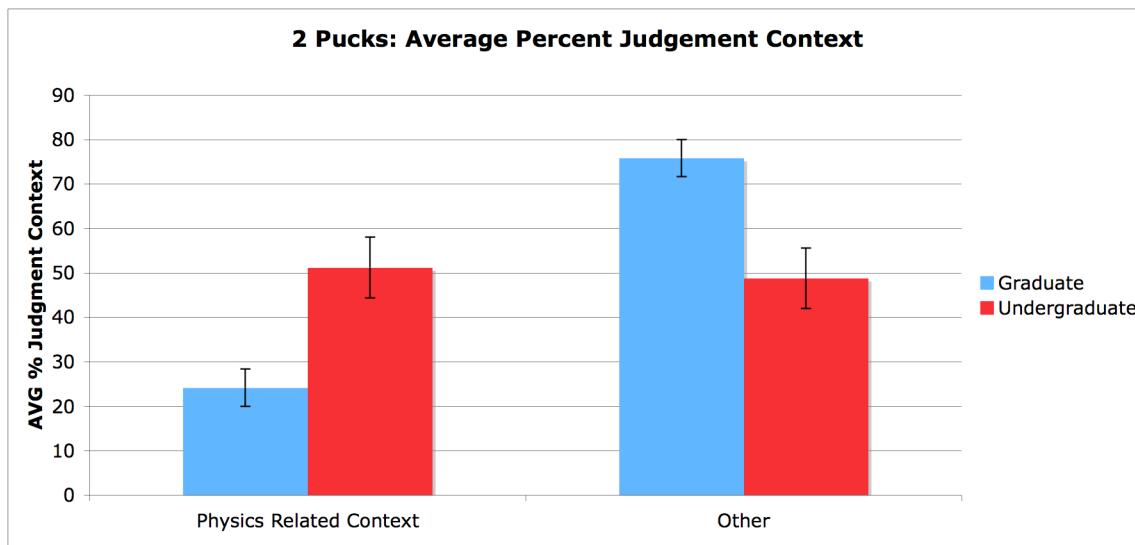


Figure 4.25 - Average percent context of judgments for “Two Pucks”

The detailed data reveals that the graduates’ bias is primarily based upon the math context (Whitney-Mann, $U=15$, $p=0.018$).

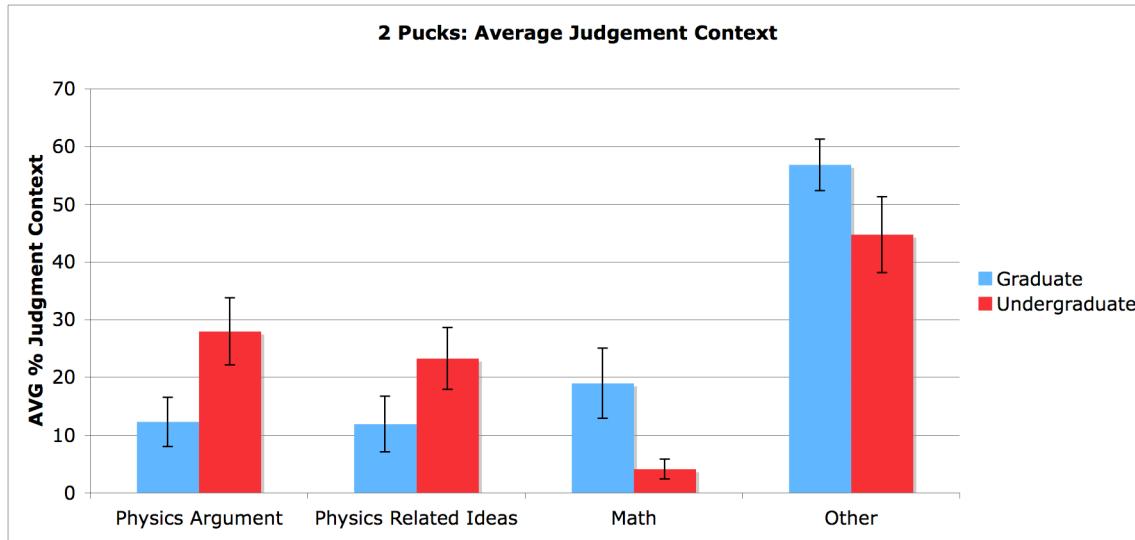


Figure 4.26 - Detailed average percent context of judgments for “Two Pucks”

This bias suggests a possible M&I training issue (the M&I trained participants might be focusing upon distinct physics items emphasized in their instruction), or perhaps stems from the difference in experience between the graduates and undergraduates. To a degree, the analysis of the judgment segments lends itself to both propositions. Two types of judgments were unique to the undergraduates: judging the system and surroundings selection and judging based upon a preconceived answer.

When the system and surroundings (a topic emphasized in the M&I curriculum) judgments were removed from the analysis, the difference between the graduate and undergraduate judgment context biases did shift to become marginally insignificant (Whitney-Mann, $U=21.5$, $p=0.066$). Of course, as noted above, judgments of this type does not solely represent the differences between the graduate versus undergraduate

participants. Instead, the system and surroundings judgments simply represent one difference between the participants for the “Two Pucks” problem.

Given that the system and surroundings selection is a topic emphasized by the M&I curriculum, the above results suggest a possible M&I training difference between the undergraduates and the graduates. However, 4 out of 6 graduate participants did make judgments on the system and surroundings for the “UCM spring” problem set (interestingly, for “UCM spring” there were no differences in judgment context for the undergraduates vs. graduates, just as when the system and surroundings judgments were removed from the “Two pucks” analysis). In essence, the M&I trained undergraduates more persistently judged the system and surroundings throughout the study, while the non-M&I graduates did not.

The other judgment distinctive to the undergraduate participants is judging based upon a preconceived answer. For the “Two Pucks” problem, 4 out of 15 undergraduates made a judgment based upon their belief as to the correct relationship between the forces acting on the pucks (see table 4.10).

Table 4.10 - Example segments of undergraduates judging “Two Pucks” based upon a preconceived answer

Undergraduates judging “Two Pucks” based upon a preconceived answer
<p>"That doesn't make sense because it would take more force to get the string to unwrap around it then it would if we were just to pull the thing forward. So f1 needs to be less than f2. Making that one wrong."</p> <p>"Um, let's see, (inaudible) I'm going to guess that that one is wrong, because it says that it requires a larger force"</p> <p>"I think I'm going to go to solution three. I still, I still think it's going to be, the same, the same amount of force."</p> <p>"Yeah, I think that f2 is going to take less of a force, so I think solution one is going to be right."</p>

Table 4.10 contains judgments based upon the undergraduate participants' assertion of the relationship between the force exerted on the first puck versus the second puck. However, this result does not in turn imply that the graduates were necessarily less susceptible to the tempting, “intuitive” incorrect solutions (as discussed earlier) given that both graduates and undergraduates rejected the correct solution based upon the “intuitive” notion that the correct solution must include an analysis of the second puck’s rotational motion. Overall, only undergraduates were judging based upon an explicitly stated belief of the relationship between the forces acting on the pucks, while both participant groups made judgments based upon the apparent need to analyze rotational motion.

CHAPTER 5: CONCLUSIONS

This chapter discusses several key findings of this research. The discussion is organized into three sections:

1. What actions emerged with the task
2. Participant attention (both the effects of Feature Variation and the Flexible Representation)
3. Future work directions

Initially, the results pertaining to the core research questions are addressed. The second section discusses those results associated with participant attention. The final section details implications and avenues for future work.

5.1: Actions Summary

This study is centered upon the following task: given three possible solutions to a physics problem, select which solution is correct and identify the errors in the other given solutions. In particular, this study addresses the overarching questions:

1. “What actions emerge during the task?”
2. “Are those actions related to the deep processing identified in the existing literature?”

Overall, the actions that emerged from this task (Planning, Predicting, Judging, Expressing Uncertainty, Comparing, and Restating) are largely consistent with those beneficial actions denoted in my review of the worked example literature (see section 2.2) as well as those actions from the reading comprehension literature (Palincsar and

Brown, 1984). Particularly satisfying is the presence of the internal consistency checks, which was predicted by the Palincsar and Brown (1984) framework, but previously unseen in the worked example literature.

In addition, given the rise in reforms using incorrect solutions (for example, see WRONG Problems in Harper, et al., 2007; Evaluation Tasks in Etkina, et al., 2006; PALS in Reif and scott, 1999), little research has been focused on what processing is associated with using incorrect solutions (Große and Renkl, 2007). As noted by Große and Renkl (2007), additional research is required for the more complex realms (particularly physics). Reassuringly, the results from this research are also reasonably consistent with the more thoroughly researched area focused exclusively upon the use of correct solutions (Atkinson, et al., 2000).

5.2: Participant Attention Summary

This section discusses two intriguing results that are related to the participants' attention: Feature Variation and Flexible Representation.

5.2.1: Flexible Representation

Multiple information representations were utilized throughout the problem sets (prose, pictorial, and mathematical). The overall data largely indicates that no specific

information representation was necessarily discounted²². Those problem sets with prose information tended to have restate segments devoted to interpreting such information (similar results emerged with the pictorial and mathematical representations). Such results suggest future work exploring the use of variable information representation to direct participant attention such as directing participant attention to interpreting graphical information.

5.2.2: Feature Variation

A particularly intriguing result is the participant discussion's association with those physics-related, solution features that were varied across solutions (such as final and initial state selections). Not only does the literature denote issues with focusing participant attention to the physics aspects of worked examples (Chi, et al., 1989 and 1991; Stark, et al., 2002), but I have also experienced similar difficulties in my early pilot work, where I employed spatial cueing (as inspired by Catrambone's, 1998, written cues and Atkinson and Derry's, 2000, temporal cues) in an effort to highlight those physics related moves.

²² As noted in section 4.1, there may be a complexity threshold. For significant processing, the data suggests that basic diagrams (conveying a low amount of information) or minimal prose may not stimulate much processing.

5.3: Future Work

While this task does not yield uniformly high levels of all actions for all participants, the presence of these actions for the majority of the participants is encouraging (especially since this study took place within the complex realm of physics). In particular, such results are encouraging for future work focused upon developing an instructional intervention. Of course, any future work focused on the development of an instructional intervention must also consider ecological validity. In addition, if the intent of such future work is to cue the full spectrum of the Palincsar and Brown (1984) actions, then the “judgers” must be addressed, given their lower levels of several components of the Palincsar and Brown (1984) framework. One avenue suggested by the existing literature might lie along a training regimen in the use of those actions with lowered levels of activity (Planning, Predicting, Comparison, and Expressing Uncertainty). Aside from the instructional design elements used in Palincsar and Brown (1984), literature reporting the use of training techniques from the related worked example literature can also be referenced (for instance, see Chi, et al., 1994; Renkl, et al., 1998; Stark, et al., 2002).

