NORTHWESTERN UNIVERSITY

What Constitutes a Mathematical Proof?

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Psychology

By

Norman Mitchell Eliaser

EVANSTON, ILLINOIS

December, 1998

ABSTRACT

What Constitutes a Mathematical Proof?

Norman Mitchell Eliaser

While much of the research in the teaching and understanding of mathematics has focused on algebra problems, little attention has focused on what problem solvers believe constitutes a legitimate mathematical proof. Four classes of claims, generated by crossing quantifier (universal or existential) with polarity (show that the statement is true or that the statement is false) were studied. Participants were asked to generate or verify proofs for claims in each of the classes. Experiments not only gauged performance and strategy choices, but also what the provers believed was an adequate proof for a given claim, what criteria they used to categorize claims, and the effects of hints about differences among claims on their performance and strategy choices. Common proof strategies involve using a numeric example, multiple numeric examples, or variables. Participants were found to have a bit of trouble representing claims using variables and recognizing that universal claims require abstract proofs.

Acknowledgments

I would like to thank my committee for being supportive throughout the process of dissertation-writing.

Lance Rips has been a great advisor, and I think the most important lesson he gave me was something that I don't think I would have heard from an advisor: "Lower your criterion." In my efforts to try to write a really good third year paper, I was indeed setting my criterion too high in order to get work done. Having the insight to just write, and then go back and revise later, I was much more productive!

Of course, the rest of my committee has been a great help too–I think my "course of sailing" was redirected very well after having my first proposal defense. A big thanks goes out to Doug Medin, Karen Fuson, and even though he was no longer here to sit on my final dissertation defense, Jerry Faries. Their comments, ideas and support have been appreciated.

Very many other people contributed in some way to the dissertation, and did so in very many different ways. I'm sorry if I unintentionally leave some of them out. My lab-mates provided a tight support network, both to ensure that I got my work done, and by either reading over and commenting on drafts (Jeremy Bailenson, Sarah Brem, and Mike Shum) or by helping listen to awful audiotapes for transcription/scoring or scoring participants' proofs (Satya Chheda and Aimee Ericson).

Even someone outside the lab got to help out with scoring tapes and it was quite appreciated. Tuti Reinhart Miller not only read over drafts of my dissertation, but she also helped act as a second rater for some of the awful audiocassettes. Tuti was one of many people who also got to hang around in the MacLab for hours with me, providing social and academic inspiration.

You folks hanging around too much in the MacLab at all hours are all to be thanked for moral support, as well as fun times. This includes, but is not limited to, people like Khytam Dawood ("I'll be back down in 10 minutes..." 45 minutes later, when I'd call upstairs, "oh, I'll be down soon"). :-)

Other support, in and out of the MacLab, comes from an assortment of other people. Adam Rosenbaum, who's spent hours working alongside me somewhere, and talked about everything from soup to nuts (and their uses in Passover cooking) with me, especially deserves a hearty thanks. Thanks also to my friends all around, including Alexandra Chartier, Suzanne Chronowitz, Nava Cohen, Shelli Feist, Mike Gelman, Steve Gnagni, Dr. Ted Karp, Robert Lebowitz, Aviva and Sinai Megibow, Anya and Scott Pakin, Dave Patrick, Dr. Judith Schwartz, Kath and Ted Schwitzner, Heath and Lisa Sherman, Andrew Shurak, Daniel Vaisrub, Scott Weiss, and Melissa Wu. For LaTeX advice, thanks to Jeff Rouder. For SASsy comments, thanks to Mike Bailey.

Unlikely as my life events were during grad school, it still wouldn't have been possible to get through some of the harder times with research without the support of Jennifer Fox. We had a lot of wonderful experiences, and a lot of surrealistic ones. To help deal with surrealism in various times of grad school as well as to answer various questions, I've had a warm and bright set of rabbis to help out. Thanks to Rabbi Michael Balinsky, Rabbi Mark Dunaevsky, Rabbi Shoshana Gelfand, Rabbi Dov Hillel Klein, and Rabbi Elliott Kleinman.

Some people all too often forgotten when dissertations are written deserve lots of thanks–the office staff. They not only provide lots of the support work for the department in the way of typing and xeroxing; they are the glue that holds the department together. Thanks to all who were here while I was here, including Karie Banach, Brenda Brown, Ulyses Grant, Tracy Halbert, Florence Hinklin, Angela Hunter, Nick Kanel, Anne Rogerson, Rachel Rufer, Joan Stahl, and especially Florence Sales! Also, thanks to all the students who participated in the experiments. You guys aren't thanked often enough either.

How many roaches should be thanked for this dissertation? Just wondering... The cleaning staff who've reduced the number of roaches in Swift Hall should be thanked.

I guess I should stick to thanking people... For the times when I had trouble deciding what to do with my academic career (do I really want to finish this dissertation stuff?!?), I would like to thank Beth Levy for being a great sounding board.

Regarding sounding boards, the best one of those I've found is Shira Karp. Shira, thanks for being there to listen, to share, and to discuss all of life's ins and outs, and for love, support, and bobcats l'shmaH.

All the credit for this dissertation, and my education in general should go to my parents. You've always wanted the best for me, and have worked hard to see to it that I get as much in the way of resources as possible. You're the best! I love you guys!

to Larry I. Jonas, The College Comedian

"How old do I look?"

Contents

List of Tables

List of Figures

Chapter 1

Introduction

What Constitutes a Mathematical Proof?

In order to prove a mathematical claim, it is necessary to know the discourse rules for proof, and be able to apply those rules. In some contexts, a single example may be sufficient to show support for a claim. In others, some abstract representation (such as a diagram, or set of equations) may be needed.

At the root if this discussion is the question, "what does it mean to someone to prove something?" Martin and Harel (1989) suggest that proving a statement may well mean convincing the reader of that statement. People not experienced in math may take the dictionary definition of "proof" of "evidence sufficent to establish a thing as true or believable." (Ref for Random House College Dictionary revised edition. Jess Stein, Editor in Chief; L.C. Hauck, Mnging Ed; P.Y. Su, Senior Defining Editor. 1984. Random House Inc. NY).

In order to convince someone of the believability of a statement, one does not necessarily need to use a rigorous mathematical proof. In order for a proof to be sound mathematically, it requires iron-clad logical reasoning based upon the given assumptions and known theorems and axioms. The difference here may described as a continuum between arguments which seem convincing to those which are deductively sound.

Suppose you wanted to convince someone that all people are born with an appendix. That person may believe you if you show him/her one cadaver with an appendix. That person may require that you show a living person with an appendix. S/he may require that you show a man and a woman with an appendix. Or the person may not believe you until you show that every single human being currently alive (who has not gone through an appendectomy) has an appendix. Any "proof" like this will not do–it is always possible (albeit unlikely) that the appendix actually appears instants, moments, or years after birth in some of these people. Yet we are convinced that people (normally) are born with an appendix. We accept some type of inductive argument for this "proof," namely that since some (non-exhaustive) set of people have appendices, all people do. It may require a larger or smaller set to convince someone, but we cannot show that literally all people have (or, still harder, are born with) an appendix. In contrast, mathematical proofs require deductive methods, in which each step is logically based upon the given information and the known axioms and theorems.

Such differences in level of rigor depend on the audience. Over the course of history, mathematicians have required different levels of rigor in their proofs. Kleiner (1991) points out that the use of variables as the symbolic notation with which we are familiar today was not developed until the 16th and 17th centuries[1], before which proofs used numeric examples rather than variables to demonstrate universal claims. Kleiner also notes that in recent years, probabilistic proofs, those in which a proposed claim is shown to be true within a certain margin of error have been used in domains such as the testing of primality of numbers.

Mathematical claims may be characterized by four classes. These classes are divided on two dimensions. The first is what I will refer to as “polarity”; the second will be termed “quantifier.” A claim’s polarity indicates whether it states that some proposition is true or false. For example, the claim that “all prime numbers greater than two are odd,” has a positive polarity. In contrast, the claim, “there are no even prime numbers greater than two” has a negative polarity.

The quantifier of a claim may be universal or existential. Existential claims are claims for which only the demonstration of a particular example is necessary (e.g., There is a number divisible by 6 which is not odd). A proof for such a claim would be the following: 12 = 2 x 6. 12 is divisible by 6. Since 12 is divisible by 2, it is even, and thus not odd. A universal claim is one which applies to all members of a class (e.g., All products computed by multiplying an even number by an odd number are even). Such a claim requires a proof using an abstraction for the general case, rather than just a specific example. Abstraction for a universal claim may involve a textual description, a diagram (e.g., a Venn diagram), or most frequently, a representation of

[1] In ancient Egypt, Ahmes used the symbol hau, which means “heap,” to represent the unknown (Cajori, 1910). He used hau in the following context: “Hau its 2/3 its 1/3 its 1/7 its whole gives 33,” which represents “$2/3\,x + x/3 + x/7 + x = 33$.” Similar uses existed in other ancient cultures. In most Western cultures, though, Kleiner’s description is true.

the problem using variables:

Let x and y be integers.

Then $2x$ is even and $2y+1$ is odd.

$(2x)(2y+1) = 4xy + 2x = 2(2xy + x)$

The product is even since it is a multiple of 2.

Students who are unfamiliar with writing mathematical proofs may run into several problems when attempting to prove claims. Besides having trouble knowing the appropriate level of rigor for their proofs, students may have problems understanding the meaning of a quantifier in a claim and knowing how to correctly represent the claim to be proven. So, for the universal claim above, a student might not realize the need to use variables to demonstrate the universal nature of the claim. Such a problem indicates that s/he does not appreciate what is to be proven.

Alternatively, the student might not know how to represent a generic even and generic odd number in proving the above claim. This would indicate a problem of representing the claim using the "tools of the trade," variables. Difficulties here involve representing or manipulating the problem rather than missing the goal of showing the claim for all cases. Both types of problems prevent students from writing successful proofs.

After outlining work that deals with the understanding of quantifiers, proofs, and mathematical problems in general, this paper addresses the questions of (1) what constitutes a mathematical proof for a problem solver, and (2) how might that conception be improved so that a person's proof techniques are improved.

1.1 Understanding Quantifiers

While not much research has addressed mathematical proof, there has been attention paid to the understanding of quantifiers (such as for all, for some, etc.) in language, and to problem solving in contexts which parallel mathematical proof. In this section, I outline some of this work which can be brought to bear on the understanding of performance on proofs. This includes work on quantified statements, which may yield insight into how quantifiers are misinterpreted in proofs, and studies on mathematical proofs, which demonstrate that people are willing to accept inductive arguments for universal claims.

Quantifiers are studied by McCawley (1993) from a linguistic standpoint. He takes a look at the various rules of inference based upon the universal and existential quantifiers. In doing so, he notes that a step (which is less than transparent) in proofs of universal claims is the line in which one picks an arbitrary element (p.45):

> The words "pick an arbitrary element" are misleading, since there isn't any "picking" involved (you don't, for example, say "Let's pick 37, since that's as arbitrary an odd number as you're likely to find"). What is going on, rather, is that a subproof is set up in which a supposition is made involving something that has not hitherto appeared in the proof. The only "information" in which that thing appears is the supposition (e.g., all that you know about n is that it is an odd number). For provers to adequately prove a universal claim, they must not only understand that the arbitrary element is not specified, other than its given properties, but also that from the fact that only some of its properties were specified,

conclusions about it can be drawn about any member of the set of elements with those properties.

This rule is known as ∀-introduction in natural deduction systems. A parallel rule called ∃-introduction is asserted for existential claims. (∃-introduction can be stated in the following way: Finding a single example which demonstrates a property constitutes proof of an existential claim.) Without applying ∀-introduction, a person cannot prove a universal claim. So, misunderstandings of these inference rules can lead to inadequate proofs (one would be drawing faulty inferences if s/he believed that a single example adequately demonstrated a universal claim, and allowed ∀-introduction).

More misunderstandings of quantified statements have been studied by Newstead and Griggs (1983). They found that participants who drew inferences from statements with the qualifier "some" believed that "Some A's are B's" implied that there were also A's which were not B's, following maxims of discourse. Thus, they did not believe that "All A's are B's" would provide support for "Some A's are B's." To the contrary–they thought that "All A's are B's" meant that "Some A's are B's" was false. If such reasoning carries over to people's understanding of proof classes, we would expect to see that people may not properly understand the quantifier used in an existential claim. People with this misunderstanding who need to prove an existential claim would never try to prove a stronger universal claim since it would not successfully support their misconception of the existential claim. This may lead to avoidance of variables in choosing a proof strategy for an existential claim, since a proof using variables would demonstrate the universal nature of the claim. There are at least two other reasons why students might be predicted to use variables infrequently.

One is that the solver may just find it easier to give one (some) example(s) than to use variables. If examples are easier, then variables would be skipped. Alternatively, people might not recognize that there are any claims which require a proof using more than just a single example.

1.2 Understanding Proofs

Without explicitly focusing on quantifiers in their study, Koedinger and Anderson (1991) show that many participants believe that a single example provides sufficient proof for a universal claim. Koedinger and Anderson gave subjects a geometric diagram along with a question that involved a geometric relation and was always phrased "If STATEMENT1 then must STATEMENT2? For example, "if angle TQU ≅ angle RQU and angle QTU ≅ angle QRU, must angle STU ≅ angle SRU?" Subjects were supposed to circle YES or NO to answer the question. The diagram (model) given always fit the antecedent STATEMENT1, but may have been only a special case of that antecedent, and was varied to lead subjects to believe or disbelieve STATEMENT2. In such a way, they varied the diagram to be a "Looks-true" or "Looks-false" model, and varied the problems so that the correct answer was YES or NO. Participants were forewarned that the diagrams could be misleading. Participants were more apt to judge the correct answer to be YES when the items were YES items in actuality and when they Looked-true. Koedinger and Anderson explain the results in the following way: Participants apply a deductive strategy by trying to find a proof and then answer yes if they can find one. If they cannot find a proof, they are said to switch to an inductive strategy– to "guess" at the answer based on the diagram. This explanation accounts for the greater difference in performance (on truth judgments made about a diagram) between the easy[2] and difficult YES problems than that for the NO problems. Since easy YES problems should yield a proof often, those yield high proportions of YES responses. In contrast, the hard YES problems probably yield proportions of YES responses similar to all problems for which there is no correct proof. Apparently, participants made guesses based on the potentially misleading diagrams rather than by working out the answer. Each diagram only represents a single instance, so a proposition describing the diagram cannot correctly be used to draw conclusions about all cases. The "must" of each question is asking the subjects if STATEMENT 1 always implies STATEMENT 2. We can conclude by subjects' answers that they were either misinterpreting the meaning of the universal quantifier implied by the word "must" or that when it is too difficult to find a proof, the participants were willing to guess based on the instance present. However, even allowing such a guess to guide their answer to the question of "Must..." shows a lack of appreciation for the quantifier. At best, they are allowing an inductive argument to take the place of a deductive one for a universal claim.

Another context in which the interpretation of the universal quantifier's definition may be tested is by asking people to rate the quality of attempted proofs. Martin and Harel (1989) sampled people who were preparing to be elementary school teachers, and had them rate mathematical proofs. The preservice teachers were given two "generalizations" (universal claims) along with three or four "attempted proofs"

[2] Easy/difficult distinctions are based upon how many conceptual steps are taken in proof planning based on Koedinger and Anderson's (1990) earlier work.

for each. An example of one of their generalizations is “If a divides b, and b divides c, then a divides c.” The teachers were asked to rate each of the proofs for the statements on a four point scale. The “attempted proofs” were inductive or deductive in style, and varied in what Martin and Harel called “form.”

The inductive forms of proofs included “example,” “pattern,” “big number,” and “example and nonexample.” The example form gave two examples which supported the claim (but left out the step demonstrating the actual property–e.g., saying that 48 is divisible by 12 without specifying the other factor). A pattern was essentially a chart in which 12 supporting instances were presented. The big number was one instance of the claim which works for a big number. The example and nonexample form shows a supporting example for the claim, as well as an example using different premises than those in the claim’s antecedent for which the consequent is not true.

The deductive types of proofs included “general proof,” “false proof,” and “particular proof.” The general and false proofs both employed variables, with the general proof adequately proving the claim, and the false proof containing some incorrect reasoning (i.e., skipping a necessary step or falsely assuming that one statement implies another). The particular proof showed a specific example for which the claim was true, and left out no steps (differentiating it from the example proof–also, the particular proof only uses one example while the example proof uses two).