Another possible avenue relevant to instructional intervention research is the organization of participants into groups. Do actions change when this task is performed in a group setting (as would be used in a laboratory or study group for instance)? If so, do participants tend to shift toward a specific group member’s profile or does the group collectively shift to a hybrid profile stemming from a combination of individual group member profiles? Will the introduction of group dynamics affect performance in

accomplishing the task correctly or change the individual participant focus? Will the use of reforms such as formal cooperative group learning (Heller and Hollabaugh, 1992) impact the results?

In terms of the secondary results (i.e., those results not directly connected to the core research question on the actions), I would particularly like to see future work focused upon the feature variation result (especially considering that this research was not specifically designed to study such an effect). As I mentioned in section 5.2.2, not only does my own work point to problems with focusing participant attention to those physics-related steps, but the literature also highlights such issues. Not only could other problem sets (not used in this study) be introduced in such work, but also the problem sets used in this study could be designed with other feature variations as a comparison to this work.

REFERENCES

- Abernethy, B., Neal, R. J., & Koning, P. (1994). Visual-Perceptual and Cognitive Differences Between Expert, Intermediate, and Novice Snooker Players. *Applied Cognitive Psychology, 8*, 185-211.
- Anderson, J. R. (1987). Skill Acquisition: Compilation of Weak-Method Problem Solutions. *Psychological Review, 94*(2), 192-210.
- Anderson, J. R., Corbett, A. T., & Reiser, B. J. (1987). *Essential Lisp*. Reading, MA: Addison-Wesley.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive Tutors: Lessons Learned. *The Journal of the Learning Sciences, 4*(2), 167-207.
- Atkinson, R. K., Derry, S. J. (2000). Computer-Based Examples Designed to Encourage Optimal Example Processing: A Study Examining the Impact of Sequentially Presented, Subgoal-Oriented Worked Examples. In B. Fishman & S. O'Conner-Divelbiss (Eds.), *Fourth International Conference of the Learning Sciences* (p. 132-133). Mahwah, NJ: Erlbaum.
- Atkinson, R., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from Examples: Instructional Principles from the Worked Examples Research. *Review of Educational Research, 70*(2), 181-214.

- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417-423.
- Baddeley, A. (2003). Working Memory: Looking Back and Looking Forward. *Nature Reviews*, 4, 829-839.
- Beichner, R. J. (2006). Instructional technology research and development in a US physics education group. *European Journal of Engineering Education*, 31(4), 383-393.
- Catrambone, R. & Holyoak, K. J. (1989). Overcoming Contextual Limitations on Problem-Solving Transfer. *Journal of Experimental Psychology*, 15(6), 1147-1156.
- Catrambone, R. (1995). Aiding Subgoal Learning: Effects on Transfer. *Journal of Educational Psychology*, 87, 5-17.
- Catrambone, R. (1996). Generalizing Solution Procedures Learned from Examples. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(4), 1020-1031.
- Catrambone, R. (1998). The Subgoal Learning Model: Creating Better Examples So That Students Can Solve Novel Problems. *Journal of Experimental Psychology: General*, 127(4), 355-376.
- Chabay, R., & Sherwood, B. (2002). *Matter and Interactions: Modern Mechanics*.

- New York, NY: John Wiley & Sons, Inc.
- Chase, W. G. & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), *Visual Information Processing* (pp. 215-281). New York, NY: Academic Press.
- Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R.J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1.) Hillsdale, NJ: Erlbaum. Collins, A., & Stevens.
- Chi, M. T. H., Bassock, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-Explanations: How Students Study and Use Examples in Learning to Solve Problems. *Cognitive Science*, 13, 145-182.
- Chi, M. T. H. & VanLehn, K. A. (1991). The Content of Physics Self-Explanations. *The Journal of the Learning Sciences*, 1(1), 69-105.
- Chi, M. T. H., de Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting Self-Explanations Improves Understanding. *Cognitive Science*, 18, 439-477.
- Clement, J. (1982). Students' preconceptions in introductory physics. *American Journal of Physics*, 50(1), 66-71.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, XX(1), 37-46.
- Conati, C. & VanLehn, K. (2000). Toward Computer-Based Support of Meta-Cognitive

- Skills: A Computations Framework to Coach Self-Explanation. *International Journal of Artificial Intelligence in Education*, 11, 389-415.
- Cooper, G. A. & Sweller, J. (1987). Effects of Schema Acquisition and Rule Automation on Mathematical Problem-Solving Transfer. *Journal of Educational Psychology*, 79(4), 347-362.
- de Groot, A. D. (1966). Perception and Memory Versus Thought: Some Old Ideas and Recent Findings. In B. Kleinmuntz (Ed.), *Problem Solving Research Method and Theory* (pp. 19-50). New York, NY: John Wiley & Sons, Inc.
- diSessa, A. A. (1993). Toward an Epistemology of Physics. *Cognition and Instruction*, 10(2&3), 105-225.
- Dufrene, R. J., Gerace, W. J., Hardiman, P. T., & Mestre, J. P. (1992). Constraining Novices to Perform Expertlike Problem Analysis: Effects on Schema Acquisition. *The Journal of the Learning Sciences*, 2(3), 307-331.
- Elyon, B. S. & Reif, F. (1984) Effects of Knowledge Organization on Task Performance. *Cognition and Instruction*, 1(1), 5-44.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: The MIT Press.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: The MIT Press.

- Etkina, E., Van Heuvelen, A., White-Brahmia, S., Brookes, D. T., Gentile, M., Murthy, S., Rosengrant, D., & Warren, A. (2006). Scientific Abilities and their Assessment. *Phys. Rev. ST Phys. Educ. Res.* 2, 020103.
- Ferguson-Hessler, M. G. M. & de Jong, T. (1990). Studying Physics Texts: Differences in Study Processes between Good and Poor Performers. *Cognition and Instruction*, 7(1), 41-54.
- Geisler, C. (2003). *Analyzing Streams of Language: Twelve Steps to the Systematic Coding of Text, Talk, and Other Verbal Data*. New York, NY: Pearson/Longman.
- Gerjets, P., Scheiter, K., & Catrambone, R. (2004). Designing Instructional Examples to Reduce Intrinsic Cognitive Load: Molar versus Modular Presentation of Solution Procedures. *Instructional Science*, 32, 33-58.
- Gertner, A. S. & VanLehn, K. (2000). ANDES: A Coached Problem Solving Environment for Physics. In Gauthier, G., Frasson, C., & VanLehn, K. (Eds.), *Proceedings of the Fifth International Conference, Intelligent Tutoring Systems* (p. 133-142). London, UK: Springer.
- Große, C. S. & Renkl, A. (2007). Finding and fixing errors in worked examples: Can this foster learning outcomes? *Learning and Instruction*, 17, 612-634.
- Halliday, D., & Resnick, R. (1981). *Fundamentals of physics*. New York: Wiley.
- Halloun, I. A., Hestenes, D. (1985). Common sense concepts about motion. *American*

- Journal of Physics*, 53(11), 1056-1065.
- Hammer, D. (1994). Epistemological Beliefs in Introductory Physics. *Cognition and Instruction*, 12(2), 151-183.
- Harper, K., Foster, T., & Maloney, D. (2007). Research-Based Alternatives to Problem Solving Workshop. *American Association of Physics Teachers Summer Meeting*. Greensboro, NC.
- Hart, S. G. & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In P. A. Hancock & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139-183). Amsterdam: Elsevier Science Publishers B.V.
- Heller, P. & Hollabaugh, M. (1992). Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups. *American Journal of Physics*, 60(7), 637-644.
- Kim, E. & Pak, S. (2002). Students do not overcome conceptual difficulties after solving 1000 traditional problems. *American Journal of Physics*, 70(7), 759-765.
- Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and Novice Performance in Solving Problems. *Science*, 208(4450), 1335-1342.
- LeFevre, J. & Dixon, P. (1986). Do Written Instructions Need Examples? *Cognition and Instruction*, 3(1), 1-30.

- Markman, E. M. (1981). Comprehension Monitoring. In W. P. Dickson (Ed.), *Children's Oral Communication Skills* (pp. 61-84). New York, NY: Academic Press.
- Miller, G. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. *The Psychological Review*, 63, 81-97.
- Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing Cognitive Load by Mixing Auditory and Visual Presentation Modes. *Journal of Educational Psychology*, 87(2), 319-334.
- Palincsar, A. S. & Brown, A. L. (1984). Reciprocal Teaching of Comprehension-Fostering and Comprehension-Monitoring Activities. *Cognition and Instruction*, 1(2), 117-175.
- Penney, C. (1989). Modality effects and the structure of short-term verbal memory. *Memory and Cognition*, 17(4), 398-422.
- Peters, P. C. (1982). Even Honors Students have Conceptual Difficulties with Physics. *American Journal of Physics*, 50(6), 501-508.
- Pirolli, P. (1991). Effects of Examples and Their Explanations in a Lesson on Recursion: A Production System Analysis. *Cognition and Instruction*, 8(3), 207-259.
- Pirolli, P. & Recker, M. (1994). Learning Strategies and Transfer in the Domain of Programming. *Cognition and Instruction*, 12(3), 235-275.

- Reif, F. & Larkin, J. H. (1991). Cognition in Scientific and Everyday Domain: Comparison and Learning Implications. *Journal of Research in Science Teaching*, 28(9), 733-760.
- Reif, F. & Scott, L. A. (1999). Teaching Scientific Thinking Skills: Students and Computers Coaching Each Other. *American Journal of Physics*, 67(9), 819-831.
- Reiner, M., Slotta, J. D., Chi, M. T. H., & Resnick, L. B. (2000). Naive Physics Reasoning: A Commitment to Substance-Based Conceptions. *Cognition and Instruction*, 18(1), 1-34.
- Renkl, A. (1997). Learning from Worked-Out Examples: A Study on Individual Differences. *Cognitive Science*, 21(1), 1-29.
- Renkl, A., Stark, R., Gruber, H., & Mandl, H. (1998). Learning from Worked-Out Examples: The Effects of Example Variability and Elicited Self-Explanations. *Contemporary Educational Psychology*, 23, 90-108.
- Renkl, A. & Atkinson, R. K. (2002) Learning From Examples: Fostering Self-Explanations in Computer-Based Learning Environments. *Interactive Learning Environments*, 10(2), 105-119.
- Renkl, A., Atkinson, R. K., Maier, U. H., & Staley, R. (2002). From Example Study to Problem Solving: Smooth Transitions Help Learning. *The Journal of Experimental Education*, 70(4), 293-315.