A surprisingly large number (80%) of the teachers gave “high ratings” (of 3 or 4 on a four-point scale) to at least one inductive argument, with over half giving a rating of 4. Martin and Harel did not find differences among ratings of the various inductive arguments.

The deductive proofs were also rated highly. Among these, over half of the participants gave high ratings to the particular proofs (which only provide a single example), and more than half also gave high ratings to the general (accurate) proof. False proofs were judged lower than the other two types, with more than half of the ratings being low (1 or 2).

It is telling that participants were often willing to accept the particular, example, and example/nonexample proofs. Either these participants did not recognize the presence of the universal quantifier or they did not know which proof techniques were appropriate for universal claims. Even when given proofs of general type and particular type, many do not contrast the two in their ratings. Apparently, an example may be sufficient by their standards to draw the universal conclusion. Having such a misunderstanding about the quantifier would prevent a prover from being successful.

1.3 Representing The Problem

1.3.1 Representing Geometry Problems

In setting up a proof, after recognizing what needs to be proven (which requires an understanding of the quantifier), the prover must create a representation for the claim. One form of abstract representation involves diagramming the suppositions described in the problem. Koedinger and Anderson (1990) posit the Diagram Configuration model (DC) for expert planning of how to prove geometric theorems in which a diagram is part of the given information. DC focuses on subgoals based upon parsing a diagram, encoding statements describing the diagram, and a search for what they call diagram configuration schemas. These schemas are clusters of geometry facts which are known about a particular prototypical image. Each schema contains what are referred to as "part-statements" and a "whole-statement." The part-statements (e.g., side AB=side AC [in a triangle]) express the relationships between parts of the configuration, while the whole-statement refers to the configuration as a whole (e.g., triangle ABC is congruent to triangle DEF). The problem solver looks at the diagram corresponding to the particular problem, and establishes a subgoal of proving that a particular whole-statement is true if the problem diagram yields part-statements which could be parts of a familiar configuration schema. In this way, the diagram configuration schemata act as abstract representations of possible proof subgoals. These more abstract representations simplify the situations being evaluated in the proof–Larkin and Simon (1987) point out that diagrams simplify search by grouping together information which is used together and allow for automatic perceptual inferences to take place.

1.3.2 Representing Word Problems

The type of work by Koedinger and Anderson in this area is quite rare–not much attention has been given to representations used for mathematical proofs in the literature, but there has been more of a focus on representations for mathematical word problems. In the experimental section of this paper I explore the task the prover has in turning claims into proofs. Some central ideas for word problems can be applied to proofs as well since the two types of problems (word problems and proofs) require representing a given problem description by a mathematical statement (usually in a written format).

The first step in any problem solving situation is building a representation of the problem (Resnick & Ford, 1981). When given a word problem, the problem-solver must figure out what the words mean–in a problem with a sentence like "Sharon is 6 inches taller than Danny, who is 1 inch shorter than Jesse" (p214), the solver must understand "who," "taller," and "shorter" and the relationships between them. So linguistic processing is necessary to rephrase the problem in one's own words. After that, various strategies can be applied to the representation in order to solve the problem. For typical problems in mathematics, Resnick and Ford believe that the important intellectual work is over once a problem's representation has been worked out.

One view of problem solving suggests that when representing and solving word problems, people employ one or more of four strategies (Tabachneck, Koedinger, & Nathan, 1994). These four strategies are termed (1) Algebra, (2) Guess and Test, (3) Verbal Mathematics, and (4) Use of Diagrams. Algebraic techniques involve changing the word problem into algebraic assignments and equations, and then performing transformations on those which would yield a solution. The guess and test method involves making successive guesses, and essentially substituting the guesses into the equation using trial-and-error. A verbal mathematics approach involves talking aloud to describe the transformations being carried out, which involves a greater demand on working memory. In contrast, diagrams allow students to have a more concrete representation of the problem, with many interrelations expressed in a picture. Transformations are then made upon the picture's elements.

It is apparent that each of these strategies, or very similar ones, could be applied to proofs. The strategy which they term algebra (i.e., the use of equations) for proofs is an effective means of demonstrating universal claims (or existential ones). Guess-and-test, a trial-and-error method is a good way to test existential claims, where only a single example is necessary. Verbal mathematics, speaking through the steps in a proof, would (as in word problems) involve more demands on working memory than a written approach. Finally, the use of diagrams is most useful in geometric proofs, or proofs which may have an isomorphic diagram form.

Within the domain of word problems, Tabachneck, et al. (1994) have shown that use of multiple strategies is more frequently successful than the application of just one. In their study, when an impasse was reached, subjects who stuck to the strategy they had initially employed always failed at completing their problems, while 79% of those who switched strategies were successful in solving the problem. Flexibility in representation, therefore, gives a problem-solver a huge advantage. Determining which representation is most beneficial is not clear-cut, though. While the verbal mathematics approach without explicit variables may be easier for a novice to understand, its greater load on working memory makes its use potentially restrictive. In contrast, an algebraic representation is an abstract representation of the problem, which is not very clearly related to the problem or its solution. At the same time, the algebraic representation is often the easiest to manipulate, especially for complex problems. In the experimental section, I closely examine the various strategies applied in written proofs.

As the Tabachneck et al. (1994) work suggests, the goal in teaching students about multiple representations is to give them greater flexibility in solving problems. Dufour-Janvier, Bednarz, and Belanger (1987) suggest that making reference to alternate problem representations in teaching gives students more tools for learning new concepts. However, they note the danger that multiple representations for the same concept can unfortunately lead to confusion because children often believe that two distinct representations correspond to two different problems. Another factor affecting interpretation of the representations used is context–students always try to make sense of what they are learning about in terms of their context (e.g., Greeno, 1989). They "will interpret the symbolic operations they learn as being about something, but what they are about in the students' understanding may be quite different from what we intend"(p.293). Namely, he posits that the formulas are often thought to be describing some abstract notations and not the quantities, variables and functions which the teacher intends to depict.

Rather than concentrating on formulas in word problem representation, Kintsch and his colleagues (e.g., Dellarosa Cummins, 1992; Dellarosa Cummins, Kintsch, Reusser, & Weimer, 1988; Kintsch & Greeno, 1985; Nathan, Kintsch, & Young, 1992) use the idea of a problem model as the key to representing a word problem. The problem model is the representation of the information contained in the text of a problem. The model involves simple schemata for sets of objects described in the word problem (e.g., apples), frames for propositions (e.g., have (Joe, five (marbles))), and higher order schemata that state which information is known and which is unknown for each problem type. They assert that not all of the initial information is maintained in the representation; the only information kept is that needed when solving for unknowns in the high level schemata which are active in working memory. The strategies applied at any particular time are productions which check the currently known information, and perform the action which moves the problem toward finding the unknown information.

Dellarosa Cummins, Kintsch, Reusser, and Weimer (1988) showed that when subjects were asked to recall word problems aloud either before or after solving them, the accuracy of solution was, as would be expected, based on correct recall of the problem's structure. Similarly, Weaver and Kintsch (1992) gave subjects pairs of word problems in which structure and equation type were both varied. Problems were of either the same or different structure (the relationships between components of the problem could be sketched the same way for two problems in a "same structure" pair). Problems with the same equation required the solution of an equation of the same form (See Table 1 for an example of each type). Subjects were able to perceive and make use of the structural similarities to solve the problems. Weaver and Kintsch found that when given a training session in which the structural similarities were highlighted, subjects' performance improves, while focusing on the equations which matched did not help very much. Hence, an appreciation of structure of the problem is necessary for setting up an adequate representation.

In describing the structure of word problems in algebra, Nathan, Kintsch, and Young (1992) distinguish between the representation for events, which they term "the situation model," and the representation created using formal relations, the "problem model" (essentially the relevant equation). Nathan, et al. describe a problem's text as being understood by their ANIMATE[3] learning environment in the following fashion: (1) a textbase consisting of propositions is formed, (2) the textbase is organized into a situation model, and (3) the textbase is then mapped into a problem model, which focuses on the algebraic problem structure. The situation model includes relationships in qualitative relative terms, while not using the numbers found in the text of the problem.

This representation of a word problem, they argue, is no different from that of a story. Problem solving difficulties occur when the problem model does not correspond to the situation model, or if the original situation model was not correctly formed. Testing this tutoring model, they found that an environment which supports situation-based reasoning leads to an improved ability to specify the solution properly. Students are also better able to produce equations which matched the text of word problems and to produce word problems which matched equations when situation-based reasoning is emphasized.

An example of a problem worked on by ANIMATE is a distance-rate-time word problem: A plane leaves Denver and travels east at two hundred miles per hour. Three hours later, a second plane leaves Denver on a parallel course and travels east at two hundred fifty miles per hour. How long will it take the second plane to overtake the first plane? (Nathan, Kintsch, & Young, 1992, p333) ANIMATE represents this as a set of propositions, each of which look like "LEAVE (Plane1, Denver, Time1)." The situation model corresponding to this problem is a representation which shows

[3]ANIMATE is a computer modeled learning environment, not aware of the problem or the student's actions, which gives equations meaning through computer animations.

the two planes flying on a parallel course, and notes that there is some moment when the second plane passes the first. Nathan, et al. (1992) do not specify the details for this representation, but say only that all that needs be done is to "represent the esssential, common parts of everyone's situation model: one moving object overtaking another"(p333). Presumably, for the situation model for the land plot problem in Table 1, the representation must have "a single dimension of one resource is larger than that dimension for the second resource while another dimension is larger for the second resource than for the first; the composite of the two dimensions is equal." Such a representation does not seem trivial to obtain. Nathan et al. (1992) hypothesize that the problem model is developed from the reader's situation model based upon which schema the problem solver recalls–the key is that the student has the context of the situation supporting his/her choice of a formula. For the problem involving two planes, ideally the "Distance = Rate x Time" schema is called up.

An extension of this type of problem model involving sets, frames and schemata might be applied to proving claims as well. In proofs, the sets might no longer be sets of objects. Rather, they would be mathematical sets, such as the odd numbers. Propositional frames would describe properties known of the sets, in an analogous fashion to the propositional frames of the arithmetic problems. The higher level schemata would similarly map onto the proof domain in terms of maintaining an awareness of which facts or properties are known and which are unknown. The situation model for a proof is a description of the higher order relationships among different sets or set elements described in the proof. Production rules would suggest which strategies to apply when certain schemata are active. In terms of the proof classes, an expert ought to have different productions for the different classes. (One only needs to find a single example for existential proofs; however in a universal proof, finding a single example would not be sufficient. Thus the productions for a universal proof would not containthe action of "try a single example," unless that were part of a more sophisticated proof technique like proof by induction.)

A different perspective also integrating schemata with problem solving is suggested by Reed (1993), who posits that five characteristics of schema theory apply to word problems. These can also provide insight into proof methods. Abstraction involves learning which aspects of a problem are necessary to solve it–finding the relevant parts of a problem. After the necessary parts are isolated, the correct values must be instantiated in the "slots" of the equation. People working on a problem should be able to predict both the values which are reasonable for solutions or other problem elements and the information which is necessary to solve the problem. People who solve problems use induction to decide how they should generalize from a single problem to a problem type. Finally, Reed notes that the way in which subjects classify problems follows a hierarchical organization; such an organization leads to problems with similar cover stories being placed in the same basic level category, and those sharing a common equation type (varying which quantities are knowns and unknowns) in subordinate categories, while the superordinate could have stories with the same solution but different cover stories (p.47).

Such an organization is used by Reed to describe the way in which students categorize problems and subsequently use the equations associated with problem categories in order to solve the problems. Suppose a student is to solve a rate problem describing one car overtaking another car. The necessary formula for this problem is Rate1 x Time1 = Rate2 x Time2. Knowing which values (the two rates and two times) are relevant, and generating this equation corresponds to Reed's abstraction process, while instantiation involves knowing how to fill the slots of the equation (variables). Thus, one function of schemata in solving a problem is to direct attention to the relevant variables. Similar schemata are necessary in mathematical proofs, namely ones in which the relevant aspects of a claim or current state of knowledge are attended to, and the necessary steps are executed based upon the facts already established.[4]

1.3.3 Development of Representations

The schemata that students have at any point are developed over practice with proofs. Schoenfeld and Herrmann (1982) studied the effect of instruction after a one-month problem solving course on problem perception. The focus of the problem-solving course was a set of heuristics for approaching problems, and students in the course were told that they ought to be sure they fully understand any problem statement before going on to solve it. Schoenfeld and Herrmann studied the differences

[4]Suppose, for instance, you are asked to show that subtracting one from a number and doubling it is the same as doubling the same number and then subtracting two. When you are required to show that two sets of operations on the same starting number yield the same result universally, you may have a schema saying to use "x" for each instance of that number. If, however, your proof task required that you describe a number as even, you may have a schema saying to use "2x" for that even number, where x is an integer. The repeated use of a number would elicit repeated use of the same variable, while evenness would elicit that "2x" representation.

in novices' and experts' problem perception among categories yielded in their sorts of math problems (contrasting novices' categories [before instruction] with those of experts [after instruction]). In their cluster analysis, more of the categories generated in the expert group correspond to divisions based on the deep problem structure than in the novice group. "Deep" problem structure was determined by a set of math experts, namely math professors. A control group, who took a computer science course of comparable workload, was not found to differ from the novice group. Schoenfeld and Herrmann's (1982) work parallels research by Chi, Feltovich, and Glaser (1981), who found that novices sort physics problems based on superficial features (often these are objects referred to in the problems, such as springs or pulleys), while experts sort on the basis of structural relationships like conservation of momentum. Similarly, as expertise increases, the use of structure in sorting proof claims would likely increase.

1.4 Outline of the course of study

Ostensibly, the most important structural element to a claim is its quantifier. While the literature reviewed above has addressed issues of representation and problem solving, only Koedinger and Anderson's work has addressed proofs, and their work has not focused on quantifiers. When confronted with a claim to be proven, a problem-solver must understand the claim (including its quantifier), choose an appropriate initial representation (e.g., use of an example, use of variables, etc.) for the proof, and operate on that representation to generate an adequate proof.

So, there are four possible problems which participants may have in adequately proving a claim. The first two deal with recognizing the quantifiers and understanding what would be required for each type of proof. These involve knowing what the initial state (givens) and goal state (a completed proof having used the right method) are:

1. They cannot distinguish between the universal and existential quantifiers.

2. They don't know which proof strategies are appropriate for each quantifier (this would be akin to a lack of understanding of McCawley's two introduction rules).

The second two deal with the abilities needed to use appropriate representations (especially for universal proofs). These involve the ability to set up the representation for the initial state and apply the correct operators to get to the goal state:

3. They cannot represent claims using variables.

4. They don't have the skill to manipulate variables.

These four possibilities are addressed by the experiments, and are reviewed in the conclusions at the end of the paper.

In order to generate a proof for a claim, it is necessary to recognize that claim's quantifier. Once a student appreciates the difference between a universal and existential claim and has established a useful representation for a given claim, s/he can apply some strategy to prove the claim.

Since universal claims require a proof using more than specific examples, students should be aware that variables would be an essential representation[5] for such proofs. Yet not all problems are ones in which using variables would be ideal. For existential proofs (either existence proofs, in which a single example is sufficient to

[5]As noted before, an isomorphic proof using a verbal description could adequately prove a universal claim. Also, note that in special cases in which the universal is restricted, enumeration by all possible examples is an acceptable proof (e.g., "Show that all even primes are divisible by 2").

demonstrate that some claim is true, or counterexample proofs, in which one is sufficient to show that some claim is false), it may be less cumbersome to use a test case than to work through a proof with variables.

The first experiment examines what strategies students employ for simple proofs in each of four proof classes, made up of existential and universal claims, both positive and negative. Positive claims are proofs of the form, "Show that there is a number for which some property is true" (existential positive) or "Show that for all numbers some property is true" (universal positive); while negative claims assume the form, "Show that there is a number for which some property is not true" (existential negative) or "Show that for all numbers some property is not true" (universal negative). After completing all the proofs, participants were asked to go back and rate how well they had proven each claim. In addition, participants were asked to return to the claims whose proofs did not make use of variables and to try to prove them with variables. For the initial proofs, I anticipated better overall performance on the existential claims than on the universals, since existential claims require only a single example for a proof while universals require the manipulation of an abstract representation such as variables. The later attempts to use variables should show how adept the participants are at using variables.

The second experiment is similar to the first, but it reduces the cognitive load on the participants, since a large number did not adequately prove many of the claims in the first experiment. Instead of generating steps in the proofs, participants rate attempted proofs (which were created by the experimenter to balance the proof classes and strategies used, as well as correctness). The strategies were the use of a single example, use of multiple examples, and use of variables. Since the participants did not have to generate the proof steps, difficulties involving generating the representation for each claim would be eliminated, and we could more clearly measure their understanding of which proof techniques are necessary or preferred for each proof class. If participants do understand that universal claims require abstract proofs, they will only rate the variable proofs as successful for those. However, if they miss that point, they may accept proofs using a single example or multiple examples as adequate for the universals.

Experiment 3 revisits the proof classes used in the first two experiments by asking participants to sort the claims into categories. The experiment explores whether explicitly giving them labels for the four proof classes allows them to better apply proof strategies. Beyond that, it compares those participants with others who are allowed to generate their own categories (self-sort) for the claims given to them based upon the strategy they would use to prove each claim. If the participants have an understanding that quantifiers play a key role in determining proof strategy, the self-sort participants ought to sort based upon quantifier. Otherwise, their categories would likely reflect surface features of the claims.

Finally, Experiment 4 provides three hints about differences among the claims which test whether focusing attention on certain aspects of the claims can easily remedy the difficulties students have. These short blurbs tell provers to focus on the quantifier, focus on the scope of the negative (when the claim is to show that something is not true for all examples, a claim's ambiguous phrasing may suggest that something is true for all but some nonempty set of examples or that something is false for every example; those two differ[6]), or give a brief example of how variables may be used in a proof. The changes in performance and strategy choices after the hint ought to give insight into exactly where difficulties in generating proofs lie.

In sum, the project studies what participants believe is an adequate proof, how that varies with the proof classes, and whether the proof classes which mathematicians can perceive are perceptible by less experienced students (or are replaced by some other taxonomy).

---

[6]An example provided by McCawley (1993) for the difference in scope of the negative is the following pair of sentences (p. 40): "John didn't criticize any candidate." "John didn't criticize every candidate."

Chapter 2

Experiment 1 – The impact of quantifier and polarity

The goal of Experiment 1 is to determine whether students have an understanding of when variables are helpful in proofs and when a simple numerical example is sufficient. Variables do not necessarily provide the simplest method for solving a problem. Instead, for certain problems, such as existence proofs or proofs using a counterexample, a single example would be sufficient. In order for participants to recognize which claims necessitate variables, they must be able to differentiate between the proof classes mentioned earlier.

Participants were given 16 claims to prove, each belonging to one of four classes. The four classes are positive universal, negative universal, positive existential, and negative existential. Positive universal claims are those which ask the participant to show that some claim is true in general (e.g., Show that every multiple of 13 will be a multiple of 13 when doubled and 13 is subtracted). Negative universal claims are similar, but ask the participant to show that some negative claim is always true (e.g., Show that every odd multiple of 5 when squared is not divisible by 2). Both the positive universal and negative universal claims require the use of variables to demonstrate the property in question. The other two classes can be proven with only a specific test case. A positive existential claim is one for which the participant can find an example demonstrating that it is true (e.g., Show that there is a number which is a common divisor of 65 and 338). A negative existential claim is similarly one for which a participant can find a test case–but one for which the negative claim is true (e.g., Show that there is a number for which doubling it does not change its parity [whether it is odd or even]). Note that these four proof classes can be viewed in a two by two arrangement in which the universal and existential quantifiers are crossed with positive and negative polarity.

After participants completed the entire packet of demonstrations, I asked them to rate how well they had proven each claim. These confidence ratings on the proofs should give us some insight into whether or not participants believe that their written answers completed the demonstrations requested (for instance, when an example was used for the universal proof class, it is possible that the participant was only trying to convince the reader that the claim "seems to be true, but I can't prove it").

The polarity of a claim should not have much of an impact on the strategy chosen to prove it. However, Clark and Chase (1972) posit that the default value for understanding any proposition is true, and that it takes more time (and effort) to negate such an interpretation when parsing a sentence whose truth value must be reported with respect to a picture (e.g., "The plus is not above the star"). Just and Carpenter (1971) similarly argue that negative sentences are the "unpreferred form," and take longer than comparable positive ("affirmative") sentences in a similar verification task (e.g., "The dots aren't red" vs. "The dots are red."). So, we might expect that parsing negatives could just be more difficult, and thus show poorer performance. Alternatively, the participants might focus on the polarity of each claim when choosing strategies. Since positive proofs require the demonstration of some property's presence, use of an example or examples which show that property's presence may be thought of as good strategies. In contrast, negative proofs require the demonstration of the absence of some property. To show a property's absence, perhaps a more abstract representation would be chosen. Note that such a difference in strategies is not warranted–the strategy selection choice should be based on the quantifier in the claim to be proven[1].

If participants understand that the universal claims require the use of variables, while the existential claims require only a test case, their use of variables ought to be greater for the universals. However, if they have difficulty understanding what is required, they are likely to use test cases to confirm many of the claims in their universal proofs. This experiment will isolate where problems exist in students who have varied amounts of mathematical experience.

A lack of understanding of the claim's quantifier would lead to participants believing that they had successfully proven the universal claims using examples. In

[1]This assumes that the claim is in a form for which the quantifier has wide scope: "(∀x) not P(x)" rather than "not (∃x) P(x)"

contrast, an inability to apply variables properly would be demonstrated by participants who believe that they did not successfully prove universal claims with one or more examples. It is possible that they cannot understand the quantifier, and thus do not use variables for universal claims. Or, they may have trouble correctly applying variables in their proofs. As a further test of the ability of participants to apply variables properly, I asked them to redo with variables all the proofs for which they did not initially use variables.

2.1 Method

2.1.1 Participants

Participants were 20 Northwestern undergraduates participating in the experiment as part of the requirements for an introductory psychology course. These students varied in their prior experience with math, and had taken anywhere from one to six courses in which they needed to do proofs.

2.1.2 Materials

Participants were given 16 claims, four from each of the proof classes (positive universal, negative universal, positive existential, negative existential). The claims varied in content from those which addressed parity (oddness or evenness) to those which addressed what happens when certain operations are performed. To balance for difficulty of proof, each participant saw one of four variants of the claims, presented in random order. As seen in Appendix A, the variants allowed similar claims to be given in each of the four proof classes. Thus, a claim's basic topic could be worded in any of the four proof classes (*Positive existential*: Show that there's some product of an even number and an odd number which is even. *Negative existential*: Show that there's some sum of an even number and an odd number which is not even. *Positive universal*: Show that every product of an even number and an odd number is an even number. *Negative universal*: Show that every sum of an even number and an odd number is not an even number.). The claims were chosen so that most of the introductory psychology students would likely have the ability to prove them.

2.1.3 Procedure

The participants received the proofs, preceded and followed by instructions. The instructions preceding the proofs were as follows:

> Please go through the following mathematical demonstrations. Don't worry too much about being able to show what each question is looking for. We are more interested in seeing what methods you use to approach mathematical problems. Also, don't worry about having to finish the whole packet in the allotted time.

After writing the proofs, participants listed the math, logic, computer science or other courses they had taken in which they had been required to use some kind of mathematical proof technique. This was to gauge how their experience may have impacted their understanding of when variables are necessary. Also, and more importantly, after finishing the booklet, the participants were given a form on which they indicated how well on a four-point scale they believed they had proven each claim given. The following numeric scale was presented for each proof: 1=shows no support, 2=is somewhat relevant, 3=shows support for, 4=clearly proves. After completing that form, participants were asked to explain their overall strategy choices:

> In some of your demonstrations, you may have chosen to use variables, while in others, you may have chosen to use numeric examples. Why did you choose to use variables when you did and numeric examples when you did? Try to answer by telling what you were thinking and don't make up an answer if you cannot recall all of what you were thinking.

Finally, participants were asked to redo all the proofs for which they did not use variables in their proofs, and were asked to redo them using variables. This task took participants between 25 and 45 minutes, which was consistently less than the allotted time of one hour.

2.2 Results and Discussion

Proofs were scored by the experimenter and a second rater on the same four-point scale that was used by the participants (1=shows no support, 2=is somewhat relevant, 3=shows support for, 4=clearly proves). A rating of 4 indicated that the proof was done correctly. Note that for universals, this requires the use of variables or an adequate descriptive proof which is isomorphic to one using variables. A rating of 3 means that there was some support for the proof; this would be the application of one or several examples for a universal proof. The difference between a 1 and a 2 was the relevance of the representation chosen to the proof. For example, a proof about squaring in which the participant doubled a number instead of squaring it would receive a 1. Interrater reliability was 94% on the ratings, and all differences were resolved through discussion between the raters. The ratings were done on this four-point scale instead of a simple all or none scale to see whether proofs which missed the mark were commonly ones in which at least some support was given to the claim.

2.2.1 Ratings of Proof Quality

A 2 (Quantifier) x 2 (Polarity) Analysis of Variance (ANOVA) was performed on the judges' ratings (see Appendix A for the four proof types). Performance was better for the existential claims ($M = 3.61$) than the universal ones ($M = 3.01$); $F(1, 19) = 56.81$, $p < 0.001$. Performance was also better for positive claims ($M = 3.40$) than negative ones ($M = 3.23$); $F(1, 19) = 4.96$, $p = 0.04$. No significant interaction between quantifier and polarity was found. The better performance on the existential claims than the universals is not surprising, as a single example is enough to prove an existential claim, but a more abstract representation is needed for a universal one. The better performance on positives suggests that participants find it easier to show that some trait is present than to show that some trait is absent.

The number of classes using mathematical proofs taken by each participant was used as a covariate in an analysis of covariance, again using quantifier and polarity. The quantifier effect held up, $F(1, 19) = 13.41$, $p < 0.001$. The polarity effect disappeared however, and even though the effects involving the covariate did not reach significance, it seems that with greater experience, participants focused less on the polarity. However, the correlation between participants' overall performance on the proofs and the number of classes taken is very low $r(19) = 0.05$, $p > 0.05$. Note that number of classes taken may not be an accurate depiction of experience, since not all participants noted some of their earlier courses (e.g., some who listed calculus did not list geometry).

An ANOVA on all of the participants' confidence ratings showed small effects for both quantifier and polarity which only approached significance. Participants believed that their performance was marginally better for existential claims ($M = 3.51$) than universals ($M = 3.31$); $F(1, 19) = 2.98$, $p = 0.10$. For polarity, positive claims ($M = 3.46$) showed barely higher confidence ratings than negatives ($M = 3.36$); $F(1, 19) = 2.77$, $p = 0.11$.

2.2.2 Analysis of Strategies

The judges classified the proof strategies that participants used for each claim. The following categories were used: example (for a single numerical example), multiple examples (for two or more numerical examples), variable (for use of letters in an attempt to demonstrate the claim), and other. Instances of use of each proof type can be found in Table 2 (these instances come from different participants).

The two judges had a 95% agreement on strategy choices. Differences were resolved by discussion (often with the selection of the strategy chosen first among apparent multiple strategies). See Table 3 for a complete breakdown of the strategies used for each of the four proof classes. Variables are used more often in proofs of universal claims (41%) than proofs of existentials (28%); $F(1, 19) = 4.20$, $p = 0.05$. Use of a single example accounted for more of the proofs of existential claims (59%) than universals (18%); $F(1, 19) = 31.53$, $p < 0.01$. Use of multiple examples, which made up most of the remaining proofs, was more common for the universals (33%) than in existentials (7%); $F(1, 19) = 15.30$, $p < 0.01$. "Multiple examples" refers to the use of two or more examples to show support for the claim. Note that this strategy will successfully prove existential claims, but will not prove universals. For each of these three strategies, no main effect was found for polarity; nor was an interaction found between quantifier and polarity.

The fact that the most common strategy for universals was variables is ideal, but its frequency is disturbingly low; under half the proofs of the universal claims were attempted using variables. Multiple examples was the most common strategy for universals after variables, suggesting that participants believed that they needed to show that the claim was true for more than just a single example. Unfortunately, use of multiple examples does not demonstrate the universality of any property.

Having asked the participants to go back and use variables to prove any claims for which they had alternative strategies, I was able to gauge whether they were capable of using variables to prove a claim. Using the least stringent criterion, I judged whether they could use variables. This criterion was that a participant was said to be capable of using variables if s/he proved at least one claim using variables, and received an experimenter rating of four (clearly proves). Two participants out of the 20 were found to be incapable of successfully proving any of the 16 claims given.

The following data reflect the same dependent measures listed above, with these two participants removed.

The performance differences held up when studying the competent participants' initial proofs. Judges rated the proofs of existential claims (M = 3.69) higher than those for universals (M = 3.07); $F(1, 19) = 52.91$, $p < 0.001$. Also, performance on positives (M = 3.47) exceeded that for negatives (M = 3.29); $F(1, 19) = 5.07$, $p = 0.04$.

Competent participants' confidence ratings of their proofs were in the same direction as the experimenter ratings of proof quality, but had effects which were only marginal. Confidence was marginally greater for existential proofs (M = 3.52) than universals (M = 3.29); $F(1, 17) = 3.091$, $p = 0.097$. Confidence was also greater for positives (M = 3.46) than negatives (M = 3.35); $F(1, 17) = 4.30$, $p = 0.05$.

Again, as in the full data set, participants used variables more frequently for the universal claims (39%) than for the existentials (25%); $F(1, 17) = 5.50$, $p = 0.03$. No main effect was found for polarity.

In terms of strategy choices, we find that participants' errors can be explained largely by which strategy they chose for the particular proofs given. The most common strategy erroneously used for each the two universal classes of claims was multiple examples. It was used for 45% of the incorrect proofs made in the positive universal class, and 40% of those in the negative universal class. Evidently, some participants believed that it was sufficient to show that some statement was true (or false) for several examples to assert its truth value for all examples. This demonstrates a lack of understanding of the universal nature of these claims. Confidence in use of multiple examples overall (both for correct and incorrect proofs) was higher in direction than that for variables (3.43 vs. 3.37 respectively). The difference between these two strategies was tested using items (only seven of the participants used both strategies, while 15 out of 16 items were done with both strategies), and we find that confidence in multiple examples (3.51) was greater than that for variables (3.16); $t(14) = 2.20$, $p = 0.05$[2]. We find the same result when looking specifically at the difference between strategies for only the universal problems. Again, confidence for multiple examples (3.51) exceeds that for variables (3.13); $t(14) = 2.20$, $p = 0.05$. Note that participants do have some idea of how good their proofs are: their confidence in correct proofs (3.64) is greater than that for incorrect proofs (3.13), $t(17) = 3.57$, $p = 0.002$.

Surprisingly, the most common strategy leading to unsuccessful proofs for both of the existential classes was use of variables (75% of errors for the positive existentials class, 67% for the negative existentials class). Usually, this involved an attempt to set up the problem using variables without following through and manipulating the resulting representation. For instance, to show that multiplying a number by four and adding three gave an odd number, a participant wrote "4x+3" and stopped. The use of variables was the second most common error strategy for negative universals (30%), and third for the positive universals (22%). Again, many of these attempts to use variables were mere representations of the stated claim using variables without any operations being performed on the resulting representations. One could argue that these restatements of the claims were not even instances of variable use, but

[2] The missing datum would strengthen the effect, as the item's mean confidence rating for variables was below the mean at 2.2.

placement of a letter into the representation. These failures with variables suggest that the provers may not have difficulty using variables to represent an aspect of the given claim (in our example, the operations performed), but run into trouble when they need to get that representation to match the rest of the claim (how can they show that the result is odd?). However, ten out of the 14 participants who used variables at least once had an average rating for their use of variables of 3 or higher, so the representations using variables were often relatively good.

The second most frequently used strategy among the errors for positive universal claims was the use of a single example (23%); this strategy was third for negative universals (22%). One cannot adequately show that some proposition is true (false) for all cases by citing a single case for which that proposition is true (false). Yet, this was commonly the approach used for universal claims, suggesting a lack of understanding of the universal quantifier.

2.2.3 Justifications for strategies

Recall that participants were asked about their overall strategy choices at the end of the experiment (when they chose the use of a single example, multiple examples, or variables). In their responses, there were four answers which were given at least twice. The most common answer, given six times (30%) was that claims asking for existence need only a single example, whereas claims using the words "all" or "every" require variables. A seventh person noted the difference in quantifier, but said that proving something false would be done with a number (numeric example). The next most common description of a strategy was the statement that using numbers makes the proof easy; 25% (five people) used numbers just for simplicity's sake. Four people (20%) said that the type of proof used varies according to the particular features of the claim, and named some such features, which were not related to the quantifier in any way. Two of the participants tried to use variables for any claim for which they were capable, and used numbers otherwise.