- Scherer, D., Dubois, P., and Sherwood, B. (2000). VPython: 3D interactive scientific graphics for students. *Computing in Science and Engineering*, 2(5), 56-62.
- Simon, H. A. & Gobet, F. (2000). Expertise Effects in Memory Recall: Comment on Vicente and Wang (1998). *Psychological Review*, 107(3), 593-600.
- Stark, R. (1999). *Lernen mit Lösungsbeispielen. Einfluß unvollständiger Lösungsbeispielen auf Beispielelaboration, Lernerfolg und Motivation*. Göttingen, Germany: Hogrefe Verlag für Psychologie.
- Stark, R., Mandl, H., Gruber, H., & Renkl, A. (2002). Conditions and effects of example elaboration. *Learning and Instruction*, 12, 39-60.
- Sternberg, R. J. (2003). *Cognitive Psychology*. (3rd Ed.) Belmont, CA: Wadsworth/Thomson Learning, Inc.
- Sweller, J. & Cooper, G. A. (1985). The Use of Worked Examples as a Substitute for Problem Solving in Learning Algebra. *Cognition and Instruction*, 2(1), 59-89.
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), 251-296.
- Tarmizi, R. A. & Sweller, J. (1988). Guidance During Mathematical Problem Solving. *Journal of Educational Psychology*, 80(4), 424-436.
- Trafton, J. G. & Reiser, B. J. (1993). The Contribution of Studying Examples and

Solving Problems to Skill Acquisition. *Proceedings of the 15th Annual Conference of the Cognitive Science Society* (p.1017-1022). Erlbaum: Hillsdale, NJ.

Tuminaro, J. & Redish, E. F. (2007). Elements of a cognitive model of physics problem solving: Epistemic games. *Phys. Rev. ST Phys. Educ. Res.*, 3, 020101.

van Gog, T., Paas, F., & van Merriënboer, J. J. G. (2004). Process-Oriented Worked Examples: Improving Transfer Performance Through Enhanced Understanding. *Instructional Science*, 32, 83-98.

VanLehn, K., Jones, R. M., & Chi, M. T. H. (1992). A Model of the Self-Explanation Effect. *The Journal of the Learning Sciences*, 2(1), 1-59.

VanLehn, K. (1996). Cognitive Skill Acquisition. *Annu. Rev. Psychol.*, 47, 513-539.

van Merriënboer, J. J. G. & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review*, 17(2), 147-177.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.) Cambridge, MA: Harvard University Press.

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301), 236-244.

Ward, M. & Sweller, J. (1990). Structuring Effective Worked Examples. *Cognition and Instruction*, 7(1), 1-39.

Yerushalmi, E. & Polinger, C. (2006). Guiding Students to Learn from Mistakes. *Physics Education*, 41(6), 532-538.

Zhu, X. & Simon, H. A. (1987). Learning Mathematics From Examples and by Doing. *Cognition and Instruction*, 4(3), 137-166.

Zhu, X., Lee, Y., Simon, H. A., & Zhu, D. (1996). Cue recognition and cue elaboration in learning from examples. *Proc. Natl. Acad. Sci. USA*, 93, 1346-1351.

APPENDICES

APPENDIX A: PROBLEM AND SOLUTION SETS

The study used four problem and solution sets. All participants were given the problem and solution sets in the same order (as presented below). For each set, the problem statement and solutions were given to the participants at the same time. The problem statement and each solution were provided on separate pages (the empty box in the upper right corner of each solution was provided as a space for the participant to indicate their selection for the correct solution). The participants were given 15 minutes to complete the task for each problem and solution set.

A.1: “Close the Door” problem and solution set

The co-researcher, Jon Gaffney, posed this problem and solutions, stemming from a question in the fall 2004 third test for the first semester M&I introductory course.

Problem: A

A door stands open, and you want to shut it by throwing something at it. You could either throw a lump of clay or a rubber ball at the door, both of which have the same mass. You know that the rubber ball will bounce back, while the lump of clay will stick to the door. Which should you pick?

Figure A.1 – “Close the door” problem statement

Solution #1



When the rubber ball contacts the door, it will bounce back with almost the same speed as before the collision, while the clay will stick to the door. This is because the rubber ball's collision is nearly elastic.

$$K_{f,ball} + K_{f,door} = K_{i,ball} + K_{i,door}$$
$$\frac{1}{2}mv_{f,ball}^2 + K_{f,door} = \frac{1}{2}mv_{i,ball}^2$$

Since the ball leaves with just about as much speed as it came in with, the door gains very little kinetic energy. However, since the clay stops when it hits the door, all of that kinetic energy can be transferred to the door:

$$K_{f,clay+door} = K_{i,clay} + K_{i,door}$$
$$K_{f,clay+door} = \frac{1}{2}mv_{i,clay}^2$$

If both the ball and the clay hit the door with the same speed, the clay will transfer more kinetic energy to the door. Therefore the clay is the better choice.

Figure A.2 – “Close the door” given solution #1

Solution #2



If we pick the ball and door as the system, $\vec{\tau}_{net} = \langle 0, 0, 0 \rangle$ N·m. Also, since angular momentum is defined around a point, let's pick the hinge of the door to be that point.

For the ball:

$$\begin{aligned}\vec{L}_{i,trans,ball} + \vec{L}_{i,rot,door} &= \vec{L}_{f,trans,ball} + \vec{L}_{f,rot,door} \\ \Delta\vec{L}_{rot,door} &= \vec{L}_{f,trans,ball} - \vec{L}_{i,trans,ball}\end{aligned}$$

And $\vec{L}_{trans} = \vec{r} \times \vec{p}$. Assume the ball initially travels along the +x-axis and collides at a 90-degree angle with a door that is initially at rest but free to rotate about the y-axis. In this case, the ball's initial angular momentum is in the +y direction. After the collision, the ball is traveling in the opposite direction so its final angular momentum is in the -y direction.

$$\begin{aligned}\Delta\vec{L}_{rot,door} &= \langle 0, rp_{i,ball}, 0 \rangle - \langle 0, -rp_{f,ball}, 0 \rangle \\ |\Delta\vec{L}_{rot,door}| &= |r(p_{i,ball} + p_{f,ball})|\end{aligned}$$

And, because the final speed of the rubber ball is nearly as much as its initial speed,

$$|\Delta\vec{L}_{rot,door}| \approx |2rmv_{i,ball}|$$

However, for the clay and the door as the system,

$$\begin{aligned}\vec{L}_{i,trans,clay} + \vec{L}_{i,rot,door} &= \vec{L}_{f,trans,clay+door} \\ \text{But, } m_{clay} &<< m_{door}\end{aligned}$$

$$\begin{aligned}\Delta\vec{L}_{rot,door} &\approx \vec{L}_{i,trans,clay} \\ |\Delta\vec{L}_{rot,door}| &\approx |rp_{i,clay}| \approx |rmv_{i,clay}|\end{aligned}$$

Since the mass of the clay is equal to the mass of the rubber ball, the door's angular momentum will change more if the ball hits it with some speed than if the clay hits it with that same speed. Therefore, the ball is the better choice.

Figure A.3 – “Close the door” given solution #2

Solution #3



Let us consider the momentum of the two objects.

Because $\vec{p} = m\vec{v}$, and the masses are equal, if the objects have the equal velocities then they will have equal momenta. Since both the clay and the rubber ball will provide the same amount of momentum to the door if they're thrown with the same initial speed, it doesn't matter which one you pick.

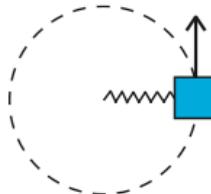
Figure A.4 – “Close the door” given solution #3

A.2: “UCM Spring” problem and solution set

This problem was largely inspired by a problem from the spring 2007 final exam for the first semester M&I introductory course.

Problem: B

A 0.5 kg block of wood is attached to one end of a spring with negligible mass. While performing an experiment far away from other objects, an astronaut holds the other end of the spring and swings the block in a circle with a constant speed. She finds that the block swings around with a period T of 1.5 seconds and that the length of the spring has stretched to a total length of 0.6 meters. The spring has a relaxed length of 0.25 m.



What is the stiffness of the spring?

Figure A.5 – “UCM Spring” problem statement

Solution #1



System: Block, spring
Surroundings: Nothing significant

Momentum Principle: $\vec{p}_f = \vec{p}_i + \vec{F}_{net}\Delta t$

For the radial component (the perpendicular component to the motion): $mv = (k_s s)T$

$$k_s = \frac{m\cancel{\omega}^{\omega r}}{sT} = \frac{m\cancel{\omega}^{\omega r} T}{sT} = \frac{2\pi mr}{sT^2}$$

$$k_s = \frac{2\pi (0.5 \text{ kg}) (0.6 \text{ m})}{(0.6 \text{ m} - 0.25 \text{ m}) (1.5 \text{ s})^2}$$

$$k_s = 2.39 \frac{\text{N}}{\text{m}}$$

Figure A.6 – “UCM Spring” given solution #1

Solution #2



System: Block

Surroundings: Spring

$$\text{Momentum Principle: } \frac{d\vec{p}}{dt} = \vec{F}_{net}$$

Since the block is moving in a circle at a constant speed, only the perpendicular component (to the motion) of $\frac{d\vec{p}}{dt}$ is non-zero. The textbook showed that this component is equal to $p\frac{v}{r}$ (recall the "kissing" circle).

$$\left| \frac{d\vec{p}}{dt} \right| = p \frac{v}{r} = p\omega = mv\omega = m(r\omega)\omega = mr\omega^2, \text{ where } \omega = \frac{2\pi}{T} = \frac{2\pi}{1.5 \text{ s}} = 4.19 \frac{\text{rad}}{\text{sec}}$$

$$\left| \frac{d\vec{p}}{dt} \right| = \left| \vec{F}_{net} \right| = \left| \vec{F}_{spring} \right| = k_s s$$

$$mr\omega^2 = k_s s$$

$$k_s = \frac{mr\omega^2}{s}$$

$$k_s = \frac{(0.5 \text{ kg})(0.6 \text{ m}) \left(4.19 \frac{\text{rad}}{\text{sec}} \right)^2}{0.6 \text{ m} - 0.25 \text{ m}}$$

$$k_s = 15.05 \frac{\text{N}}{\text{m}}$$

Figure A.7 – “UCM Spring” given solution #2

Solution #3



System: Block

Surroundings: Spring

$$\text{Momentum Principle: } \frac{d\vec{p}}{dt} = \vec{F}_{\text{net}}$$

$$\text{For the radial component (the perpendicular component to the motion): } \left| \frac{d\vec{p}}{dt} \right| = \left| \vec{F}_{\text{spring}} \right| = k_s s$$