These results suggest that some of the provers understand the quantifier's import, some have difficulties both with understanding that a universal claim requires an abstract proof, and others have trouble setting up or manipulating a representation for a claim using variables. It seems that the trouble is not with manipulating the representation with variables per se as much as it is with knowing when a representation allows the conclusion. Just writing, "4x+3" doesn't show that multiplying a number by 4 and adding 3 yields an odd number. If someone doesn't write, "4x+3 = 2(2x) + 2(2) + 1 = 2(2x+1)+1 which is odd because it's one more than a multiple of 2," it may very well be that s/he could do the manipulation, but did not think of that final representation as being a useful one.

Perhaps if the task can be simplified so that participants would not have two stumbling blocks (those of choosing a representation and then manipulating it), they would have more resources to recognize what constitutes an adequate proof. Both of these potential impediments are removed in Experiment 2.

Chapter 3

Experiment 2 – The weight of cognitive load

Simply giving participants claims to prove may be too much of a burden, especially if they do not have much experience with proofs. Participants may be able to recognize the differences between the proof classes, but not have ample resources to focus on those differences when generating proofs. The differences might be more apparent if participants are asked to judge prewritten proofs. Thus, in this experiment, unlike the first, the participants did not have to generate the steps themselves, so the cognitive load of this experiment was lower than that in Experiment 1. Without the need to generate each proof, the participants are better able to show their understanding of what is required for each proof type.

Participants were given a series of claims, each one with an attempted proof. The attempted proofs belonged to one of three strategies: example, multiple examples, and variables. Note that an attempted proof which uses a strategy of either a single example or multiple examples would completely prove an existential claim if the proof contains one or more supporting cases, but would be insufficient for universal claims. In contrast, if the attempted proof is a supportive variable proof, it would be appropriate for all existential or universal claims. Since variables would work for all types of claims, but the other two strategies do not, there is an unbalanced number of times for which applying the different strategies would provide successful proofs. To remedy this, I took two steps. First, new, unsupportive proofs were created for all the crossings of strategy, quantifier, and polarity. As an example, for the claim, "Show that there is an odd number less than 25 which is a perfect square (a number whose square root is an integer)," the following was an unsupportive proof:

> 11 is less than 25.
>
> $11^2 = 121$.
>
> So, 11 is a perfect square.

Creating unsupportive proofs for all three strategies still would leave more successful variable proofs than example or multiple example proofs. To balance the proportion of times that attempts using each strategy were sucessful (at 50%), I provided three times as many supportive example and supportive multiple examples proofs for the existential claims as the proof attempts in other conditions. Participants rated how well each claim supported the proof on the same type of four point scale as used in Experiment 1. The problem with using the same exact scale was the top rating. In Experiment 1, the best proofs were rated as 4 for "Clearly proves."

Here, since the participants did not write the proofs themselves, clarity of proof could mean either how well the proof supports the claim or how clearly written and easy to follow the proof itself is. To disambiguate the two, the rating of 4 was changed to "Completely proves," and participants were asked to make additional ratings of clarity of writing for each proof. Note that ratings of 3 and 4 are ratings suggesting that the proof is somewhat supportive, in that 3 is "Shows support for." In contrast, ratings of 2, "Is somewhat relevant," and 1, "Shows no support for" are non-supportive. Talk-aloud protocols were collected. All participants were asked to state aloud all of their thoughts on why they are making each rating. In addition, as in Experiment 1, after completing all the proofs, I asked participants for a statement about when they believed that applying variables was the preferred method as opposed to giving a demonstration for one or more numeric examples. In this experiment, with the reduction in cognitive load, we expect that people will be able to determine easily which proofs are stronger or weaker ones. In addition, since they will not need to generate the proofs themselves, participants ought to be better able to focus on the quantifier of a claim as a crucial feature in determining the proof strategy applied.

3.1 Method

3.1.1 Participants

Participants were 18 Northwestern undergraduates participating in the experiment as part of the requirements for an introductory psychology course.

3.1.2 Materials

Participants received two booklets, each with a set of 16 claims; order of presentation was counterbalanced. Each claim was followed by an "attempted proof" (written by the experimenter) using one of the following strategies: (a) an example, (b) multiple examples, or (c) a proof using variables. Instances of each of these proof strategies for "Show that for all whole numbers, doubling a number will not yield an odd number, are respectively (a) "2(4) = 8 is even," (b) "2(3) = 6, which is even. 2(4) = 8, which is even," and (c) "Let k be a whole number. 2k is even since 2 is a factor of 2k."

I presented an equal proportion of correct proofs for each strategy type, which meant presenting three times (i.e., three) as many supportive single example and multiple example proofs as in the other cells (i.e., one) in the design (unsupportive single example, unsupportive multiple examples, supportive variables, and unsupportive variables). For a full set of claims and "attempted proofs," see Appendix B. The order of presentation of the two sets of claims was counterbalanced.

3.1.3 Procedure

The participants were given a set of claims, each one with an attempted proof, and asked to rate how well each proof supports the claim. Ratings were on a four-point scale from "Shows no support" to "Completely proves." Participants also rated how clearly written each proof is on a four point scale from "Impossible to understand," to "Simple to understand." Students' responses indicated their beliefs about when each of the strategies are useful, and when they are successful. These are the complete instructions:

> In this experiment, you will be presented with two booklets. One will contain a set of mathematical claims; the other will be a ratings booklet. In the claim booklet, each claim will be followed by an attempted proof. Your task will be to read each claim aloud. Then, you will read the proof aloud. Please judge how well the proof supports the claim.
>
> You will make a rating for each of the proofs on a four point scale, ranging from "Shows no support" to "Completely proves." We ask that you both circle the number in the ratings booklet for the appropriate claim, as well as state your rating aloud. Please state why you are giving the rating, and why you think the method used may or may not be preferred for each one of the claims. The ratings booklet also asks you to express your opinion about the clarity of the proofs. You will make a rating for each on a four point scale, ranging from "Impossible to understand" to "Simple to understand." Again, please circle the number in the booklet, and state aloud what your rating is, and why you have made that rating.
>
> We are interested in what you think about when you make your decisions. In order to find out what you are thinking, I am going to ask you to THINK ALOUD as you work on the problems given. What I mean by think aloud is that I want you to tell me EVERYTHING you are thinking from the time you start each problem until you have given your final comment on that problem. I would like you to talk aloud CONSTANTLY from the time you read the problem until you have finished describing

> why you have made the ratings. I don't want you to try to plan out what you say or try to explain to me what you are saying. Just act as if you are alone in the room speaking to yourself. It is most important that you keep talking. If you are silent for any long period of time, I will ask you to talk. Your comments will be audiotaped, and I may prompt you for more information about the proofs as you go along.

Participants were prompted to tell what they would have done differently in order to completely prove the claim whenever they did not give a proof a rating of 4 (completely proves). Also, as in Experiment 1, after going through all of the proofs, participants were asked to tell, in broad terms, when they decided variables were preferred and when they decided that numeric examples were.

3.2 Results and Discussion

3.2.1 Participants' Ratings

A 2x2x2x3 within subjects (repeated measures) ANOVA was performed on participants' ratings of the proofs. The independent variables were support (supportive vs. nonsupportive proofs), quantifier, polarity, and strategy (example, multiple examples, and variables).

Participants rated existential claims' proofs better (3.60) than universal claims' proofs (1.78), $F(1, 17) = 387.70$, $p < 0.01$. This was expected, as a greater proportion of existential proofs were good proofs (in order to adequately counterbalance the proportion of times that each of the three strategies led to successful proofs).

Their ratings were higher for positives (2.89) than negatives (2.49), suggesting that participants were more readily convinced of the presence of a property than its absence, $F(1, 17) = 47.84$, $p < 0.01$.

Participants successfully recognize that the non-supportive proofs are worse than the supportive ones, but they rate the supportive ones higher (2.80) than the non-supportive ones (2.58), $F(1, 17) = 11.81$, $p < 0.01$.

There was an interaction between quantifier and strategy such that while proofs using variables were judged somewhat better (a difference of 0.16) than single or multiple examples for existentials, multiple examples were thought to be better than using a single example or variables for universals (a difference of at least 0.3), $F(2, 34) = 3.85$, $p = 0.03$. Such an interaction shows that participants do not know that variables are required for proofs of universal claims.

Here is an example of a participant's transcript in which he accepts a proof using multiple examples, "Um..Show that there is an odd multiple of 3 uh which when squared is divisible by 5. 15 is an odd multiple of 3 since it's 3x5. 15 squared = 225. 225 is divisible by 5. 45 is an odd multiple of 3 since since it's 3x15. 45 squared = 2025. 2025 is divisble by 5. So there's an odd multiple of 3 which when squared is divisible by 5. Um...15 is an odd multiple of 3. 15 squared is 225, and it's divisible by 5, so that shows um confirms that proves that. Um...Another example 45 is an odd multiple of 3, 3x15. 45 squared is 225, I mean 2025. 2025 is divisible by 5 so there that proves it clearly." Similarly, many other participants were convinced fully by proofs using multiple examples.

Ratings also showed a curious interaction between polarity and strategy such that variables were preferred over single or multiple examples for positive claims, but multiple examples were preferred for negative ones, $F(2, 34) = 20.01$, $p < 0.01$.

Support and strategy also showed an interaction such that multiple examples and examples were preferred for the nonsupportive proofs, variables were preferred for the supportive ones $F(2, 34) = 18.63$, $p < 0.01$. This is surprising since there are equal numbers of supportive proofs from each of the three strategies. Perhaps when the proof was non-supportive, participants were more likely to endorse a simpler representation than when the proof gave support for the claim.

The three-way interaction between quantifier, polarity and support shows that while the larger rating decline between positive and negative proofs occurs for the existentials (3.91 to 3.46) than universals (1.50 to 1.44) in the nonsupportive proofs while the rating decline between positives and negatives is greater for universals (2.59 to 1.59) than existentials (3.56 to 3.46) in supportive proofs, $F(1, 17) = 13.77$, $p < 0.01$. This reflects a difference in focus. When the proofs are nonsupportive, the largest differences in ratings can be explained by quantifier. However, when the proofs are supportive, and interpreting them becomes easier, focus is shifted to the polarity as well as the quantifier (the dropoff between positive and negative universal is about the same as that between the positive existential and positive universal).

A similar explanation can be given to the three way interaction between quantifier, support and strategy, $F(2, 34) = 6.43$, $p < 0.01$. The pattern described above for the support x strategy interaction holds for nonsupportive proofs, but for supportive ones, use of a single example is the preferred strategy for existentials while multiple examples are preferred for universals.

Finally, the four-way interaction was significant as the three-way effect between quantifier, support and strategy can be broken down differently for positive and negative polarities, $F(2, 34) = 10.53$, $p < 0.01$. This can mainly be explained by the supportive universal proofs, for which variables are the preferred strategy for the positives.

3.2.2 Participant Strategies

Participants differ in their beliefs about when the best strategy was variables, numeric examples or a single numeric example. Their answers about the overall approach used were categorized as follows: Ten people (56%) used quantifiers (understood that universal claims require variables, or some similar abstract representation, while the existential claims require minimally one example), six (33%) preferred numbers (felt that using numeric examples [either single or multiple] are always easier, and thus preferred that strategy to variables), while two (11%) had a relatively diverse picture of what strategies were appropriate, apparently based upon features of the individual problems.

Beyond the answer to the single question at the end of the experiment, I looked at the strategies applied *during* the course of the experiment based on the think-aloud protocols. The participants can be studied on a coarse or fine level of detail. On the coarse level, each person's overall understanding of which strategy should be applied during the experiment was scored. The overall approach used by each participant ought to agree closely with their self-reported approaches described above. Each participant was classified into one of six global approaches as follows. Upon seeing the first universal claim, some of the participants chose the strategy needed for that proof specifically because of the universal nature of the claim (e.g., "since it claims that...talks about every even number again, some kind of variable would have to be used in the equation."). These people can be divided into two categories, those who said you need to use variables in the proof, and those who said you need multiple examples in the proof. Similarly, some participants did make the realization that the quantifier was important for strategy choice, but not until later on in the experiment. Again, they can be divided into those who believe variables are necessary for universals and those who believe that multiple examples are necessary. The remaining two categories were for the participants who did not acknowledge the quantifier throughout their ratings. One group was comprised of those who simply said that using numbers was easier, and therefore an example or multiple examples should be used to prove all of the claims. The final category for an overall approach was that of participants who believed that the particular features of each claim were necessary to determine what strategy should be applied to the proof of that claim. In order to ensure reliability, a second rater joined the experimenter for this categorization. The two raters had a 78% agreement rate (14/18), and resolved disputes through discussion. Three of the disagreements involved whether or not participants mentioning more than one example for a universal meant that they believed more than one was necessary. Since these three participants also said that multiple examples helped to prove existential claims, they were considered not to have acknowledged the quantifier. Only 11% of the participants consistently followed the strategy positing that numbers were easier than variables, and 6% believed that strategy choices were based upon features of the individual problems irrelevant to quantifier. The remaining 83% acknowledged quantifier at least once in their description of the best way to do the proofs for each claim. When reviewing only the first of the "attempted proofs" for a universal claim, 22% asserted that variables were necessary for universal claims, and 17% asserted that multiple examples were necessary. While 11% eventually mentioned that variables were necessary, 17% eventually mentioned that multiple examples were necessary.

The participants' ability to correctly rate the attempted proofs (measured by the difference between the experimenter ratings for the proofs and the participant ratings) was not found to vary greatly with overall strategy. Participants who said they used quantifier in some (not necessarily correct) way in order to determine how well the proofs supported the claims rated the claims on average 0.07 higher than the experimenter while those who did not use quantifier rated them 0.11 higher than the experimenter, $t(16) = 0.41$, $p > 0.05$.

When comparing the overall approach as scored by the experimenter and second rater to that given in the self-report, we see that one of the people who described (in the self-report) a context-dependent approach which changes based upon the problem actually spoke (during the task) about numbers being simpler. Also, five of the people who reported using numbers because they were easier did at some point in the experiment cite the quantifier as the basis for determining which strategy of proof to apply. So, we see that the self-reports often missed the use of quantifier when the very same participants sometimes focus on it. This gives indirect evidence that participants, while sometimes aware of quantifier, give insufficient import to quantifier in determining their overall approach to selecting strategies to prove claims.

At a more fine level of study, each participant's pattern of responding can be examined. Recall that participants were asked to report what strategy they would have applied for each claim if the proof given was not sufficient. So, whenever a proof was rated as completely proving the claim, the participant had accepted the strategy used. If the proof was rated as anything else, the participant was required to state how s/he would do the proof. Thus, for each claim, the participant endorsed some strategy. For existential claims, if the participant endorsed using a supportive example, then s/he has the right idea. If the participant endorsed another strategy for an existential claim, s/he may not understand that only one supportive example is necessary. For universal claims, if the participant endorsed a supportive variable claim, then s/he has the right idea. If s/he endorsed a supportive example or supportive multiple example proof, then s/he does not understand the need for abstraction in proofs of universal claims. Table 4 shows proportions of how well each participant understood which strategy was optimal, giving proportions for clearly understood, may not have understood and did not understand. As can be seen from the table, only four out of the 18 participants (22%) used variables for universal claims more than 50% of the time.

Thus, even with the reduction in cognitive load, it seems that while most participants were aware that more than a single example is necessary for the universal quantifier, some were not always aware of the quantifier. As in Experiment 1, participants often do not know that multiple examples are insufficient to prove a universal claim.

Chapter 4

Experiment 3 – Sorting claims into categories

So far, the first two experiments have started with the assumption that the proof classes are noticeable by the participants. Looking at participants' answers for when they used a single example, multiple examples, or variables, we find some evidence that they recognize the difference between the universal and existential quantifiers. Just because experienced mathematicians and logicians are used to distinguishing between claims based upon quantifier, it is not safe to make the assumption that less experienced students are capable of viewing them the same way without some reminder about the differences between claims. This experiment strives to ascertain whether the proof classes as described are a reasonable way in which students would categorize the claims.

In this experiment, there are two tasks, a self-sort and a guided proof class sort. In the self-sort condition, participants were asked to sort the claims based on the strategies they would use to prove those claims. The sorts were performed on all claims before participants worked on the requested demonstrations. If provers are capable of sorting them into structural categories which are equivalent to the four described classes, then their strategy choices for the proofs they write should map neatly onto those classes. If not, then the strategy choices may be messier.

In the proof class condition, participants were told explicitly about the four proof classes, and then asked to carry out the requested demonstrations. If they can distinguish between the proof classes, then they might have distinct strategies for working with each category of claim to be proven.

The results of these two groups can be compared. If the structural categories made by the first group can be matched (at least roughly) with those given to the second group, we would expect that the strategies applied for universals would be applied in the same way by both groups, as would those for existentials. Here, as in the first two experiments, participants were asked why they chose to use numeric examples for some proofs and variables for others.

4.1 Method

4.1.1 Participants

Participants were 22 Northwestern undergraduates participating in the experiment as part of the requirements for an introductory psychology course.

4.1.2 Materials

The claims given were 24 of the same types of claims as in Experiment 1 (see Appendix C). As in the first experiment, claims vary by Quantifier (existential vs. universal) and Polarity (positive vs. negative). Each claim was presented on a quarter of an 8.5 x 11" sheet of paper, so that they could be easily sorted.

4.1.3 Procedure

All participants were told that they were to sort claims and then prove them. Half of the participants (self-sort) were told to sort the claims into structural categories based on how they believed the claims would be proven:

> There are several different classes of mathematical claims which can be defined by the strategies that would be used to prove those claims. Look at these claims, and consider what type of strategy you would use to prove each of them. Your task will be to sort all of the claims into classes based on which strategy you would use to prove them. Label each pile with an appropriate strategy name on the top right of the first sheet.

The other half (proof class) read the following passage, describing the four proof classes:

> There are four main classes of mathematical claims which can be described as follows. Two of the classes are positive, in that the claims listed are to be proven. The other two are called negative, because the claims listed

> are to be disproven. Each of those two categories (positive and negative) may further be broken down into two classes. In one of those, something is said to be true or false for all possible cases of something, whereas in the other class, something is said to be true or false for at least a single case.
>
> Your task will be to sort all of the claims into these four classes. Place each class into a separate pile. Label each pile on the top right of the first sheet. After you have completed the task, please write down on this piece of paper any difficulties you have encountered while sorting the proofs into these classes. When you have the classes separated into piles and have written down any difficulties you feel you had, please ask for the stapler, staple the piles, and get the next set of instructions.

The participants sorted the claims into their categories, and then stapled each category into a mini-packet. Next, both groups were asked to complete the demonstrations to the best of their ability. After finishing the set of demonstrations, each participant was given the same form as in Experiment 1, asking how well they believed each demonstration proved the claim given, as well as the question about when they chose to use variables in their demonstrations and when they chose to use numeric examples.

4.2 Results and Discussion

Both cluster analysis and multidimensional scaling (MDS) were used for each of the two sort conditions in order to determine the how the claims were sorted across subjects. The data from each participant's sort took the form of a symmetric matrix of proximities, in which 0 stood for a pair of claims which were in the same category and 1 stood for a pair of claims which were in different categories. Within each condition, the individual subject matrices were averaged.

4.2.1 Hierarchical cluster analysis

A hierarchical cluster analysis was performed, using average linking (see Figures 1a and 1b). For the proof class condition, the top level split yields two main clusters–one has only universal claims, the other only existential ones. The universal class breaks up into two clusters, one of which has all negatives, the other of which has all the positives and one negative–the one negative uses the phrase "is not odd," which could be thought of as "is even," and thus be construed as positive. The existential class similarly breaks up into two clusters, one of which has all negatives, while the other has all the positives and two negatives.

For the self-sort condition, the top level split is clearly based on quantifier, splitting the universal claims into one cluster and the existential ones into another. Within those clusters, the further splits do not usually fall along the lines of polarity. The labels given to the categories generated by self-sort participants do not correspond to the quantifier (see below in Category Names).

4.2.2 Multidimensional Scaling

The same matrices were used as in the cluster analysis. Measurements were taken to be ordinal, and scaling was based on Euclidean distance in SPSS's ALSCAL procedure. R^2 was used to measure the goodness of fit. Kruskal and Wish (1978) recommend that the maximum number of permissible dimensions not exceed one quarter of the number of items, and that each added dimension should account for at least a 5% increase in variance accounted for by the model. Following Kruskal and Wish's (1978) rule for determining how many dimensions are appropriate for the proof class sorts, the three-dimensional solution, which accounts for 48% of the variance, was selected.

For the three-dimensional solution, two separate multiple regressions were run, one using quantifier and one using polarity as dependent variables, with values along the three dimensions for each data point as independent variables. For quantifier, only the first dimension was predictive, with a regression coefficient of -0.23, $t(20) = 2.56$, $p = 0.02$; it correlated -0.46 with quantifier. For polarity, again only the first dimension was predictive, with a regression coefficient of 0.19, $t(20) = 2.10$, $p = 0.05$; it correlated 0.41 with polarity. So, the first dimension seems to run along a continuum from positive to negative, at the same time moving from universals to existentials. The other dimensions do not offer any apparent interpretation. Evidently, the participants did attempt to follow the instructions given about sorting the proofs into classes based upon quantifer and polarity, but they could not clearly view the two as orthogonal.

For the self-sort data, according to Kruskal and Wish's (1978) rule, the four-dimensional solution, which accounts for 52% of the variance, is appropriate (one-, two-, and three-dimensional solutions account for 28%, 35%, and 45% respectively).

As in the Proof Class sort, the first dimension predicts quantifier, with a coefficient of .28, $t(19) = 3.72$, $p < 0.01$; their correlation was 0.63. However, for polarity, only the fourth dimension's coefficient (.20) approaches significance, $t(19) = 1.74$, $p = 0.10$, with a correlation of 0.36. The other dimensions have no apparent interpretation here. Clearly, quantifier is an important dimension along which the participants divided the claims in both conditions. To a lesser extent, polarity appears to play a role in how some of the participants categorize claims. The categories used are described in greater depth as part of the discussion of strategy usage that follows.

Using individual differences scaling, it is not readily apparent how to describe the different sorting strategies of the Proof Class and Self-Sort participants, although the dimension of quantifier very clearly is important to both groups.

4.2.3 Strategy Usage

Interrater agreement on the strategies used by participants was 89%. Disputes were resolved through discussion. A 2x2x2 (quantifier x polarity x sorting condition) ANOVA was performed for the proportion of use for each of the three most common strategies–example, variable, and multiple examples.

Example use was greater for existentials (62%) than for universals (26%), $F(1, 20) = 29.93$, $p < 0.01$. It was also used more often for negatives (50%) than for positives (37%), $F(1, 20) = 11.70$, $p = 0.03$. No interaction between quantifier and polarity was found. No effects involving sorting condition reached significance.

For variable use, there were no significant main effects, but there was a marginal interaction between quantifier and polarity. While variable use was not much lower for positive than negative existentials (a 2% difference), it was higher for positive than negative universals (a 5% difference), $F(1, 20) = 3.57$, $p = 0.07$.

The use of multiple examples was greater for universals (39%) than for existentials (9%), $F(1, 20) = 28.92$, $p < 0.01$. Multiple examples also showed an interaction between quantifier and polarity such that while there was not much of a difference in usage between existential positives and negatives (8%, 9% respectively), there was quite a difference between universal positives (44%) and negatives (33%), $F(1, 20) = 6.00$, $p = 0.02$. Again, no effects involving sort condition were found.

The effects having to do with quantifier match those earlier in this paper, but the polarity effects and interactions are novel effects. The interactions seen here suggest that participants may have difficulty with the scope of the negative. By scope, I am referring to the difference between "there is a case for which something is not true" and "something is not true for all cases." For universal negatives, it may be difficult for participants to disambiguate the two. As a concrete example, "show that every sum of an even number and an odd number is not an even number," could be interpreted as "show that there's some even number and some odd number whose sum is not an even number" or as "show that whenever you add an even number and an odd number, you'll always get an odd number." If participants mistook universal negatives for existential negatives, that may explain the greater use of variables and multiple examples for universal positives relative to universal negatives.

As in the first experiment, confidence was higher in direction for multiple examples (3.4) than for variables (3.2), but the only significant difference in confidence between a pair of the three most common strategies was that between multiple examples (3.4) and a single example (3.1), $t(18) = 4.34$, $p < 0.01$.

4.2.4 Category Names

The category names generated by the participants in the self-sort conditions varied greatly for the same types of categories. To simplify the taxonomy, only one name was selected for each type of category. Table 5 lists all of these categories (those used by two or more participants), along with the average number of problems for which each category accounts. Many of the category names were names of operations (e.g., "multiplication," "multiplication and subtraction," "squaring"). In order to simplify the taxonomy, all such groupings have been labeled "operation(s)." Operation(s) accounted for 21% of all problems, and was used by five people as a category label. Five people also used some type of category which represented "multiple examples," and five created a category of claims based on "facts about parity." Four people used "single example" as a category, and four people used "steps," in which multiple operations would be performed on a number. "Steps" differs from operation(s) in that the focus of the category name was on the fact that there were steps or a series of operations performed rather than just the name or names of operations which were applied. "Trial and error" was used as a category by three people, and "variables" was used as a category by only two. In sum, the categories generated by the participants focused largely on the operations and combinations of operations which would be performed on a number, numbers, or variable that would be used in the proof of the claim rather than on whether a single number, multiple ones, or variables would be used for representing the claim. The representation chosen largely determined the performance; since participants used either single or multiple examples for many universal claims, they had no chance to give an adequate proof in those cases.

4.2.5 Performance

Interrater agreement on the ratings of proof quality was 80%. Again, any disagreements were resolved through discussion. Proof quality was higher for the existential proofs (M = 3.22) than universal proofs (M = 2.41), $F(1, 20) = 38.25$, $p < 0.01$. No significant differences were found for polarity or sorting condition. Nor were any of the interactions using proof quality as a dependent measure found to be significant. Using the number of mathematically oriented classes as a covariate, and repeating the analysis, the difference between existential and universal proofs holds up, $F(1, 19) = 5.72$, $p = 0.03$.

A polynomial trend analysis on the participants' ratings of their proof quality using our rating of quality of proof as an independent variable showed that partipants did show linear concordance with the raters on proof quality, $F(3, 33) = 4.46$, $p = 0.01$. No quadratic or cubic effects were found. So, it seems that the participants do have some idea of which of their proofs are better, and which are worse.

4.2.6 Overall strategy used

Participants' answers about when they used a single numeric example, multiple ones, or variables shed more light on their awareness of the importance of quantifier.

The difference between problems asking to demonstrate something for all cases and those asking for an example for which something was to be demonstrated was acknowledged by 54% of the participants, two thirds of whom mentioned that variables were needed for the universal proofs. Seven of the 12 people who mentioned quantifier were in the proof class condition; the other five were in the self-sort condition. Hence, it does not seem that the different sorting instructions clearly changed the provers' answers regarding the importance of a claim's quantifier in strategy choice. The only other described approach which appeared more than twice (out of the 22 participants) was that numbers are just easy to use for proofs, and thus were applied. People who gave this answer (23% of the sample) did not acknowledge that the quantifier mattered.

4.2.7 Summary

The sorting tasks here in Experiment 3 illustrate that distinguishing claims based upon quantifier is a feasible task for participants. Not only can they sort claims based on quantifier when explicitly instructed to do so, but they use quantifier in sorting when told to sort claims based on the strategies needed for proving them. Strategy choices also reflect knowledge of the quantifier: variables are used more for the universals and application of a single example is used more for existentials. Also, as seen earlier, multiple examples were used more often for universals than for existentials, and are incorrectly thought of as adequate proof for the universals.

Acknowledgment that the quantifier of a claim matters in determining which representation to choose for its proof, along with the ability to choose and manipulate the appropriate representation have been the two foci of the experiments so far. We have seen in the first three experiments that participants have a problem recognizing the need for an abstract representation in proofs of universal claims or with setting up such a representation appropriately. In the fourth experiment, the focus is on giving participants hints to make them aware of certain features of the claims or the method for representing claims using variables.

Chapter 5

Experiment 4 – Here's a clue

In order to improve performance, students must be made aware of the various distinctions in claim types discussed so far. In order to make them aware of these distinctions, three small blurbs were used as hints for focusing on the distinctions between problems. In the earlier experiments, most of the participants did understand the difference between universal and existential claims, but there were still some who missed the distinction (e.g., 17% of the people in Experiment 2 never cited quantifier in their answers about the appropriateness of a proof strategy). Thus, one hint focused on "quantifier," and emphasized that universal claims need to be proven for all possible cases.

Another hint type showed people how to represent the claim in terms of variables (for example that 2x could be a generic even number, while 2y+1 could be an odd number). This would be helpful for people who have trouble understanding how to represent a claim in terms of variables, and possibly push those who feel that using numbers is just easier into applying variables more often when their use is necessary.

The third type focused on the sense of scope, namely emphasizing the difference between negative universal and negative existential claims, clarifying the scope of the negative applied in the claim. This scope problem is discussed among the results of Experiment 3, where both the multiple examples and variables strategies are used less for universal negatives than universal positives.

After reading the focus hint, participants were given more proofs, the performance on which indicates the degree to which focusing attention on such features of the claim could repair the difficulties in participants' proof techniques. These hints ought to help participants who have difficulties making the distinctions described.

5.1 Method

5.1.1 Participants

Participants were 28 undergraduates taking Introduction to Psychology at Northwestern University, participating in the experiment in partial fulfillment of a class requirement.

5.1.2 Materials

The 24 claims to be proven included six from each of the four proof classes, presented in random order (See Appendix D for these claims).

5.1.3 Procedure

Participants were given packets with 12 claims (early claims) to be proven, and were told to work on demonstrating the claims to the best of their ability. As in earlier experiments, the participants were asked how well they believed each demonstration proved the claim. After collecting those packets, the experimenter gave each participant one of the three hints, along with a packet of another 12 claims (late claims) to be proven. After proving the second set of 12 claims, participants rated how well they had proven the claims. Also, as in the earlier experiments, they listed the classes they had taken and told when they have chosen to use variables, multiple examples, or a single example.

In the quantifier condition, participants read a description of proof classes similar to what the proof class group in Experiment 3 read having to do with the quantifiers:

> Proof claims may be broken down into two classes. In one of those, something is claimed to be true or false for all possible cases of something, whereas in the other class, something is claimed to be true or false for at least a single case. So, for the former, one needs to show that the statement is true or false for every possible case, while the latter only requires finding a single relevant case.

Participants in the variable representation condition were given the following passage:

> In proving mathematical claims, it is important to know how to properly represent the problems in terms of variables. For example, 2x could be a generic even number, while 2y+1 could be an odd number. Importantly,

2x+1 would not be an appropriate odd number if 2x were already chosen to be an even number (2x+1 is no longer "any" odd number–it happens to be the particular one which is equal to 1 + the even number discussed in this context).

In the scope condition, participants read the following:

Two types of proof claims are often confused. One is of the form, "Show that something is not true for all possible cases," and the other is of the form, "Show that there is a case for which something is not true." The former has the negative applied to all cases, whereas the latter only applies the negative to a single case.

5.2 Results and Discussion

The experimenter and second rater had an 83% agreement on the strategies used by the participants, and an 81% agreement on their ratings of proof quality. Any disagreements were resolved through discussion.

Means for performance (judges ratings) can be found in Table 6. A 3x2x2x2 (Hint x Block [before vs. after hint] x Quantifier x Polarity) ANOVA was carried out on these scores. Performance was better overall for people in the Quantifier ($M = 3.35$) and Scope ($M = 3.39$) conditions than for the Variable ($M = 3.19$) condition, $F(2, 25) = 3.59$, $p = 0.04$. The hints had a marginal negative impact on the participants, who dropped from a mean score of 3.35 before receiving a hint to 3.28 afterwards, $F(1, 25) = 3.43$, $p = 0.08$. This is not likely the result of fatigue, as a linear trend of performance dropoff with time over the 24 proofs was not found to be significant, $F(1, 25) < 1$.

As in the earlier experiments, performance was better on the existential proofs ($M = 3.80$) than the universal ones ($M = 2.82$), $F(1, 25) = 151.58$, $p < 0.01$. However, surprisingly, performance was better on the negative proofs ($M = 3.36$) than the positive ones ($M = 3.27$), $F(1, 25) = 5.97$, $p = 0.02$. The absolute difference for polarity is quite small for a four-point scale, though (only .09).