The analytical solution of a spring mass system is: $A \cos(\omega t)$

Substitute the analytical solution into the Momentum Principle:

$$\left| \frac{dp}{dt} \right| \approx \left| \frac{d(mx)}{dt} \right| = \left| m \frac{d}{dt} \frac{ds}{dt} \right| = \left| m \frac{d}{dt} \frac{d[A \cos(\omega t)]}{dt} \right| = \left| -\omega m A \frac{d}{dt} (\sin(\omega t)) \right| = \left| -\omega^2 m A \cos(\omega t) \right|$$

$$\left| \frac{dp}{dt} \right| \approx \left| -\omega^2 m A \cos(\omega t) \right| = k_s s, \text{ or } k_s = \omega^2 m$$

$$\omega = \sqrt{\frac{k_s}{m}}, \text{ where } \omega = \frac{2\pi}{T} = \frac{2\pi}{1.5 \text{ s}} = 4.19 \frac{\text{rad}}{\text{sec}}$$

$$k_s = \omega^2 m = \left(4.19 \frac{\text{rad}}{\text{sec}} \right)^2 (0.5 \text{ kg})$$

$$k_s = 8.78 \frac{\text{N}}{\text{m}}$$

Figure A.8 – “UCM Spring” given solution #3

A.3: “Fission” problem and solution set

This problem directly stems from Problem 4.15 in the introductory course’s textbook (Chabay and Sherwood, 2002, p.161).

Problem: C

A uranium nucleus (U-236) may undergo a fission process resulting in two palladium nuclei (Pd-118). Each Pd-118 nucleus has a charge of $46e$ and a rest mass of 117.894 u , where $e = 1.6 \times 10^{-19} \text{ C}$ and $\text{u} = 1.6603 \times 10^{-27} \text{ kg}$. U-236 has a charge of $92e$ and a rest mass of 235.996 u . (There may also be some free neutrons after the fission process, but neglect those for this problem).

What is the distance between the Pd-118 nuclei centers just after the fission process (when they are initially at rest)?

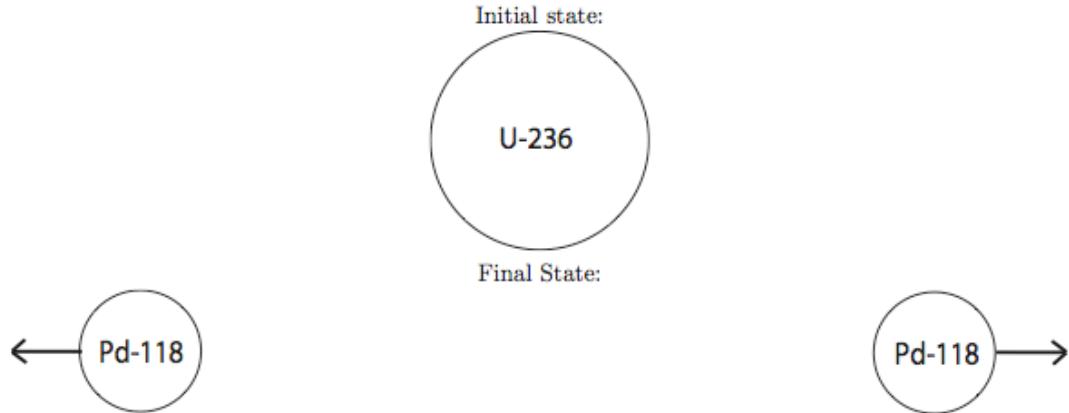
No figures in solutions drawn to scale

Figure A.9 – “Fission” problem statement

Solution #1



System: All particles
Surroundings: Nothing significant



Energy Principle: $E_f = E_i + W$

$$E_{rest,f} + U_f = E_{rest,i} + \mathcal{W}$$

$$\frac{1}{4\pi\epsilon_0} \frac{(Q_{Pd-118})(Q_{U-236})}{r} = E_{rest,i} - E_{rest,f}$$

$$\left(9 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2}\right) \frac{(46 \times 1.6 \times 10^{-19} \text{ C})(92 \times 1.6 \times 10^{-19} \text{ C})}{r} = m_{U-236}c^2 - 2m_{Pd-118}c^2$$

$$r = \frac{9.8 \times 10^{-25} \text{ N} \cdot \text{m}^2}{m_{U-236}c^2 - 2m_{Pd-118}c^2}$$

$$r = \frac{9.8 \times 10^{-25} \text{ N} \cdot \text{m}^2}{(235.996 \times 1.6603 \times 10^{-27} \text{ kg}) \left(3 \times 10^8 \frac{\text{m}}{\text{s}}\right)^2 - 2(117.894 \times 1.6603 \times 10^{-27} \text{ kg}) \left(3 \times 10^8 \frac{\text{m}}{\text{s}}\right)^2}$$

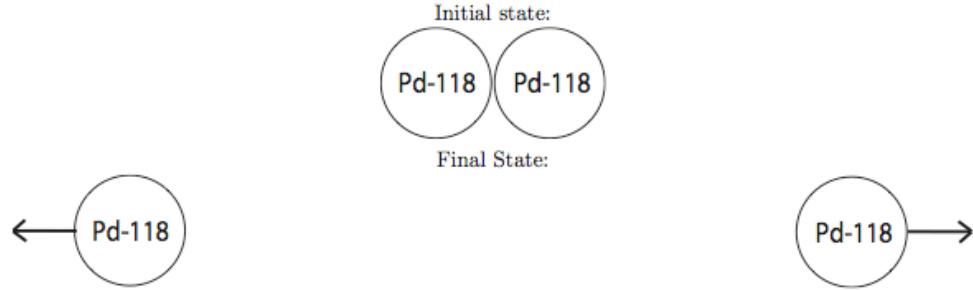
$$r = 5.553 \times 10^{-17} \text{ m}$$

Figure A.10 – “Fission” given solution #1

Solution #2



System: All particles
 Surroundings: Nothing significant



Energy Principle: $\Delta E = W$

$$E_f - E_i = \mathcal{W}$$

$$E_f = E_i$$

$$E_{rest,f} + \cancel{U_f}^{\approx 0} = E_{rest,i} + U_i$$

$$\frac{1}{4\pi\epsilon_0} \frac{(Q_{Pd-118})^2}{r} = m_{U-236}c^2 - 2m_{Pd-118}c^2$$

$$r = \left(9 \times 10^9 \frac{N \cdot m^2}{C^2} \right) \frac{(46 \times 1.6 \times 10^{-19} C)^2}{(235.996 \times 1.6603 \times 10^{-27} kg) \left(3 \times 10^8 \frac{m}{s} \right)^2 - 2(117.894 \times 1.6603 \times 10^{-27} kg) \left(3 \times 10^8 \frac{m}{s} \right)^2}$$

$$r = 1.569 \times 10^{-14} m$$

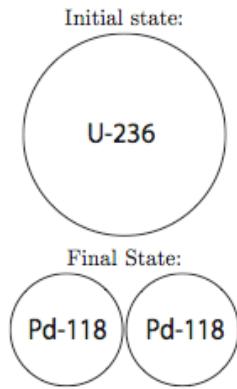
Figure A.11 – “Fission” given solution #2

Solution #3



System: All particles

Surroundings: Nothing significant



Energy Principle: $\Delta E = W$

$$\Delta K + \Delta U + \Delta E_{rest} = W$$

$$U_f - U_i + E_{rest,f} - E_{rest,i} = 0$$

$$\frac{1}{4\pi\epsilon_0} \frac{(Q_{Pd-118})^2}{r} + 2m_{Pd-118}c^2 - m_{U-236}c^2 = 0$$

$$r = \frac{1}{4\pi\epsilon_0} \frac{(Q_{Pd-118})^2}{m_{U-236}c^2 - 2m_{Pd-118}c^2}$$

$$r = \left(9 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2} \right) \frac{(46 \times 1.6 \times 10^{-19} \text{ C})^2}{(235.996 \times 1.6603 \times 10^{-27} \text{ kg}) \left(3 \times 10^8 \frac{\text{m}}{\text{s}} \right)^2 - 2(117.894 \times 1.6603 \times 10^{-27} \text{ kg}) \left(3 \times 10^8 \frac{\text{m}}{\text{s}} \right)^2}$$

$$r = 1.569 \times 10^{-14} \text{ m}$$

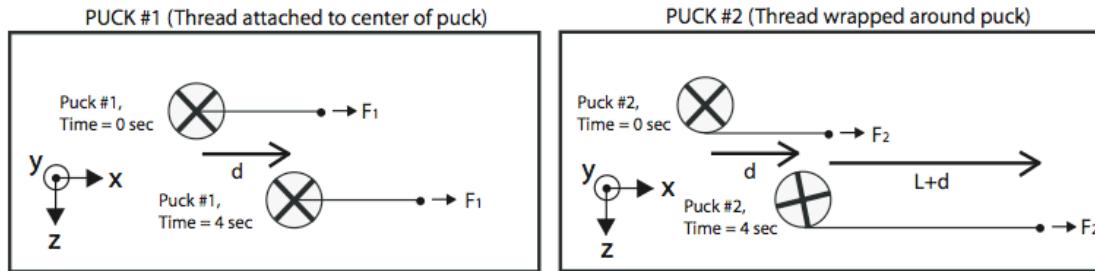
Figure A.12 – “Fission” given solution #3

A.4: “Two Pucks” problem and solution set

This problem was inspired by a demonstration of Bruce Sherwood, where two disks (mounted on low friction carts) were pulled by the same tensional force. One disk had the string attached at the center and the other disk had the string wound around the edge. When pulled by the common tensional force, both disks moved the same distance in the same time interval.

Problem: D

Two identical pucks (mass $m = 0.15 \text{ kg}$, radius $R = 0.05 \text{ m}$, and moment of inertia $I = 1.875 \times 10^{-4} \text{ kg} \cdot \text{m}^2$) are sitting at rest on a sheet of ice (neglect friction due to the interactions between the puck and ice). One puck has a thread (of negligible mass) attached to its center, while the other puck has a thread (also of negligible mass) wrapped around it. The threads are pulled by a constant force (of magnitudes F_1 and F_2 respectively) for 4 seconds at the free end of the thread (shown by a small circle below), at which time both pucks have a speed of 1.6 m/s . Over the 4 seconds, a length, L , of thread unwinds from the second disk.