While there was a small increase in performance between the early and late existential proofs (0.08), there was a dropoff in performance between the early and late universal proofs (0.21), $F(1, 25) = 5.07$, $p = 0.03$. There was also a small increase in performance between early and late positive proofs (0.04), but a dropoff between early and late negative proofs (0.19), $F(1, 25) = 8.00$, $p = 0.01$. Figure 2 illustrates the three way interaction between block, quantifier, and polarity showing that while performance on the positive claims does not show a marked change between early and late tests, the universal negatives show a large dropoff (from 3.12 to 2.68) while the existential negatives do not change by much, $F(1, 25) = 5.62$, $p = 0.03$.

Performing an analysis of covariance using number of classes as a covariate, the type of hint effect holds up, as does the effect for quantifier, the block by polarity interaction, and the block by polarity by quantifier interaction. The effect for polarity is totally wiped out, $F(1, 24) = 0.65$, $p = 0.43$. Similarly, the block by quantifier interaction disappears, $F(1, 24) = 0.74$, $p = 0.40$.

So, one robust effect is that for quantifier. In all four experiments, we see that people were better at generating existential proofs than universal ones. The interactions involving performance can best be understood by studying the changes in strategies with hint.

Clearly, these hints did not work. Performance was especially damaged on universals and negatives, presumably the harder types of claims. To see what the provers believed about their own proofs, a 2x2x2 (Block x Quantifier x Polarity) ANOVA was performed on their confidence ratings for their own proofs.

The confidence for existential proofs was correctly higher ($M = 3.81$) than that for universals ($M = 3.21$), $F(1,27) = 39.59$, $p < 0.01$. Confidence was also higher for negatives ($M = 3.55$) than for positives ($M = 3.46$), $F(1,27) = 5.77$, $p = 0.02$. While confidence showed a slight dropoff for negative proofs after the hint (from 3.59 to 3.51), it showed a gain for positives (from 3.37 to 3.59), $F(1,27) = 8.93$, $p < 0.01$. There was a marginal interaction between quantifier and polarity showing a larger difference between the positive and negative universal (Ms = 3.14, 3.27 respectively) than that between the positive and negative existential (Ms = 3.78, 3.83 respectively), $F(1,27) = 3.24$, $p = 0.08$. The three-way interaction with block suggests that this greater difference in universals comes mainly from the early claims, as later ones actually show negative universals to have a lower overall rating than the positives ($M = 3.15$ for the post-hint negative universals as opposed to $M = 3.24$ for the post-hint positive universals), $F(1,27) = 4.12$, $p = 0.05$. The pattern of confidence ratings suggests clearly that there is less confidence in proofs of universal claims, and the other results hint at very slight changes in certainty after the hint. It appears that since all the confidence ratings are high, above 3, that the provers seem generally confident in their proofs.

While performance did not improve (declined nonsignificantly) after the hint, and confidence ratings do not necessarily reflect where particpants had trouble, the changes in strategy choices are telling about what constitutes the participants' difficulties. In order to study strategy changes with the various kinds of hints, nine separate 2x2x2 ANOVAs were run, one for each of the three hint types using the proportion of usage for a single example, variables, and multiple examples respectively as dependent measures while block, quantifier, and polarity were the independent variables. See Figures 3-5 for the changes in strategy choices with each type of hint.

5.2.1 Quantifier Hint

In the quantifier hint, the focus was on the difference between existential and universal claims, so any shift in strategy after the hint ought to highlight that distinction. As would be expected globally, the proportion of time that Example was the strategy was greater for existential claims (70%) than for universals (26%), $F(1,9) = 26.37$, $p < 0.01$. No other effects were found for use of the Example strategy. The Variable strategy shows only one effect–an interaction between polarity and block such that there is an increase in usage of variables for the positive claims (from 6.5% to 23.5%), but a decline in their usage for negative claims (23.5% to 13.5%), $F(1, 9) = 6.63$, $p = 0.03$. Since polarity is orthogonal to the quantification described in the hint, this result is difficult to interpret. The multiple examples strategy is used more often for the universals (37%) than for the existentials (6.5%), $F(1, 9) = 11.12$, $p = 0.01$. No other effects were found, so it is safe to conclude that the quantifier hint in which participants are told to be aware of differences between existential and universal claims does not impact their strategy choice much.

5.2.2 Variable Hint

For the variable hint, in which participants were given a few demonstrations of how to use variables in proofs, it would seem reasonable to expect greater usage of variables in proofs. Ideally, the increase in usage would be for universals, where variables are needed more than existentials. Here, the use of the Example strategy is greater for existentials (58%) than universals (15%), $F(1, 8) = 33.88$, $p < 0.01$. People shift away from using a single example for existentials after the hint (dropping from 67% to 49%), but use them more for universals (8% to 22%), $F(1, 8) = 20.47$, $p < 0.01$. The use of Variables in the variable hint condition shows an interaction between block and polarity such that there is a slight dropoff in usage for negative claims (from 35% to 32%), but an increase for positive claims (from 20% to 41%), $F(1, 8) = 8.28$, $p = 0.02$. There is a marginal interaction between block and quantifier such that there's an increase in variable use for existentials (from 20% to 41%), with a slight dropoff for universals (from 39% to 35%), $F(1, 8) = 3.94$, $p = 0.08$. The increases shown for positive claims and existentials suggest that participants have become more familiar with applying variables to their problems, and are more apt to do so; the corresponding drops for negatives and universals are not large ones. The use of multiple examples in the variable hint condition shows a marginal dropoff with the hint from 16% to 11%, $F(1, 8) = 3.56$, $p = 0.10$. Multiple examples are used more frequently for universals (22%) than for existentials (5%), $F(1, 8) = 5.64$, $p = 0.04$. There is a marginal interaction between block and polarity such that there's a slight dropoff in use of multiple examples for positives (from 21% to 19%), but an increase for negatives (from 7% to 23%), $F(1, 8) = 3.63$, $p = 0.09$. It seems that the variable hint did make the participants more aware of the notion that they should use variables, but unfortunately not selectively for the universal problems. It is also possible that since the variable blurb was about odd and even numbers, that the participants' focus was shifted to the parity of numbers being a cue for variable use, rather than the idea that variables provide a handy representation for any claim. However, proportion of variable use doesn't support this idea–variables are used just as much before as after the hint about variables for universal negatives (33%) and for existential negatives (also 33%).

5.2.3 Scope Hint

For the scope hint, in which the provers are focused on the difference between the existential negatives and universal negatives, any strategy changes would likely occur among these proof classes. Hopefully those changes would reflect a greater use of variables for the negative universals. Overall, we find the expected greater use of the Example strategy for existentials (72%) than for universals (14%), $F(1, 8) = 197.46$, $p < 0.01$. While none of the other first- or second-order effects reach significance, the three-way interaction between block, quantifier, and polarity (see Figure 6) indicates that the difference in use of a single example between the existentials and universals is greater for positives than negatives before the hint and becomes greater for negatives after the hint, $F(1, 8) = 19.39$, $p < 0.01$. This makes sense, since the scope hint focused people on the differences between the existential negative and the universal negative.

An unfortunate finding for our scope hint is that the use of Variable as a strategy drops off after participants received the hint (from 34% to 28%), $F(1, 8) = 5.71$, $p = 0.04$. Variables are, as would be expected, used more frequently for universals (46%) than existentials (17%), $F(1, 8) = 8.37$, $p = 0.02$. Variable usage shows an interaction between block and polarity such that their use increases for positives (24% to 37%), but drops off for negatives (44% to 19%), $F(1, 8) = 12.25$, $p = 0.01$. No other effects are found for Variables.

In the scope hint condition, the use of multiple examples is greater for universals (31%) than for existentials (4%), $F(1, 8) = 8.72$, $p = 0.02$. The scope hint pushed participants to increase their use of multiple examples for universals (24% to 37%), but not for existentials (which stayed at 4%), $F(1, 8) = 7.73$, $p = 0.02$. The interaction of block with polarity was marginal, with a slight decrease in use of multiple examples for the positives (21% to 19%), but an increase in its use for the negatives (7% to 23%), $F(1, 8) =4.12$, $p = 0.08$. I found the presence of a three-way interaction, with negative universals showing a large gain across block (7% to 41%), while the other proof classes' claims changed very little (at most a drop of 8% for positive universals).

Evidently, the focus of the scope hint on the difference between negative existentials and negative universals did make the participants alter their strategies. Participants realized that a single example was not sufficient to prove a universal negative claim. After the hint, they relied more upon multiple examples to demonstrate that the claim is true in a broader setting. Either they lacked the ability to use variables to set up proofs, or believed that multiple examples are in some sense more convincing than variables, because there was a general shift away from variables after the scope hint.

5.2.4 Summary

As in the first three experiments, the provers here were asked upon completion of all the proofs what their overall strategy was. Overall, 18 out of 28 participants (64%) mention quantifier as important in their choice of strategy. Twelve of those (43% of the total) had the correct idea that the universal proofs required variables while existentials require only a single numerical example. The other six have misconstruals of what to do with the quantifier, including three who believed that they needed multiple examples for universals. The other strategies used by more than a single prover were to use numbers because they're easier (18%), to just choose the easiest strategy (7%), or base the decision on the specific features of the claim (7%). Awareness of quantifier does not seem to be enhanced greatly by the hint (except perhaps the scope hint), as eight out of nine mentioned it in the scope hint condition, five out of ten mentioned it in the quantifier hint condition, and four of nine mentioned it in the variable hint condition. So, it does seem that many participants at least have some notion regarding the need for a different proof technique for universal claims than for existentials. However, many have a misconception about how much can be inferred from a proof using multiple examples. An additional stumbling block for some provers is that they either do not know how to set up the proof or do not want to exert extra effort by using variables.

Chapter 6

General Discussion

People's ideas of what constitutes a proof vary. In Experiment 1, participants were asked to prove claims from the four proof classes formed by crossing quantifier (existential or universal) with polarity (positive or negative). Their proofs made heavy use of examples–either single or multiple instances for which the claim was true. Such proofs successfully prove existential claims, but fail to prove universal claims. For universal claims, it is necessary to use an abstract representation–either variables or a verbal description–which demonstrates the truth of the claim for all possible instances. While the single most common strategy for universal claims was variables, there were almost as many proofs for universals using multiple examples. The apparent belief that multiple examples provide a sufficient proof for universals is a kind of naive induction on the part of the participants. If you consider the proportion of usage of single or multiple examples in universal proofs to be a measure of use of some inductive strategy, you can account for 50% of all the proofs, as opposed to the 41% for which variables are used. This acceptance of (less than sufficient) inductive proofs is consistent with the findings of Koedinger and Anderson (1991) and Martin and Harel (1989), whose participants were willing to accept single or multiple instances or examples as sufficient proofs of positive universal statements. Here, for the existential claims, using a single example was the preferred strategy (59%), followed by the use of variables (28%). So, it does seem that for many of the existential claims, participants were aware that only a single example was required.

Due mainly to participants' strategy choices, and to a lesser extent, their execution of strategies, performance was better for existentials than universals. Interestingly, performance was slightly better for the positive claims than the negatives, and this difference did not interact with quantifier. The participants may have some sense of how well their proofs were written, rating their own existential proofs as marginally better than their universal ones, and rating their positive ones barely better than their negative ones.

If participants may have too much cognitive load to notice or use the quantifier when required to write proofs, it stood to reason that giving them "attempted proofs" to rate would lessen their load. Then, with more resources to devote to the quantifier, if they understood the need for abstraction in universal proofs, they ought to rate the inductive proofs as insufficient, but accept proofs using variables. In Experiment 2, I gave participants completed "attempted proofs" of various claims. The proofs used the three most commonly used strategies of the first experiment–single example, multiple examples, and variables. The strategies' frequencies were balanced to allow each of them to work for 50% of the proofs in which they appeared.

Participants' overall ratings were in accord with the ratings given to the proofs by the experimenter beforehand. However, they differ in their beliefs about which strategies may be appropriate for each claim. At the conclusion of the experiment, 56% of the participants said that they selected the preferred strategy based upon quantifier. Even more, 83%, mentioned quantifier during the course of the experiment as being crucial to their choice of preferred strategy for a given proof. So, there is evidence of an awareness that the quantifier is important when choosing a proof strategy, even if there are many occasions on which it was not acknowledged and the overall strategy described at the end did not take quantifier into account. Unfortunately, even though quantifier is noticed by most participants, the universal quantifier's need for an abstract proof is not as well known.

With an open mind to the idea that participants may not have the same taxonomy of proofs, Experiment 3's sorting task gives insight into how participants view categories of proofs. Participants were in one of two conditions, proof class or self sort. The proof class participants were given descriptions of the four proof classes formed by crossing quantifier with polarity, and asked to sort claims into those four categories. Self sort participants were told to sort the claims into piles based upon the strategy they would use when proving the claims. Both groups were then asked to prove the claims.

The hierarchical cluster analysis shows that the proof class participants sorted the claims based upon quantifier, as the top level split was between the existential and universal claims. Within that, the next level split was based upon polarity. So, it seems that given the instructions to sort by proof class, the participants had the ability to do so. The self-sort participants sorted their claims by quantifier, but the clusters within each of the sets are not based upon polarity; rather, the sorts were done based upon the types of operations required by each proof type.

This experiment also reaffirmed the results of the earlier ones to the extent that performance was again better for existential proofs than universal proofs, but no effect was found for polarity here. Again, the difference in performance can mainly be attributed to the selection of strategies.

A major goal in research on proof techniques ought to be the improvement of the training students receive on how to prove claims. With that in mind, the fourth experiment gives participants 12 claims to prove, one of three short hint passages, and then another 12 claims to prove. The types of hint focus on the quantifier, use of variables, and the scope of the negative. The quantifier hint instructs the participants to be aware of the difference between claims which assert something to be true or false for all possible cases and those which assert that something is true for at least one case. The variable hint explains briefly how certain problems may be represented using variables. Finally, the scope hint explains the difference in scope between the universal and existential negative ("Show that something is not true for all possible cases" vs. "Show that there is a case for which something is not true").

While the various hints were not found to improve performance on the proofs, interesting changes in strategies took place, especially for the scope hint. In the scope hint condition, there was a shift away from using single examples or variables toward using multiple examples. The dropoff in use of examples is greater for the negative universals, showing some success for the hint's focus, the difference between the negative universals and the negative existentials. The variable hint yielded increases in variable usage for positive and for existential claims, suggesting that making participants aware of how variables may be used to represent the claims may lead to increased use, but not to successful selection of the variable strategy for all universal claims. The quantifier hint did not have much of an effect upon strategy choices. Training for proofs has been difficult for other researchers too–note that Koedinger and Anderson (1991) did not see improved performance after training on proofs either.

Throughout all four experiments, many participants noted the quantifier, but still were unable to correctly execute proofs of the appropriate type for the universal claims. So, even with an appreciation of the quantifier's importance, more work needs to be done on improving student's knowledge of how to apply variables in proofs. Students should not continue to choose "numbers" for their proofs just because they are "easier than variables."

Recall the four possible types of proof difficulties for participants mentioned in the introduction:

1. They fail to distinguish between the universal and existential quantifiers.
2. They don't know which proof strategies are appropriate for each quantifier.
3. They cannot represent claims using variables.
4. They don't have the skill to manipulate variables.

The first of these, the idea that participants cannot distinguish between universal and existential quantifiers, can be soundly rejected. Looking at all of the experimental results, the difference is noticed. Use of a single example is consistently the strategy of choice for existential claims and not for universals. In the second experiment, when participants did not have to generate their own proofs, but critiqued proofs which were already written, the numbers clearly show awareness of quantifier: 56% cited the difference in quantifier as important in their judgments, and 83% mention quantifier as they were going through their ratings. In the third experiment too, a clear distinction between existential and universal claims is found in the cluster analysis–not only do the participants in the proof class condition split along quantifier, but so do those in the self sort condition.

Evidence for the second type of difficulty is strewn throughout the experiments. While use of a single example is appropriately the consistent favored strategy for existential claims, participants use multiple examples more than variables for universals. They also show greater confidence in use of multiple examples in proofs than using variables. In Experiment 2, they rate use of multiple examples for universal proofs higher than variables. The fourth experiment, in which the hints emphasized differences between problem types, the scope hint, which focused on the difference between the existential negative and universal negative, lowered the frequency of variable use overall. The scope hint also increased the use of multiple examples for universals, suggesting that with a greater awareness of the quantifier in a negative claim, participants were more likely to use the wrong strategy.