Top view of pucks sliding on ice (looking down on the ice from a location on the +y axis). *Not drawn to scale*

You find that the distance, d , is 3.2 m and the length, L , is 6.4 m . In order for the pucks to be traveling at the same speed after the 4 seconds, will the second puck need to be pulled with a stronger force? That is, will F_2 need to be larger than F_1 ?

Figure A.13 – “Two Pucks” problem statement

Solution #1



For the first puck (thread attached at the center of the puck):

System: Puck, Earth

Surroundings: Thread

Energy Principle: $E_f = E_i + W$

$$K_f + U_f = K_i + U_i + W$$

$$\frac{1}{2}mv_f^2 + mgh_f = \frac{1}{2}mv_i^2 + mgh_i + W$$

$$\frac{1}{2}mv^2 + mgh_f - mgh_i = F_1d$$

$$\frac{1}{2}mv^2 + mg(h_f - h_i) = F_1d$$

$$F_1 = \frac{\frac{1}{2}mv_f^2}{d} = \frac{\frac{1}{2}(0.15 \text{ kg})(1.6 \frac{\text{m}}{\text{s}})^2}{3.2 \text{ m}} = 0.06 \text{ N}$$

For the second puck (thread wrapped around the puck):

System: Puck

Surroundings: Thread

Energy Principle: $E_f = E_i + W$

$$K_{trans,f} + K_{rot,f} - K_{trans,i} - K_{rot,i} = W$$

$$\frac{1}{2}mv_f^2 + \frac{1}{2}I\omega_f^2 = \frac{1}{2}mv_i^2 + \frac{1}{2}I\omega_i^2 + F_2(L + d)$$

$$F_2 = \frac{\frac{1}{2}mv_f^2 + \frac{1}{2}I\omega_f^2}{L + d} = \frac{\frac{1}{2}mv_f^2 + \frac{1}{2}I\left(\frac{v_f}{R}\right)^2}{L + d}$$

$$F_2 = \frac{\frac{1}{2}(0.15 \text{ kg})\left(1.6 \frac{\text{m}}{\text{s}}\right)^2 + \frac{1}{2}(1.875 \times 10^{-4} \text{ kg} \cdot \text{m}^2)\left(\frac{1.6 \text{ m}}{0.05 \text{ m}}\right)^2}{(6.4 + 3.2) \text{ m}} = 0.03 \text{ N}$$

The first puck requires a larger force (0.06 N as opposed to only 0.03 N) to reach the same speed as the second puck.

Figure A.14 – “Two Pucks” given solution #1

Solution #2



For the second puck (thread wrapped around the puck):

System: Puck

Surroundings: Thread

Angular Momentum Principle: $\vec{L}_f - \vec{L}_i = \vec{\tau}_{\text{net}}\Delta t$

$$\vec{L}_f - \vec{L}'_i = \vec{\tau}_{\text{net}}\Delta t$$

$$\vec{L}_{\text{trans},f} + \vec{L}_{\text{rot},f} = \vec{\tau}_{\text{net}}\Delta t$$

$$\vec{r} \times \vec{p}_f + I\vec{\omega}_f = \vec{r} \times \vec{F}_2\Delta t$$

$$\langle 0, p_f^R p_{fx}^R, 0 \rangle + I \langle 0, 0, \omega_f \rangle = \langle 0, p_f^R F_2^R, 0 \rangle \Delta t$$

(The puck will spin counter-clockwise as the thread unwinds, so by the Right Hand Rule $\vec{\omega}_f$ is in the y direction)

For the y-components: $Rp_{fx} + I\omega_f = RF_2\Delta t$

$$Rm v_f \cancel{+} I \left(\frac{v_f}{R} \right) = RF_2\Delta t, v_{fx} = v_f \text{ since the pucks travel in the x direction.}$$

$$F_2 = \frac{Rmv_f + I \left(\frac{v_f}{R} \right)}{R\Delta t}$$

$$F_2 = \frac{(0.05 \text{ m})(0.15 \text{ kg}) \left(1.6 \frac{\text{m}}{\text{s}} \right) + (1.875 \times 10^{-4} \text{ kg} \cdot \text{m}^2) \left(\frac{1.6 \frac{\text{m}}{\text{s}}}{0.05 \text{ m}} \right)}{(0.05 \text{ m})(4 \text{ sec})}$$

$$F_2 = 0.09 \text{ N}$$

For the first puck (thread attached at the center of the puck):

System: Puck

Surroundings: Thread

Angular Momentum Principle: $\vec{L}_f - \vec{L}_i = \vec{\tau}_{\text{net}}\Delta t$

$$\vec{L}_{\text{trans},f} - \cancel{\vec{L}_{\text{trans},i}} = \vec{\tau}_{\text{net}}\Delta t$$

$$\vec{r} \times \vec{p}_f = \vec{\tau}_{\text{net}}\Delta t$$

$$\langle 0, p_f^R p_{fx}^R, 0 \rangle = \langle 0, p_f^R F_1^R, 0 \rangle \Delta t$$

For the y-components: $\cancel{R}p_{fx} = F_1\Delta t$

$$mv_f \cancel{+} F_1\Delta t$$

$$(0.15 \text{ kg}) \left(1.6 \frac{\text{m}}{\text{s}} \right) = 0.24 \text{ kg} \cdot \frac{\text{m}}{\text{s}} = F_1\Delta t$$

$$F_1 = \frac{0.24 \text{ kg} \cdot \frac{\text{m}}{\text{s}}}{\Delta t} = \frac{0.24 \text{ kg} \cdot \frac{\text{m}}{\text{s}}}{4 \text{ sec}} = 0.06 \text{ N}$$

The second puck requires a larger force (0.09 N as opposed to only 0.06 N) to reach the same speed as the first puck.

Figure A.15 – “Two Pucks” given solution #2

Solution #3



For the first puck (thread attached at the center of the puck):

System: Puck

Surroundings: Thread, Earth, ice sheet

Momentum Principle: $\vec{p}_f = \vec{p}_i + \vec{F}_{net}\Delta t$

$$\vec{F}_{net} = \langle F_1, F_{ice} - F_{Earth}, 0 \rangle$$

$$\langle p_f, 0, 0 \rangle = \langle p_i, 0, 0 \rangle + \langle F_1, F_{ice} - F_{Earth}, 0 \rangle \Delta t$$

$$\langle mv_f, 0, 0 \rangle = \langle F_1, F_{ice} - F_{Earth}, 0 \rangle \Delta t$$

For the y-components: $0 = (F_{ice} - F_{Earth}) \Delta t$	For the x-components: $mv_f = F_1 \Delta t$
$F_{ice} - F_{Earth} = 0$	$F_1 = \frac{mv_f}{\Delta t}$
$F_{ice} = F_{Earth}$	$F_1 = \frac{(0.15 \text{ kg}) (1.6 \frac{\text{m}}{\text{s}})}{4 \text{ sec}} = 0.06 \text{ N}$

For the second puck (thread wrapped around the puck):

System: Puck

Surroundings: Thread, Earth, ice sheet

Momentum Principle: $\vec{p}_f = \vec{p}_i + \vec{F}_{net}\Delta t$

$$\langle p_f, 0, 0 \rangle = \langle p_i, 0, 0 \rangle + \langle F_2, F_{ice} - F_{Earth}, 0 \rangle \Delta t$$

$$\langle mv_f, 0, 0 \rangle = \langle F_2, F_{ice} - F_{Earth}, 0 \rangle \Delta t$$

For the x-components: $mv_f = F_2 \Delta t$	For the y-components: $0 = (F_{ice} - F_{Earth}) \Delta t$
$(0.15 \text{ kg}) (1.6 \frac{\text{m}}{\text{s}}) = F_2 \Delta t$	$F_{ice} - F_{Earth} = 0$
$F_2 \Delta t = 0.24 \text{ kg} \cdot \frac{\text{m}}{\text{s}}$	$F_{ice} = F_{Earth}$
$F_2 = \frac{0.24 \text{ kg} \cdot \frac{\text{m}}{\text{s}}}{\Delta t}$	
$F_2 = \frac{0.24 \text{ kg} \cdot \frac{\text{m}}{\text{s}}}{4 \text{ sec}} = 0.06 \text{ N}$	

Both forces need to have equal magnitudes (0.06 N) in order for the pucks to have the same speed at 4 seconds.

Figure A.16 – “Two Pucks” given solution #3

APPENDIX B: RESEARCHER SCRIPTS

The researchers conducting the data collection sessions used the same scripts to introduce the participants to the task and the think aloud protocol. Think aloud protocol warm-up exercises were also included in the introduction. At the discretion of the researcher, the participant completed some or all of the warm-up exercises to become familiarized with the think aloud protocol.

B.1: Welcome and introduction script – lead-in to the protocol warm-up exercises

Welcome to my research study, and thank you for participating! I greatly appreciate your help. [*Ensure that the first part of the consent form is signed and dated*]

Please note the video camera [*point out video camera*] that will record the session. Unless you look up at it, your face will not be seen. This is the microphone; it is very sensitive, so please be careful not to brush against it. Only NC State researchers will see the videotape unless you give me permission to show the tape to others. Once the session is done, you can fill out the last part of the consent form specifying how the video can be used.

If you have a cell phone here, please turn it off.

I am now going to begin the video recording.

During this study, I will give you physics problems to look at. For each problem that I give you, I will also give you three written solutions. Only one of the solutions will be correct. The other solutions will have errors. Your task is to figure out which solution is correct and to find the errors in the incorrect solutions. I have intentionally picked these problems to be **difficult**. So don't worry if you struggle with this task in any way. Because I am interested in how you identify the correct solution and find the errors, I need to use challenging problems and solutions otherwise we wouldn't get much useful information from this study. I just ask that you try your best. If you are not certain which solution is correct, then indicate as well as you can which one might be the correct solution. If you are not sure that you have found all errors, just try to find as many as possible.

There are others also being interviewed right now. Once these interviews have finished, we will bring all of you together to work as a team to discuss your decisions. During that session, you will be asked to present your decision and reasoning behind it to the other members of your group for one of the problems you will see in this interview. You will have one opportunity to "pass" if you are assigned a problem you do not want to present, so you should bear that in mind if there is one problem in particular that you would rather not present. I want to emphasize again that these problems *are* challenging, so it is OK if you simply need to take a guess and present that as justification to the group. All we ask is that you try your best.