The third difficulty type, having problems setting up the claims using variables, is a difficulty for a minority of the participants. In the first experiment, only two out of the 20 people could not set up any of the proofs correctly using variables (This includes the original proofs of the given claims and those done when asked to go back and use variables for all proofs for which they had not used variables). Similarly, most of the participants throughout the experiments seemed to have the ability to adequately represent a generic number using a variable.

Finally, participants can generally manipulate the representations which use variables once they have gotten past the stage of setting up the representation. The initial hurdle of setup may be explained by an inadequate goal state for the representation; if participants are unaware of what resulting expression could prove the claim, they may not have the chance to show their abilities by manipulating the representation at hand.

To sum up, the majority of participants can tell the difference between claims having a universal and an existential quantifier, so (1) is not a common problem. The main difficulty revealed by this line of research has been (2). Even with the awareness that something different ought to be done to prove universal claims, participants often rely upon multiple examples rather than variables for those proofs. There are a few people who have difficulties setting up claims using variables, making (3) a difficulty which might require more attention. As for (4), once participants set up their proofs, if they have chosen to use variables, they seem to be facile at algebraic manipulations.

6.1 Future research

While we have gained insight into students' perceptions of what constitutes a proof, many questions remain to be addressed. Clearly, people's understanding of the importance of the quantifier in determining the strategy chosen for a proof needs to be reinforced in their mathematics education. The way in which this training is given ought to receive more attention. Training which is more intensive than a hint of a paragraph or two should be tested. The variable hint in Experiment 4 only focused on how to use variables to set up problems with odds and evens. The training on how to apply variables should attend more to the concept that a variable in a proof represents a generic number, which allows us to draw conclusions about generic members of the set from which that number is pulled, laying out the rule for $\forall$-introduction.

The design of such a training study would parallel that of Experiment 4. One training type will focus on "quantifier," meaning that it would emphasize that universal claims need more than a single supporting example. Another training type will focus on sense of scope, namely stating the difference between negative and counterexample claims, clarifying the scope of the negative applied in the claim. The third type will show people how to represent the claim in terms of variables. For each training type, participants would have to answer several questions ensuring that they understood the training blurb. In quantifier training, they would have to identify whether each of several claims are "for all" or "there is" claims. Similarly for the scope training, they would have to identify whether each of several claims is one which is false for all instances or one which is false for (at least) an example. For variable training, participants would have to set up several claims using variables. For all conditions, the participants would be given an answer sheet with correct answers, and a chance to ask any questions they may have about the answers. Other training types may be introduced if other types of errors commonly occur. A follow-up set of proofs would indicate whether such training could repair the difficulties in participants' proof techniques.

6.2 Conclusion

The proof classes form a taxonomy for mathematical claims. A claim's proof class should yield insight into the strategy (or strategies) which would be appropriate for proving it. An essential key to successfully proving a claim is the understanding of its quantifier. For universal claims, unless an elegant verbal explanation is available, variables are necessary for representing the claim. However, for existential claims, variables are not necessary, but may offer assistance in proving a claim.

Representation with variables is not yet fully meaningful to students who have not gained an appreciation for the distinction between the universal and existential quantifiers. They first need to learn that a universal statement such as "All multiples of 6 are even" cannot be proven by an example, "12 is even," or even four examples and an ellipsis, "12 is even, 18 is even, 24 is even, 30 is even..." Instead, a proof might be the following: "6 is the product of 2 and 3, so all multiples of 6 can be expressed as 2(3)(x) where x is an integer. Since 2(3x) is a multiple of 2, it is even. Hence all multiples of 6 are even."

Note that the preceding proof did not require the variable. The following (equivalent) explanation would have sufficed: "6 is the product of 2 and 3, so all multiples of 6 are multiples of 2, meaning they are all even." Variables are convenient tools to simplify representations for problems in which the descriptive proofs are more cumbersome.

Understanding the meaning behind the quantifier is the key to finding the correct strategy for a proof. Apparently, most people have an appreciation for the difference between the universal and existential quantifiers, but many do not know what distinguishes a successful universal proof from a successful existential one.

Tables

Table 1
Weaver & Kintsch (1992) word problem forms

An example of “same structure” pair is comprised of the first and second problems; an example of a “same equation” pair is the first and third:

> The area of one plot of land is the same as the area of a second plot of land. The length of the first plot is 500 m, while the length of the second plot is 400 m. The width of the second plot is 25 m more than the width of the first plot. What is the width of the first plot?
>
> There are 2 electrical circuits, A and B. The current in circuit A is 12 amps, while the current in circuit B is 9 amps. The resistance of B is 2 more than that of A. The voltage in the two circuits is the same. What is the voltage in the two circuits?
>
> Ernie invests some money at 10% interest and $350 more than that at 8% interest. Both earn the same interest. How much was invested at each interest rate?

Table 2
Instances of each of the commonly used strategy types in Experiment 1

Example:
Show that there is a number which when multiplied by 4 and 3 is subtracted gives an answer which is not even.
number = 5
(5x4)-3 = 20-3 = 17. (not even)

Multiple Example:
Show that every odd multiple of 5 when squared is not divisible by 2.
odd multiples of 5: 5, 15, 25...
$5^2 = 25 \rightarrow$ not divisible by 2.
$15^2 = 225 \rightarrow$ not divisible by 2. Therefore shown.

Variables:
Show that there is a number for which you can get the same result as squaring it if you do the following: Subtract one from the original number, and then multiply that by the original number plus one. Add one to that result.
$[(x-1)(x+1)] + 1 = x^2-1+1 = x^2$
$x^2 = x^2$

Table 3
Strategy choices (given in percentages) for all proofs (n=320) in Experiment 1

	Existential		Universal	
Strategy	Positive	Negative	Positive	Negative
Single Example	58.8	58.8	17.6	17.3
Variables	30.0	26.3	38.8	44.0
Multiple Examples	3.8	8.8	34.1	32.0
Other	7.5	6.3	9.3	6.7

Strategy choices for correct proofs (n=166) in Experiment 1

	Existential		Universal	
Strategy	Positive	Negative	Positive	Negative
Single Example	67.6	69.4	0.0	0.0
Variables	22.1	14.5	90.5	100.0
Multiple Examples	4.4	11.3	0.0	0.0
Descriptive	0.0	0.0	9.5	0.0
Other	5.9	4.8	0.0	0.0

Strategy choices for incorrect proofs (n=154) in Experiment 1

	Existential		Universal	
Strategy	Positive	Negative	Positive	Negative
Single Example	8.3	22.2	23.4	21.7
Variables	75.0	66.7	21.9	30.0
Multiple Examples	0.0	0.0	45.3	40.0
Descriptive	0.0	0.0	3.1	6.7
Other	16.7	11.1	6.3	1.7

Table 4
Understanding of optimal strategies in Experiment 2

	Existential		Universal	
Participant	May not have understood	Understood	Did not under-stand	Understood
1	0.600	0.400	0.750	0.250
2	0.500	0.500	0.833	0.167
3	0.400	0.600	0.833	0.167
4	0.350	0.650	0.750	0.250
5	0.450	0.550	0.750	0.250
6	0.350	0.650	0.834	0.167
7	0.650	0.350	0.833	0.167
8	0.450	0.550	0.833	0.167
9	0.450	0.550	0.416	0.583
10	0.450	0.550	0.917	0.083
11	0.450	0.550	0.333	0.667
12	0.500	0.500	1.000	0.000
13	0.400	0.600	0.833	0.167
14	0.400	0.600	0.834	0.167
15	0.500	0.500	0.834	0.167
16	0.250	0.750	0.333	0.667
17	0.400	0.600	0.000	1.000
18	0.400	0.600	0.750	0.250

Table 5
Categories generated in self-sort condition of Experiment 3

Category name	Number of people using category	Average number of problems	Proportion of problems accounted for by category
operation(s)	5	15.00	0.21
multiple examples	5	9.20	0.13
parity	5	7.60	0.10
example	4	9.25	0.10
steps	4	5.00	0.05
trial and error	3	6.33	0.05
variables	2	6.00	0.03

Table 6
Judges' rating of performance on proofs in Experiment 4
(ratings on four-point scale described above)

		Early		
Hint	Positive existential	Negative existential	Positive universal	Negative universal
Quantifier	3.63	3.93	2.90	3.20
Scope	3.74	4.00	2.85	3.22
Variable	3.89	3.41	2.48	2.93

		Late		
Hint	Positive existential	Negative existential	Positive universal	Negative universal
Quantifier	3.93	3.87	2.70	2.67
Scope	3.78	3.89	2.89	2.78
Variable	3.74	3.81	2.67	2.59

Figures

Figure 1a. Cluster analysis for Proof Class sort in Experiment 2

Figure 1b. Cluster analysis for Self Sort in Experiment 2.

Figure 2. Performance as it varies based upon block, quantifier and polarity in Experiment 4.

Figure 3. Strategy usage in Experiment 4 for Quantifier Hint.

Figure 4. Strategy usage in Experiment 4 for Scope Hint.

Figure 5. Strategy usage in Experiment 4 for Variable Hint.

Figure 6. Proportion of use of single example strategy in Experiment 4.

References

Cajori, F. (1910). A history of elementary mathematics. New York: The Macmillan Co.

Chi, M.T.H., Feltovich, P.J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. Cognitive Science, 5, 121-152.

Clark, H.H., & Chase, W.G. (1972). On the process of comparing sentences against pictures. Cognitive Psychology, 3, 472-517.

Dellarosa Cummins, D. (1992) Role of analogical reasoning in the induction of problem categories. Journal of Experimental Psychology: Learning, Memory, & Cognition, 18, 1103-1124.

Dellarosa Cummins, D., Kintsch, W., Reusser, K. & Weimer, R. (1988). The role of understanding in solving word problems. Cognitive Psychology, 20, 405-438.

Dufour-Janvier, B., Bednarz, N., & Belanger, M. (1987) In C. Janvier (Ed). Problems of representation in the teaching and learning of mathematics. Hillsdale, NJ: Erlbaum.

Greeno, J.G. (1989). Situations, mental models and generative knowledge. In D. Klahr & K. Kotovsky (Eds.), Complex information processing: The impact of Herbert A. Simon. Hillsdale, NJ: Erlbaum.

Just, M.A., & Carpenter, P.A. (1971). Comprehension of negation with quantification. Journal of Verbal Learning And Verbal Behavior, 10, 244-253.

Kintsch, W., & Greeno, J.G. (1985). Understanding and solving word arithmetic problems. Psychological Review, 92, 109-129.

Kleiner, I. (1991). Rigor and proof in mathematics: a historical perspective. Mathematics Magazine, 64, 291-315.

Koedinger, K.R., & Anderson, J.R. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. Cognitive Science, 14, 511-550.

Koedinger, K.R., & Anderson, J.R. (1991). Interaction of deductive and inductive reasoning strategies in geometry novices. In Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society, Hillsdale, NJ: Erlbaum.

Kruskal, J.E., & Wish, M. (1978). Multidimensional scaling. Beverly Hills: Sage Publications.

Larkin, J.H., & Simon, H.A. (1987). Why a diagram is (sometimes) worth ten thoussand words. Cognitive Science, 11, 65-99.

Martin, W.G., & Harel, G. (1989). Proof frames of preservice elementary teachers. Journal for Research in Mathematics Education, 20, 41-51.

McCawley, J.D. (1993). Everything that linguists have always wanted to know about logic but were ashamed to ask, 2nd edition. Chicago: The University of Chicago Press.

Nathan, M.J., Kintsch, W. & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. Cognition and Instruction, 9, 329-389.

Newstead, S.E., & Griggs, R.A. (1983). Drawing inferences from quantified statements: A study of the square of opposition. Journal of Verbal Learning and Verbal Behavior, 22, 535-546.

Reed, S.K. (1993). A schema-based theory of transfer. In D.K. Detterman & R.J. Sternberg (Eds.), Transfer on trial: Intelligence, cognition, and instruction. Norwood, NJ: Ablex.

Resnick, L.B., & Ford, W.W. (1981). The psychology of mathematics for instruction. Hillsdale, NJ: Erlbaum.

Schoenfeld, A.H., & Hermann, D.J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solvers. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 484-494.

Tabachneck, H.J.M., Koedinger, K.R., & Nathan, M.J. (1994). Toward a theoretical account of strategy use and sense-making in mathematics problem solving. In Proceedings of the 17th Annual Conference of the Cognitive Science Society (pp. 836-841).

Weaver III, C.A., & Kintsch, W. (1992). Enhancing students comprehension of the conceptual structure of algebra word problems. Journal of Educational Psychology, 84, 419-428.

Appendices

Appendix A: Claims used in Experiment 1

Only one of the four proofs for each example was presented.

1

Positive existential:

Show that there is a number that when multiplied by 2 and then 4 is added, the result will be even.

Negative existential:

Show that there is a number that when multiplied by 3 and then 5 is addded, the result will not be even

Positive universal:

Show that for every number, taking a number and multiplying it by 2 and then adding 4 yields an even number.

Negative universal:

Show that for every number, taking a number and multiplying it by 2 and then subtracting 6 gives an even number.

2

Positive existential:

Show that there's a number for which tripling it and adding one changes the parity (whether it is odd or even)

Negative existential:

Show that there's a number for which tripling it does not change the parity (whether it is odd or even)

Positive universal:

Show that for every number, tripling it and adding one changes the parity (whether it is odd or even)

Negative universal:

Show that for every number, tripling it does not change the parity (whether it is odd or even)

3

Positive existential:

Show that there's a number which is divisible by 10 which is even.

Negative existential:

Show that there's a number which is divisible by 8 which is not odd

Positive universal:

Show that every number divisible by 4 is even

Negative universal:

Show that every number divisible by 6 is not odd

4

Positive existential:

Show that there's some product of an even number and an odd number which is even.

Negative existential:

Show that there's some sum of an even number and an odd number which is not even.

Positive universal:

Show that every product of an even number and an odd number is an even number

Negative universal:

Show that every sum of an even number and an odd number is not an even number

5

Positive existential:

Show that there are positive whole numbers for which if one number is divisible by a second, and the second is also divisible by the third one, then the second is divisible by the first.

Negative existential:

Show that there are positive whole numbers for which if one number is divisible by a second, and the second is also divisible by the third one, then the second is not always divisible by the first.

Positive universal:

Show that for all positive whole numbers, if one number is divisible by a second one and the second is also known to be divisible by some third number, then the first number is divisible by the third one.

Negative universal:

Show that for all positive whole numbers, if one number is divisible by a second one and the second is also known to be divisible by some third number, then the third one is not divisible by the first if they are not the same number .

6

Positive existential:

Show that there is a number for which you can get the same result as squaring it if you do the following: Subtract one from the original number, and then multiply that by the original number plus one. Add one to that result.

Negative existential:

Show that there is a number for which you cannot get the same result as squaring it if you do the following: Subtract one from the original number, and then multiply that by the original number minus one. Add one to that result.

Positive universal:

Show for every number that you can get the same result as squaring the number if you do the following: Subtract one from the original number, and then multiply that by the original number plus one. Add one to that result.

Negative universal:

Show for every number that you cannot get the same result as squaring the number if you do the following: Subtract one from the original number, and then multiply that by the original number minus one. Add one to that result.

7

Positive existential:

Show that there are a pair of consecutive whole numbers whose product is even.

Negative existential:

Show that there are a pair of consecutive whole numbers whose product is not odd.

Positive universal:

Show that for every pair of consecutive whole numbers, their product is even.

Negative universal:

Show that for every pair of consecutive whole numbers, their product is not odd.

8

Positive existential:

Show that there's a number for which if you subtract 2, multiply that by 3, add 6, divide by 3, and then add 13, you do get 11.

Negative existential:

Show that there's a number for which if you subtract 2, multiply that by 3, add 6, divide by 3, and then add 13, you do not get 11.

Positive universal:

Show that for every number, if you subtract 2, multiply that by 3, add 6, divide by

3, and then add 13, you get the original number plus 13.

Negative universal:

Show that for every number, if you subtract 2, multiply that by 3, add 6, divide by 3, and then add 13, you don't get the original number plus 6.

9

Positive existential:

Show that there's a number that will yield 6 if you do the following: Multiply the number by 3, add 9, multiply by 2/3, and subtract twice the original number.