At the end of the group session today, a researcher will be happy to explain the answers of all of the problems in this interview as well as the problems in the group interview. I can't explain the answers in between the problems or at the end of this interview.

During this study, I am interested in what you are thinking about. So, I will ask you to think aloud as you work through the problems and go over the solutions.

- By this I mean that I would like you to say everything that you are thinking, talking constantly, from the time you first see the problem and solutions until you are finished with the problem and solutions.
- You should not plan ahead what you are going to say or explain to me what you are saying.
- Act as if you are alone in this room talking through the problem to yourself.
- It is very important that you keep talking, so if you are silent for a period of time, then I will remind you to keep talking.

We will go through some warm-up exercises to get you used to the process, and I will give you suggestions to help you improve. Then I will describe your task for this session in more detail before we begin. Any questions?

B.2: Think aloud protocol warm-up exercises

The warm-up exercises were presented to the participants one at a time.

Let's do some warm-up exercises to practice thinking aloud.

- Without writing anything down, what is the product of 111 and 12 (remember to keep talking!):

$$\begin{array}{r} 111 \\ \times 12 \\ \hline \end{array}$$

- Below is the name of an animal that is only partially spelled out. Without writing anything down, what is the name of that animal? (remember to keep talking!)

- H - R -

- Think about the path that you would take from D.H. Hill library's exit to your favorite place to eat on Hillsborough Street. How many times would you turn to get there? (remember to keep talking!)

- How many doors (including interior doors such as a closet door) are in your parent's house? (remember to keep talking!)

- Think about all of the classes you are now taking. How many homework problems have you been assigned over the past two weeks for all of your classes combined? (remember to keep talking!)

6. How can you make the following equation true by drawing only one straight line? (remember to keep talking!)

$$5 + 5 + 5 = 550$$

B.3: Follow-up script to the warm-up exercises – the final instructions to the participants

Very good job on the think-aloud training! Remember that the most important thing for you to do is to tell us everything you're thinking. You can't tell us too much!

Here is a red pen that you can use to mark-up the solutions [*gesture to red pen*] and to mark the corner of the correct solution [*gesture to the square on one of the solutions as an example – KEEP MOST OF IT COVERED*]. Here is a reference sheet that you can use at any time [*gesture to the reference sheet*]. You can also use this calculator at any time [*gesture to the calculator*]. If you want to write anything, please use the whiteboard [*gesture to the whiteboard and marker*]. Please don't erase anything; just cross out your work. We have extra whiteboards [*gesture to whiteboards*]; we will give you a clean whiteboard whenever you want one.

You will have up to 15 minutes for each problem to determine which solution is correct and the errors in the other solutions. This timer will show you how much time is left [*motion to timer*]. We can move onto the next problem and solutions set at any time before that 15 minutes is up; just let me know when you're done working with each problem and we can go on.

As a reminder, we are asking you to determine which solution is correct and what is wrong with the other solutions. Do you have any questions about your task during this session?

Ok. Here's your first problem [*present the participant the first problem, with the three possible solutions*].

B.4: Reference Sheet

As mentioned in the follow-up script (see section B.3), the participants were provided with a reference sheet (largely generated from the same reference sheet that the first semester course of the Matter and Interactions introductory sequence at North Carolina State University provides for exams).

Reference Sheet

Fundamental principles:

$$\Delta \vec{p} = \vec{F}_{net} \Delta t \quad \text{or} \quad \frac{d\vec{p}}{dt} = \vec{F}_{net}$$

$$\Delta E = W_{surr} + Q$$

$$\Delta \vec{L}_A = \vec{\tau}_{net,A} \Delta t \quad \text{or} \quad \frac{d\vec{L}_A}{dt} = \vec{\tau}_{net,A}$$

Cross Product: $\vec{A} \times \vec{B} = ((A_y B_z - A_z B_y), (A_z B_x - A_x B_z), (A_x B_y - A_y B_x))$

$$\vec{L}_{trans,A} = \vec{r}_A \times \vec{p} \text{ (point particle)} \quad \vec{\tau}_A = \vec{r}_A \times \vec{F}$$

Multiparticle systems:

$$\vec{r}_{cm} = \frac{m_1 \vec{r}_1 + m_2 \vec{r}_2 + \dots}{m_1 + m_2 + \dots}$$

$$\vec{P}_{tot} \approx M_{tot} \vec{v}_{cm} \quad (v \ll c)$$

$$K_{tot} = K_{trans} + K_{rel}$$

$$K_{rel} = K_{rot} + K_{vib}$$

$$K_{trans} \approx \frac{1}{2} M_{tot} v_{cm}^2 \quad (v \ll c)$$

$$I = m_1 r_{1\perp}^2 + m_2 r_{2\perp}^2 + \dots$$

$$K_{rot} = \frac{L_{rot}^2}{2I} = \frac{1}{2} I \omega^2$$

$$\vec{L}_{rot} = I \vec{\omega}$$

$$\vec{L}_A = \vec{L}_{trans,A} + \vec{L}_{rot}$$

Other physical quantities:

$$\gamma \equiv \frac{1}{\sqrt{1 - \left(\frac{|\vec{v}|}{c}\right)^2}}$$

$$E^2 - (pc)^2 = (mc^2)^2$$

$$\vec{F}_{grav} = -G \frac{m_1 m_2}{|\vec{r}|^2} \hat{r}$$

$$U_{grav} = -G \frac{m_1 m_2}{|\vec{r}|}$$

$|\vec{F}_{grav}| \approx mg$ near Earth's surface

$\Delta U_{grav} \approx mg \Delta y$ near Earth's surface

$$\vec{F}_{elec} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{|\vec{r}|^2} \hat{r}$$

$$U_{elec} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{|\vec{r}|}$$

$|\vec{F}_{spring}| = k_s s$ opposite to the stretch

$U_{spring} = \frac{1}{2} k_s s^2$ for ideal spring

$$U_i \approx \frac{1}{2} k_{si} s^2 - E_M \text{ approx. interatomic pot. energy}$$

$$\Delta E_{thermal} = mC\Delta T$$

$$E_N = -\frac{13.6\text{eV}}{N^2} \text{ where } N = 1, 2, 3 \dots \text{ (Hydrogen atom energy levels)}$$

$$E_N = N\hbar\omega_0 + E_0 \text{ where } N = 0, 1, 2 \dots \text{ and } \omega_0 = \sqrt{\frac{k_{si}}{m_a}} \text{ (Quantized oscillator energy levels)}$$

$$\left| \left(\frac{d\vec{p}}{dt} \right)_{\perp} \right| = \frac{pv}{R} \approx \frac{mv^2}{R} \text{ (if } v \ll c \text{) where } R = \text{radius of kissing circle}$$

Figure B.1 – First page of reference sheet

$$\omega = \frac{2\pi}{T}$$

$$Y = \frac{F/A}{\Delta L/L} \text{ (macro)}$$

$$x = A \cos \omega t$$

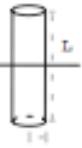
$$Y = \frac{k_{si}}{d} \text{ (micro)}$$

$$\omega = \sqrt{\frac{k_s}{m}}$$

$$\text{speed of sound } v = d \sqrt{\frac{k_{si}}{m_0}}$$

$\hat{f} = \langle \cos \theta_x, \cos \theta_y, \cos \theta_z \rangle$ unit vector from angles

Moment of inertia for rotation about indicated axis

			
$I = \frac{2}{5}MR^2$	$I = \frac{1}{2}MR^2$	$I = \frac{1}{12}ML^2$	$I = \frac{1}{12}ML^2 + \frac{1}{4}MR^2$

$$\Omega = \frac{(q+N-1)!}{q!(N-1)!}$$

$$\Delta S = \frac{Q}{T} \text{ (small } Q)$$

$$S \equiv k \ln \Omega$$

$$\text{prob}(E) \propto \Omega(E) e^{-\frac{E}{kT}}$$

$$\frac{1}{T} \equiv \frac{\partial S}{\partial E}$$

Constant	Symbol	Approximate Value
Speed of light	c	$3 \times 10^8 \text{ m/s}$
Gravitational constant	G	$6.7 \times 10^{-11} \text{ N} \cdot \text{m}^2/\text{kg}^2$
Approx. grav field near Earth's surface	g	9.8 N/kg
Electron mass	m_e	$9 \times 10^{-31} \text{ kg}$
Proton mass	m_p	$1.7 \times 10^{-27} \text{ kg}$
Neutron mass	m_n	$1.7 \times 10^{-27} \text{ kg}$
Electric constant	$\frac{1}{4\pi\epsilon_0}$	$9 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$
Proton charge	e	$1.6 \times 10^{-19} \text{ C}$
Electron volt	1 eV	$1.6 \times 10^{-19} \text{ J}$
Avogadro's number	N_A	$6.02 \times 10^{23} \text{ atoms/mol}$
Plank's constant	h	$6.6 \times 10^{-34} \text{ joule} \cdot \text{second}$
$\hbar = \frac{h}{2\pi}$	\hbar	$1.05 \times 10^{-34} \text{ joule} \cdot \text{second}$
specific heat capacity of water	C	4.2 J/kg
Boltzmann constant	k	$1.38 \times 10^{-23} \text{ J/K}$

milli	m	1×10^{-3}
micro	μ	1×10^{-6}
nano	n	1×10^{-9}
pico	p	1×10^{-12}

kilo	K	1×10^3
mega	M	1×10^6
giga	G	1×10^9
tera	T	1×10^{12}

Figure B.2 – Second page of reference sheet

APPENDIX C: FOCUS CODING BREAKPOINTS

The focus-coding pass denoted whether the participant was focused upon the approach or answer of a solution when predicting, judging, comparing, or expressing uncertainty. For the purposes of this research, specific breakpoints were used to separate the step (or steps) that was considered to be the answer from the approach for each solution. The breakpoint in the below solutions is denoted by a dashed line.

C.1: “Close the Door” solutions

Solution #1



When the rubber ball contacts the door, it will bounce back with almost the same speed as before the collision, while the clay will stick to the door. This is because the rubber ball's collision is nearly elastic.

$$K_{f,ball} + K_{f,door} = K_{i,ball} + K_{i,door}$$
$$\frac{1}{2}mv_{f,ball}^2 + K_{f,door} = \frac{1}{2}mv_{i,ball}^2$$

Since the ball leaves with just about as much speed as it came in with, the door gains very little kinetic energy. However, since the clay stops when it hits the door, all of that kinetic energy can be transferred to the door:

$$K_{f,clay+door} = K_{i,clay} + K_{i,door}$$
$$K_{f,clay+door} = \frac{1}{2}mv_{i,clay}^2$$

If both the ball and the clay hit the door with the same speed, the clay will transfer more kinetic energy to the door. Therefore the clay is the better choice.