Negative existential:

Show that there's a number that will not yield 5 if you do the following: Multiply the number by 3, add 9, multiply by 2/3, and subtract twice the original number.

Positive universal:

Show that every number will yield 6 if you do the following: Multiply the number by 3, add 9, multiply by 2/3, and subtract twice the original number.

Negative universal:

Show that every number will not yield 5 if you do the following: Multiply the number by 3, add 9, multiply by 2/3, and subtract twice the original number.

10

Positive existential:

Show that there's a number that will yield itself minus four when the following operations are performed on it: Take the number, subtract 1 from it, multiply by 4, and then subtract 3 times the original number.

Negative existential:

Show that there's a number that will not yield itself when the following operations are performed on it: Take the number, subtract 1 from it, multiply by 4, and then subtract 3 times the original number.

Positive universal:

Show that every number will yield itself minus four when the following operations are performed on it: Take the number, subtract 1 from it, multiply by 4, and then subtract 3 times the original number.

Negative universal:

Show that every number will not yield itself when the following operations are performed on it: Take the number, subtract 1 from it, multiply by 4, and then subtract 3 times the original number.

11

Positive existential:

Show that there's a multiple of 13 that will be a multiple of 13 when doubled and then 13 is subtracted.

Negative existential:

Show that there's a multiple of 4 that will not be a multiple of 4 when doubled and then 13 is subtracted.

Positive universal:

Show that every multiple of 13 will be a multiple of 13 when doubled and then 13 is subtracted.

Negative universal:

Show that every multiple of 4 will not be a multiple of 4 when doubled and then 13 is subtracted.

12

Positive existential:

Show that there is an odd multiple of 3 which when squared is divisible by 5.

Negative existential:

Show that there is an odd multiple of 3 which when squared is not divisible by 5.

Positive universal:

Show that every odd multiple of 5 when squared is divisible by 5.

Negative universal:

Show that every odd multiple of 5 when squared is not divisible by 2.

13

Positive existential:

Show that there is a number which when multiplied by 4 and then 3 is subtracted gives an answer which is odd.

Negative existential:

Show that there is a number which when multiplied by 4 and then 3 is subtracted gives an answer which is not even.

Positive universal:

Show that every number which when multiplied by 4 and then 3 is subtracted gives an answer which is odd.

Negative universal:

Show that every number which when multiplied by 4 and then 3 is subtracted gives an answer which is not even.

14

Positive existential:

Show that there's a number for which the following process doubles the original number: Add one to the number. Double that result and take two away from it.

Negative existential:

Show that there's a number for which the following process does not double the original number: Subtract one from the number. Double that result, and take two away from it.

Positive universal:

Show that, for every number, the following process doubles the original number: Add one to the number. Double that result and take two away from it.

Negative universal:

Show that, for every number, the following process does not double the original number: Subtract one from the number. Double that result, and take two away from it.

15

Positive existential:

Show that there are three odd numbers whose sum is odd.

Negative existential:

Show that there are three odd numbers whose sum is not even.

Positive universal:

Show that every three odd numbers have a sum which is odd.

Negative universal:

Show that every three odd numbers have a sum which is not even.

16

Positive existential:

Show that there is a positive number which the following will double: Take the number, multiply it by 3, subtract 3, divide that by 3, and then add the original number plus 1.

Negative existential:

Show that there is a positive number which the following will not triple: Take the number, multiply it by 3, subtract 3, divide that by 3, and then add the original number plus 1.

Positive universal:

Show that every positive number will double if the following is done: Take the number, multiply it by 3, subtract 3, divide that by 3, and then add the original number plus 1.

Negative universal:

Show that every positive number will not triple if the following is done: Take the number, multiply it by 3, subtract 3, divide that by 3, and then add the original number plus 1.

Appendix B: "Attempted Proofs" used in Experiment 2 (two sets of claims 1-16, for a total of 32 proofs)

Supportive Existential Positive Examples

1. Show that there is a pair of two even numbers whose product is even.

6 and 10 are both even.

6 x 10 = 60.

60 is even.

2. Show that there is a number which the following process would double: Add one to the number. Double that result and take two away from it.

Let 3 be the number.

2 (3+1)-2 =

2(4)-2 =

8-2 =

6

6 is 3 doubled.

3. Show that there is a number which when squared becomes the sum of a set of three consecutive odd numbers.

$3^2 = 9 = 1 + 3 + 5$.

1, 3, and 5 are three consecutive odd numbers, and $3^2 = 9$, which is their sum.

Supportive Existential Positive Multiple Examples

1. Show that there is an odd multiple of 3 which when squared is divisible by 5.

15 is an odd multiple of 3, since it's 3x5.

$15^2 = 225$.

225 is divisible by 5.

45 is an odd multiple of 3, since it's 3x15.

$45^2 = 2025$.

2025 is divisible by 5.

2. Show that there are three odd numbers which have an odd sum.

3+5+7 = 15.

11+13+15 = 39.

23+25+27 = 75.

For each of these trios of odd numbers, the sum was odd.

3. Show that there is a number for which doubling it and adding one changes the parity (whether it is odd or even).

Let 6 be the number. 6 is even.

2(6)+1 = 13.

13 is odd, so the parity changed.

Let 18 be the number. 18 is even.

$2(18) + 1 = 37$.

37 is odd, so the parity changed.

Supportive Existential Positive Variables

Show that there is a number for which you can get the same result as squaring it if you do the following: Subtract one from the original number, and then multiply that by the original number plus one. Add one to that result.

Let x be the number.

(x-1)(x+1)+1

$= x^2 + x - x - 1 + 1$

$= x^2$

So, doing all of that to x gives x squared.

Supportive Existential Negative Example

1. Show that there is a positive number which is not larger than its square.

2 is a positive number.

22=4.

2 is not larger than 4.

2. Show that there is a sum of an even number and an odd number that is not an even number.

$4 + 3 = 7$

Four is even, three is odd, and their sum, 7, is not even.

3. Show that there is a positive number which will not triple if the following is done: Take the number, multiply it by 3, subtract 3, divide that by 3, and then add the original number plus 1.

Let 2 be the number.

$\frac{(3(2)-3)}{3} + 2 + 1 =$

$\frac{6-3}{3} + 3 =$

$\frac{3}{3} + 3 =$

$1 + 3 =$

4

The process doubled 2; it didn't triple it.

Supportive Existential Negative Multiple Examples

1. Show that there is a sum of five odd numbers which is not even.

1+3+5+9+13 = 31.

3+7+11+33+101 = 155.

5+15+25+37+47 = 129.

All of those sums are odd.

2. Show that there is a number divisible by 30 which is not odd

60 is a number divisible by 30 which is not odd.

120 is a number divisible by 30 which is not odd.

3. Show that there's a number that will not yield 5 if you do the following: Multiply the number by 3, add 9, multiply by $\frac{2}{3}$ and subtract twice the original number.

Try 5:

$\frac{2}{3}$ (3(5) + 9) - 2(5) =

$\frac{2}{3}$ (15 + 9) - 10 =

$\frac{2}{3}$ (24) - 10 =

16 - 10 =

6

6 isn't 5.

Try 0:

$\frac{2}{3}$ (3(0) + 9) - 2(0) =

$\frac{2}{3}$ (9) - 0 =

6

6 isn't 5.

Try 1:

$\frac{2}{3}$ (3(1) + 9) - 2(1) =

$\frac{2}{3}$ (3 + 9) - 2 =

$\frac{2}{3}$ (12) - 2 =

8 - 2 =

6

6 isn't 5

Try 3:

$\frac{2}{3}$ (3(3) + 9) - 2(3) =

$\frac{2}{3}$ (9 + 9) - 6 =

$\frac{2}{3}$ (18) - 6 =

12 - 6 =

6

6 isn't 5.

Supportive Existential Negative Variables

Show that there is a number which when multiplied by 4 and 3 is subtracted gives an answer which is not even.

Let x be a whole number.

4 (x) - 3 =

4x - 3 =

2 (2x - 1) - 1, which is odd, since it's one less than a multiple of 2.

Supportive Universal Positive Example

Show that for all positive whole numbers: if one number is divisible by a second, and the second is also divisible by a third one, then the first is always divisible by the third.

Say 24 is the first, 8 is the second, and 4 is the third.

24 is divisible by 8 since $24 \div 8 = 3$.

8 is divisible by 4 since $8 \div 4 = 2$.

And, the first number, 24 is divisible by the third one, $24 \div 4 = 6$.

Supportive Universal Positive Multiple Examples

Show that every even number greater than 4 is a multiple of an odd number.

36 is even, and is a multiple of 9.

14 is even, and is a multiple of 7.

Supportive Universal Positive Variables

Show that every multiple of 13 will be a multiple of 13 when doubled and 13 is subtracted.

2(13k) - 13

= 13 (2k -1).

Since the result has 13 as a factor, it is a multiple of 13.

Supportive Universal Negative Example

Show that all numbers with three as a factor do not have 28 as a multiple.

3(4) = 12.

28 is not a multiple of 12.

Supportive Universal Negative Multiple Examples

Show that for every pair of consecutive whole numbers, their product is not odd.

2(3) = 6. 6 is not odd.

4(5) = 20. 20 is not odd.

7(8) = 56. 56 is not odd.

Supportive Universal Negative Variables

Show that for every number, if you subtract 2, multiply that by 3, add 6, divide by 3, and then add 13, you don't get the original number plus 6.

Let x be the number

$\frac{3(x-2)+6}{3} + 13 =$

$\frac{3x-6+6}{3} + 13 =$

$\frac{3x}{3} + 13 =$

x+13

x+13 is the original number plus 13, not the original number plus 6.

Unsupportive Existential Positive Example

Show that there is an odd number less than 25 which is a perfect square (a number whose square root is an integer).

11 is less than 25.

$11^2 = 121$.

So, 11 is a perfect square.

Unsupportive Existential Positive Multiple Examples

Show that there is a number which when divided by two yields an even number.

$\frac{6}{2} = 3$.

6 is an even number.

$\frac{10}{2} = 5$.

10 is an even number.

Unsupportive Existential Positive Variables

Show that there is a number between 1 and 10 for which adding it to 23 will give a value between 20 and 29.

23 +x = 43.

x=20, which is not a value in between 1 and 10.

Unsupportive Existential Negative Example

Show that there is a number that when multiplied by 3 and then 5 is added, the result will not be even

3(5) + 5 = 15 + 5 = 20

20 is even.

Unsupportive Existential Negative Multiple Examples

Show that there's a number that will not yield itself minus three when the following operations are performed on it: Take the number, subtract 1 from it, multiply by 4, and then subtract 3 times the original number.

Let 5 be the number.

(4(5-1) - 3) 5 =

(4(4) - 3) 5 =

(16 - 3) 5 =

13 x 5 =

65

That's not the same as 5-3, which would be 2.

Let 3 be the number.

(4(3-1) - 3) 3 =

(4(2) - 3)3 =

(8 - 3) 3 =

5 x 3 =

15

That's not the same as 3-4, which is -1.

Unsupportive Existential Negative Variables

Show that there are a pair of perfect squares whose sum is not a perfect square.

$x^2 + y^2 = (x + y)^2$

x+y is not a perfect square.

Unsupportive Universal Positive Example

Show that every even number can be written as the sum of two odd numbers and an even number.

2 + 4 + 6 = 12

12 is even. 2, 4, and 6 are also even.

Unsupportive Universal Positive Multiple Examples

Show that every number keeps its parity (whether it is odd or even) when it's squared.

3 x 2 = 6

5 x 2 = 10

Both 3 and 5 are odd, while 6 and 10 are even.

Unsupportive Universal Positive Variables

Show that every even multiple of 5 has a 0 as its right-most digit.

Try a multiple of 5 which is odd.

5x = 15

x=3

Unsupportive Universal Negative Example

Show that all odd numbers subtracted from 28 will yield a result which is not even.

31 - 28 = 3

31 is odd.

3 is not even.

Unsupportive Universal Negative Multiple Examples

Show that for all whole numbers, doubling a number will not yield an odd number.

2(.5) = 1

2(1.5) = 3

1 and 3 are odd numbers.

Unsupportive Universal Negative Variables

Show that all odd numbers are not multiples of even numbers.

Let x and y be even numbers. Then xy is the product of x and y. So xy is even.

Appendix C: Claims used in Experiment 3

(These claim numbers correspond to those in Figures 1a and 1b)

1. Show that there are a pair of perfect squares whose sum is not a perfect square.

2. Show that there is an odd multiple of 3 which when squared is divisible by 5.

3. Show that every even number can be written as the sum of two odd numbers and an even number.

4. Show that every three odd numbers have a sum which is odd.

5. Show that for every number, if you subtract 2, multiply that by 3, add 6, divide by 3, and then add 13, you don't get the original number plus 6.

6. Show that there is an odd number less than 25 which is a perfect square (a number whose square root is an integer).

7. Show that every even number can't be written as the sum of five odd numbers.

8. Show that there is a positive number which is not larger than its square.

9. Show that there's a number that will not yield itself minus three when the following operations are performed on it: Take the number, subtract 1 from it, multiply by 4, and then subtract 3 times the original number.

10. Show that every number keeps its parity (whether it is odd or even) when it's squared.

11. Show that every sum of an even number and an odd number is not an even number

12. Show that there are positive whole numbers such that if one number is divisible by a second, and the second is also divisible by a third one, then the first is divisible by the third.

13. Show that every number divisible by 30 is not odd

14. Show that there is a multiple of 28 that does not have 3 as a factor .

15. Show that there is a number for which you can get the same result as squaring it if you do the following: Subtract one from the original number, and then multiply that by the original number plus one. Add one to that result.

16. Show that for every pair of consecutive whole numbers, their product is even.

17. Show that there is a number between 1 and 10 for which adding it to 23 will give a value between 20 and 29.

18. Show for every pair of two even numbers that their product is even.

19. Show that every multiple of an even number is not odd.

20. Show that there is a number for which doubling it and adding one changes the parity (whether it is odd or even).

21. Show that there is a number which when multiplied by 4 and 3 is subtracted gives an answer which is not even.

22. Show that every multiple of 13 will be a multiple of 13 when doubled and 13 is subtracted.

23. Show that there's a number that will not yield 5 if you do the following: Multiply the number by 3, add 9, multiply by $\frac{2}{3}$ and subtract twice the original number.

24. Show that subtracting any odd number from 28 will yield a number which is not even.

Appendix D: Claims used in Experiment 4

1. Show that there are a pair of perfect squares whose sum is not a perfect square.

2. Show that there is an odd multiple of 3 which when squared is divisible by 5.

3. Show that every even number can be written as the sum of two odd numbers and an even number.

4. Show that every three odd numbers have a sum which is odd.

5. Show that for every number, if you subtract 2, multiply that by 3, add 6, divide by 3, and then add 13, you don't get the original number plus 6.

6. Show that there is an odd number less than 25 which is a perfect square (a number whose square root is an integer).

7. Show that every even number can't be written as the sum of five odd numbers.

8. Show that there is a positive number which is not larger than its square.

9. Show that there's a number that will not yield itself minus three when the following operations are performed on it: Take the number, subtract 1 from it, multiply by 4, and then subtract 3 times the original number.

10. Show that every number keeps its parity (whether it is odd or even) when it's squared.

11. Show that every sum of an even number and an odd number is not an even number

12. Show that there are positive whole numbers such that if one number is divisible by a second, and the second is also divisible by a third one, then the first is divisible by the third.

13. Show that every number divisible by 30 is not odd

14. Show that there is a multiple of 28 that does not have 3 as a factor .

15. Show that there is a number for which you can get the same result as squaring it if you do the following: Subtract one from the original number, and then multiply that by the original number plus one. Add one to that result.

16. Show that for every pair of consecutive whole numbers, their product is even.

17. Show that there is a number between 1 and 10 for which adding it to 23 will give a value between 20 and 29.

18. Show for every pair of two even numbers that their product is even.

19. Show that every multiple of an even number is not odd.

20. Show that there is a number for which doubling it and adding one changes the parity (whether it is odd or even).

21. Show that there is a number which when multiplied by 4 and 3 is subtracted gives an answer which is not even.

22. Show that every multiple of 13 will be a multiple of 13 when doubled and 13 is subtracted.

23. Show that there's a number that will not yield 5 if you do the following: Multiply the number by 3, add 9, multiply by $\frac{2}{3}$ and subtract twice the original number.

24. Show that subtracting any odd number from 28 will yield a number which is not even.