Figure C.1 – “Close the door” given solution #1 with breakpoint

Solution #2



If we pick the ball and door as the system, $\vec{\tau}_{net} = \langle 0, 0, 0 \rangle$ N·m. Also, since angular momentum is defined around a point, let's pick the hinge of the door to be that point.

For the ball:

$$\begin{aligned}\vec{L}_{i,trans,ball} + \vec{L}_{i,rot,door} &= \vec{L}_{f,trans,ball} + \vec{L}_{f,rot,door} \\ \Delta\vec{L}_{rot,door} &= \vec{L}_{f,trans,ball} - \vec{L}_{i,trans,ball}\end{aligned}$$

And $\vec{L}_{trans} = \vec{r} \times \vec{p}$. Assume the ball initially travels along the $+x$ -axis and collides at a 90-degree angle with a door that is initially at rest but free to rotate about the y -axis. In this case, the ball's initial angular momentum is in the $+y$ direction. After the collision, the ball is traveling in the opposite direction so its final angular momentum is in the $-y$ direction.

$$\begin{aligned}\Delta\vec{L}_{rot,door} &= \langle 0, rp_{i,ball}, 0 \rangle - \langle 0, -rp_{f,ball}, 0 \rangle \\ |\Delta\vec{L}_{rot,door}| &= |r(p_{i,ball} + p_{f,ball})|\end{aligned}$$

And, because the final speed of the rubber ball is nearly as much as its initial speed,

$$|\Delta\vec{L}_{rot,door}| \approx |2rmv_{i,ball}|$$

However, for the clay and the door as the system,

$$\begin{aligned}\vec{L}_{i,trans,clay} + \vec{L}_{i,rot,door} &= \vec{L}_{f,trans,clay+door} \\ \text{But, } m_{clay} &<< m_{door}\end{aligned}$$

$$\begin{aligned}\Delta\vec{L}_{rot,door} &\approx \vec{L}_{i,trans,clay} \\ |\Delta\vec{L}_{rot,door}| &\approx |rp_{i,clay}| \approx |rmv_{i,clay}|\end{aligned}$$

Since the mass of the clay is equal to the mass of the rubber ball, the door's angular momentum will change more if the ball hits it with some speed than if the clay hits it with that same speed. Therefore, the ball is the better choice.

Figure C.2 – “Close the door” given solution #2 with breakpoint

Solution #3



Let us consider the momentum of the two objects.

Because $\vec{p} = m\vec{v}$, and the masses are equal, if the objects have the equal velocities then they will have equal momenta. Since both the clay and the rubber ball will provide the same amount of momentum to the door if they're thrown with the same initial speed, it doesn't matter which one you pick.

Figure C.3 – “Close the door” given solution #3 with breakpoint

C.2: “UCM Spring” solutions

Solution #1



System: Block, spring

Surroundings: Nothing significant

Momentum Principle: $\vec{p}_f = \vec{p}_i + \vec{F}_{net}\Delta t$

For the radial component (the perpendicular component to the motion): $mv = (k_s s)T$

$$k_s = \frac{mv^r}{sT} = \frac{m\omega r}{sT} = \frac{\frac{2\pi}{T}mr}{sT^2}$$

$$k_s = \frac{2\pi(0.5 \text{ kg})(0.6 \text{ m})}{(0.6 \text{ m} - 0.25 \text{ m})(1.5 \text{ s})^2}$$

$$k_s = 2.39 \frac{\text{N}}{\text{m}}$$

Figure C.4 – “UCM Spring” given solution #1 with breakpoint

Solution #2



System: Block

Surroundings: Spring

Momentum Principle: $\frac{d\vec{p}}{dt} = \vec{F}_{net}$

Since the block is moving in a circle at a constant speed, only the perpendicular component (to the motion) of $\frac{d\vec{p}}{dt}$ is non-zero. The textbook showed that this component is equal to $p\frac{v}{r}$ (recall the "kissing" circle).

$$\left| \frac{d\vec{p}}{dt} \right| = p \frac{v}{r} = p\omega = mv\omega = m(r\omega)\omega = mr\omega^2, \text{ where } \omega = \frac{2\pi}{T} = \frac{2\pi}{1.5 \text{ s}} = 4.19 \frac{\text{rad}}{\text{sec}}$$

$$\left| \frac{d\vec{p}}{dt} \right| = \left| \vec{F}_{net} \right| = \left| \vec{F}_{spring} \right| = k_s s$$

$$mr\omega^2 = k_s s$$

$$k_s = \frac{mr\omega^2}{s}$$

$$k_s = \frac{(0.5 \text{ kg})(0.6 \text{ m}) \left(4.19 \frac{\text{rad}}{\text{sec}} \right)^2}{0.6 \text{ m} - 0.25 \text{ m}}$$

$$k_s = 15.05 \frac{\text{N}}{\text{m}}$$

Figure C.5 – “UCM Spring” given solution #2 with breakpoint

Solution #3



System: Block

Surroundings: Spring

$$\text{Momentum Principle: } \frac{d\vec{p}}{dt} = \vec{F}_{net}$$

$$\text{For the radial component (the perpendicular component to the motion): } \left| \frac{d\vec{p}}{dt} \right| = \left| \vec{F}_{spring} \right| = k_s s$$

The analytical solution of a spring mass system is: $A \cos(\omega t)$

Substitute the analytical solution into the Momentum Principle:

$$\left| \frac{dp}{dt} \right| \approx \left| \frac{d(mx)}{dt} \right| = \left| m \frac{d}{dt} \frac{ds}{dt} \right| = \left| m \frac{d}{dt} \frac{d[A \cos(\omega t)]}{dt} \right| = \left| -\omega m A \frac{d}{dt} (\sin(\omega t)) \right| = \left| -\omega^2 m A \cos(\omega t) \right|$$

$$\left| \frac{dp}{dt} \right| \approx \left| -\omega^2 m A \cos(\omega t) \right| = k_s s, \text{ or } k_s = \omega^2 m$$

$$\omega = \sqrt{\frac{k_s}{m}}, \text{ where } \omega = \frac{2\pi}{T} = \frac{2\pi}{1.5 \text{ s}} = 4.19 \frac{\text{rad}}{\text{sec}}$$

$$k_s = \omega^2 m = \left(4.19 \frac{\text{rad}}{\text{sec}} \right)^2 (0.5 \text{ kg})$$

$$k_s = 8.78 \frac{\text{N}}{\text{m}}$$

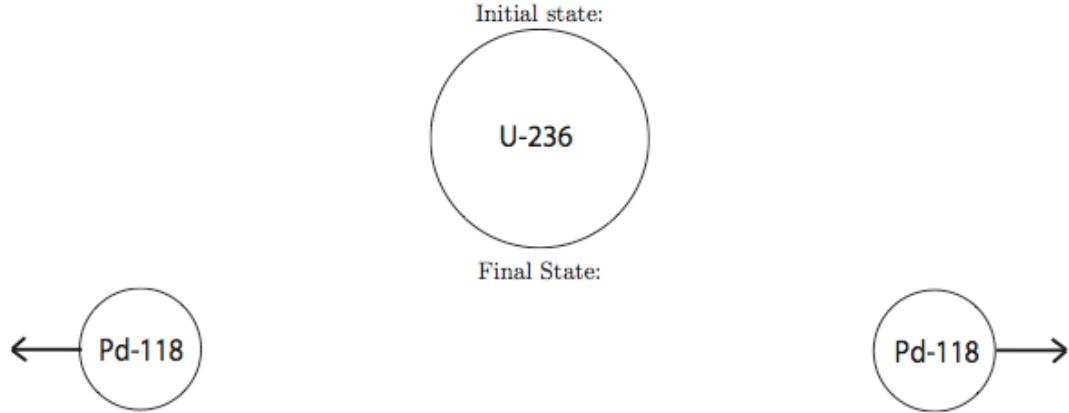
Figure C.6 – “UCM Spring” given solution #3 with breakpoint

C.3: “Fission” solutions

Solution #1



System: All particles
Surroundings: Nothing significant



Energy Principle: $E_f = E_i + W$

$$E_{rest,f} + U_f = E_{rest,i} + \mathcal{W}$$

$$\frac{1}{4\pi\epsilon_0} \frac{(Q_{Pd-118})(Q_{U-236})}{r} = E_{rest,i} - E_{rest,f}$$

$$\left(9 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2}\right) \frac{(46 \times 1.6 \times 10^{-19} \text{ C})(92 \times 1.6 \times 10^{-19} \text{ C})}{r} = m_{U-236}c^2 - 2m_{Pd-118}c^2$$

$$r = \frac{9.8 \times 10^{-25} \text{ N} \cdot \text{m}^2}{m_{U-236}c^2 - 2m_{Pd-118}c^2}$$

$$r = \frac{9.8 \times 10^{-25} \text{ N} \cdot \text{m}^2}{(235.996 \times 1.6603 \times 10^{-27} \text{ kg}) \left(3 \times 10^8 \frac{\text{m}}{\text{s}}\right)^2 - 2(117.894 \times 1.6603 \times 10^{-27} \text{ kg}) \left(3 \times 10^8 \frac{\text{m}}{\text{s}}\right)^2}$$

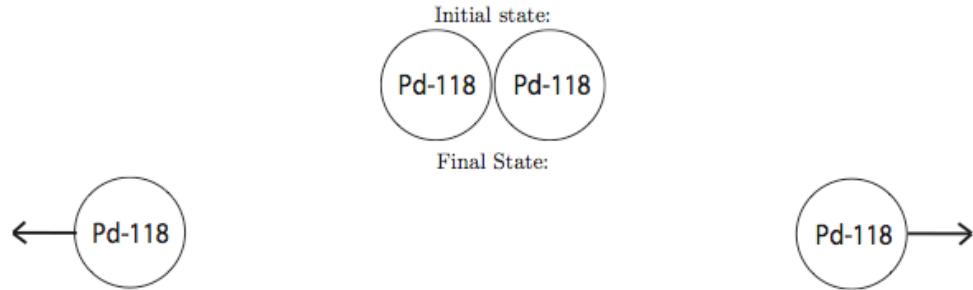
$$r = 5.553 \times 10^{-17} \text{ m}$$

Figure C.7 – “Fission” given solution #1 with breakpoint

Solution #2



System: All particles
 Surroundings: Nothing significant



Energy Principle: $\Delta E = W$

$$E_f - E_i = \mathcal{W}$$

$$E_f = E_i$$

$$E_{rest,f} + \cancel{U_f}^{\approx 0} = E_{rest,i} + U_i$$

$$\frac{1}{4\pi\epsilon_0} \frac{(Q_{Pd-118})^2}{r} = m_{U-236}c^2 - 2m_{Pd-118}c^2$$

$$r = \left(9 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2} \right) \frac{(46 \times 1.6 \times 10^{-19} \text{ C})^2}{(235.996 \times 1.6603 \times 10^{-27} \text{ kg}) \left(3 \times 10^8 \frac{\text{m}}{\text{s}} \right)^2 - 2(117.894 \times 1.6603 \times 10^{-27} \text{ kg}) \left(3 \times 10^8 \frac{\text{m}}{\text{s}} \right)^2}$$

$$r = 1.569 \times 10^{-14} \text{ m}$$

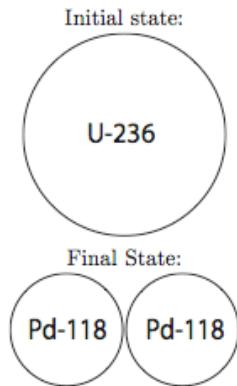
Figure C.8 – “Fission” given solution #2 with breakpoint

Solution #3



System: All particles

Surroundings: Nothing significant



Energy Principle: $\Delta E = W$

$$\Delta K + \Delta U + \Delta E_{rest} = W$$

$$U_f - U_i + E_{rest,f} - E_{rest,i} = 0$$

$$\frac{1}{4\pi\epsilon_0} \frac{(Q_{Pd-118})^2}{r} + 2m_{Pd-118}c^2 - m_{U-236}c^2 = 0$$

$$r = \frac{1}{4\pi\epsilon_0} \frac{(Q_{Pd-118})^2}{m_{U-236}c^2 - 2m_{Pd-118}c^2}$$

$$r = \left(9 \times 10^9 \frac{\text{N} \cdot \text{m}^2}{\text{C}^2} \right) \frac{(46 \times 1.6 \times 10^{-19} \text{ C})^2}{(235.996 \times 1.6603 \times 10^{-27} \text{ kg}) \left(3 \times 10^8 \frac{\text{m}}{\text{s}} \right)^2 - 2(117.894 \times 1.6603 \times 10^{-27} \text{ kg}) \left(3 \times 10^8 \frac{\text{m}}{\text{s}} \right)^2}$$

$$r = 1.569 \times 10^{-14} \text{ m}$$

Figure C.9 – “Fission” given solution #3 with breakpoint

C.4: “Two Pucks” solutions

Solution #1



For the first puck (thread attached at the center of the puck):

System: Puck, Earth

Surroundings: Thread

Energy Principle: $E_f = E_i + W$

$$K_f + U_f = K_i + U_i + W$$

$$\frac{1}{2}mv_f^2 + mgh_f = \frac{1}{2}mv_i^2 + mgh_i + W$$

$$\frac{1}{2}mv^2 + mgh_f - mgh_i = F_1 d$$

$$\frac{1}{2}mv^2 + mg(h_f - h_i) = F_1 d$$

$$F_1 = \frac{\frac{1}{2}mv_f^2}{d} = \frac{\frac{1}{2}(0.15 \text{ kg})\left(1.6 \frac{\text{m}}{\text{s}}\right)^2}{3.2 \text{ m}} = 0.06 \text{ N}$$

For the second puck (thread wrapped around the puck):

System: Puck

Surroundings: Thread

Energy Principle: $E_f = E_i + W$

$$K_{trans,f} + K_{rot,f} - K_{trans,i} - K_{rot,i} = W$$

$$\frac{1}{2}mv_f^2 + \frac{1}{2}I\omega_f^2 = \frac{1}{2}mv_i^2 + \frac{1}{2}I\omega_i^2 + F_2(L + d)$$

$$F_2 = \frac{\frac{1}{2}mv_f^2 + \frac{1}{2}I\omega_f^2}{L + d} = \frac{\frac{1}{2}mv_f^2 + \frac{1}{2}I\left(\frac{v_f}{R}\right)^2}{L + d}$$

$$F_2 = \frac{\frac{1}{2}(0.15 \text{ kg})\left(1.6 \frac{\text{m}}{\text{s}}\right)^2 + \frac{1}{2}(1.875 \times 10^{-4} \text{ kg} \cdot \text{m}^2)\left(\frac{1.6 \frac{\text{m}}{\text{s}}}{0.05 \text{ m}}\right)^2}{(6.4 + 3.2) \text{ m}} = 0.03 \text{ N}$$

The first puck requires a larger force (0.06 N as opposed to only 0.03 N) to reach the same speed as the second puck.

Figure C.10 – “Two pucks” given solution #1 with breakpoint

Solution #2



For the second puck (thread wrapped around the puck):

System: Puck

Surroundings: Thread

Angular Momentum Principle: $\vec{L}_f - \vec{L}_i = \vec{\tau}_{net}\Delta t$

$$\vec{L}_f - \vec{L}'_i = \vec{\tau}_{net}\Delta t$$

$$\vec{L}_{trans,f} + \vec{L}_{rot,f} = \vec{\tau}_{net}\Delta t$$

$$\vec{r} \times \vec{p}_f + I\vec{\omega}_f = \vec{r} \times \vec{F}_2\Delta t$$

$$\langle 0, p_x^R \vec{p}_{fx}, 0 \rangle + I \langle 0, \omega_f, 0 \rangle = \langle 0, p_x^R \vec{F}_2, 0 \rangle \Delta t$$

(The puck will spin counter-clockwise as the thread unwinds, so by the Right Hand Rule $\vec{\omega}_f$ is in the y direction)

For the y-components: $Rp_{fx} + I\omega_f = RF_2\Delta t$

$$Rm\cancel{v_f} + I \left(\frac{v_f}{R} \right) = RF_2\Delta t, v_{fx} = v_f \text{ since the pucks travel in the x direction.}$$

$$F_2 = \frac{Rm\cancel{v_f} + I \left(\frac{v_f}{R} \right)}{R\Delta t}$$

$$F_2 = \frac{(0.05 \text{ m})(0.15 \text{ kg}) \left(1.6 \frac{\text{m}}{\text{s}} \right) + (1.875 \times 10^{-4} \text{ kg} \cdot \text{m}^2) \left(\frac{1.6 \frac{\text{m}}{\text{s}}}{0.05 \text{ m}} \right)}{(0.05 \text{ m})(4 \text{ sec})}$$

$$F_2 = 0.09 \text{ N}$$

For the first puck (thread attached at the center of the puck):

System: Puck

Surroundings: Thread

Angular Momentum Principle: $\vec{L}_f - \vec{L}_i = \vec{\tau}_{net}\Delta t$

$$\cancel{\vec{L}_{trans,f}} - \vec{L}_{trans,i} = \vec{\tau}_{net}\Delta t$$

$$\vec{r} \times \vec{p}_f = \vec{\tau}_{net}\Delta t$$

$$\langle 0, p_x^R \vec{p}_{fx}, 0 \rangle = \langle 0, p_x^R \vec{F}_1, 0 \rangle \Delta t$$

For the y-components: $\cancel{Ap_{fx}} = AF_1\Delta t$

$$mv_f \cancel{v_f} F_1 \Delta t$$

$$(0.15 \text{ kg}) \left(1.6 \frac{\text{m}}{\text{s}} \right) = 0.24 \text{ kg} \cdot \frac{\text{m}}{\text{s}} = F_1 \Delta t$$

$$F_1 = \frac{0.24 \text{ kg} \cdot \frac{\text{m}}{\text{s}}}{\Delta t} = \frac{0.24 \text{ kg} \cdot \frac{\text{m}}{\text{s}}}{4 \text{ sec}} = 0.06 \text{ N}$$

The second puck requires a larger force (0.09 N as opposed to only 0.06 N) to reach the same speed as the first puck.

Figure C.11 – “Two pucks” given solution #2 with breakpoint

Solution #3



For the first puck (thread attached at the center of the puck):

System: Puck

Surroundings: Thread, Earth, ice sheet

Momentum Principle: $\vec{p}_f = \vec{p}_i + \vec{F}_{net}\Delta t$

$$\vec{F}_{net} = \langle F_1, F_{ice} - F_{Earth}, 0 \rangle$$

$$\langle p_f, 0, 0 \rangle = \langle p_i, 0, 0 \rangle + \langle F_1, F_{ice} - F_{Earth}, 0 \rangle \Delta t$$

$$\langle mv_f, 0, 0 \rangle = \langle F_1, F_{ice} - F_{Earth}, 0 \rangle \Delta t$$

<u>For the y-components:</u> $0 = (F_{ice} - F_{Earth}) \Delta t$ $F_{ice} - F_{Earth} = 0$ $F_{ice} = F_{Earth}$	<u>For the x-components:</u> $mv_f = F_1 \Delta t$ $F_1 = \frac{mv_f}{\Delta t}$ $F_1 = \frac{(0.15 \text{ kg}) (1.6 \frac{\text{m}}{\text{s}})}{4 \text{ sec}} = 0.06 \text{ N}$
---	---

For the second puck (thread wrapped around the puck):

System: Puck

Surroundings: Thread, Earth, ice sheet

Momentum Principle: $\vec{p}_f = \vec{p}_i + \vec{F}_{net}\Delta t$

$$\langle p_f, 0, 0 \rangle = \langle p_i, 0, 0 \rangle + \langle F_2, F_{ice} - F_{Earth}, 0 \rangle \Delta t$$

$$\langle mv_f, 0, 0 \rangle = \langle F_2, F_{ice} - F_{Earth}, 0 \rangle \Delta t$$

<u>For the x-components:</u> $mv_f = F_2 \Delta t$ $(0.15 \text{ kg}) (1.6 \frac{\text{m}}{\text{s}}) = F_2 \Delta t$ $F_2 \Delta t = 0.24 \text{ kg} \cdot \frac{\text{m}}{\text{s}}$ $F_2 = \frac{0.24 \text{ kg} \cdot \frac{\text{m}}{\text{s}}}{\Delta t}$ $F_2 = \frac{0.24 \text{ kg} \cdot \frac{\text{m}}{\text{s}}}{4 \text{ sec}} = 0.06 \text{ N}$	<u>For the y-components:</u> $0 = (F_{ice} - F_{Earth}) \Delta t$ $F_{ice} - F_{Earth} = 0$ $F_{ice} = F_{Earth}$
--	---

Both forces need to have equal magnitudes (0.06 N) in order for the pucks to have the same speed at 4 seconds.

Figure C.12 – “Two pucks” given solution #3 with breakpoint