



Sources of Students' Difficulties with Proof By Contradiction

Jeffrey M. Rabin¹ · David Quarfoot¹

Accepted: 20 July 2021
© The Author(s) 2021

Abstract

The literature on proof by contradiction (PBC) is nearly unanimous in claiming that this proof technique is “more difficult” for students than direct proof, and offers multiple hypotheses as to why this might be the case. To examine this claim and to evaluate some of the hypotheses, we analyzed student work on proof construction problems from homework and examinations in a university “Introduction to Proof” course taught by one of the authors. We also conducted stimulated-recall interviews with student volunteers probing their thought processes while solving these problems, and their views about PBC in general. Our results suggest that the knowledge resources students bring to bear on proof problems, and how these resources are activated, explain more of their “difficulties” than does the logical structure of the proof technique, at least for this population of students.

Keywords Proof by contradiction · Hypothesis · Resources

Introduction

Proof is one of the key characteristics of mathematics as a discipline and serves a variety of functions within the mathematical community. Beyond providing a logical basis for a given claim, proof can help explain a theorem, systematize ideas, partially unmask the complex act of discovery, and communicate findings (de Villiers, 1990). As such, students are regularly trained in the production of proofs and asked, across grade levels, to make sense of them. At the college level, “Introduction to Proof” courses have been created to facilitate this growth, and extensive research has been conducted on both the teaching and learning of proof (Hanna & de Villiers, 2012).

✉ David Quarfoot
dquarfoot@ucsd.edu

Jeffrey M. Rabin
jrabin@ucsd.edu

¹ Department of Mathematics, University of California, San Diego, USA

Proof by contradiction (PBC) is an essential form of proof across all mathematical content areas, for example in proving the nonexistence of mathematical objects having certain properties. (The term *indirect proof* (IP) is often used to include both PBC and proof by contrapositive. Our focus in this paper is strictly on PBC and we will try to use that specific term consistently.) Mathematics education researchers are nearly unanimous in maintaining that PBC is “more difficult” for students than direct proof (DP) (Tall, 1979; Brown, 2018). This is somewhat puzzling, since people without specialized training use similar reasoning in everyday life (Reid & Dobbin, 1998). The common form of argument, “If X were true, then how do you explain Y?” is clearly intended as a PBC that X is false. Our thorough review of the literature (Quarfoot & Rabin, *in press*) revealed limited and conflicting evidence for the asserted difficulty of PBC relative to DP. We undertook the study described in this paper to collect evidence that might support the claimed difficulty of PBC within the mathematical context, and test various hypotheses that might explain it.

The literature on PBC, reviewed in Quarfoot and Rabin (*in press*), often treats the claim of “difficulty” very broadly, without circumscribing it by relevant contextual features. Some studies address students’ proof *construction*, while others are concerned with their proof *comprehension* or the degree of *conviction* produced by the proof in the statement that is proved. The age and mathematical background of the students are not always emphasized as relevant variables: secondary versus tertiary students, STEM majors versus preservice teachers, those in an Introduction to Proof course versus senior mathematics majors. The pedagogical approach used in the students’ initial exposure to PBC is likely also relevant, as are the specific mathematical content areas of the majority of examples they see.

Our study investigates proof *construction* by students enrolled in an Introduction to Proof course taught by one of the authors. We analyze their work on a “naturalistic” sample of proof problems: those that were assigned as homework or exam problems in the course. The purpose of this choice was to broadly represent the range of problems that shaped their understanding of the PBC technique rather than a few researcher-designed tasks whose idiosyncratic features might dominate the results. As described below, our data sources include the homework and exam solutions from all students who agreed to participate in the study ($N = 72$), and interviews with six student volunteers. We are not aware of previous research on proof construction involving this many participants or this variety of tasks.

Our study was initially designed around three main hypotheses we developed, from our own experience and sources in the literature, that might explain why PBC would be uniquely difficult in a mathematical context despite being a familiar mode of reasoning in everyday life. A fourth hypothesis emerged from the early stages of data analysis. Subsequently, as laid out in the companion paper (Quarfoot & Rabin, *in press*), we undertook a systematic literature search to catalog and organize the full range of hypotheses that have appeared in the existing PBC literature. Our study was not designed to address all of these, but we have tried to identify the implications of our data for them where possible. Using the labels from the framework presented in Fig. 1 below, our initial hypotheses, with brief descriptions, were as follows.

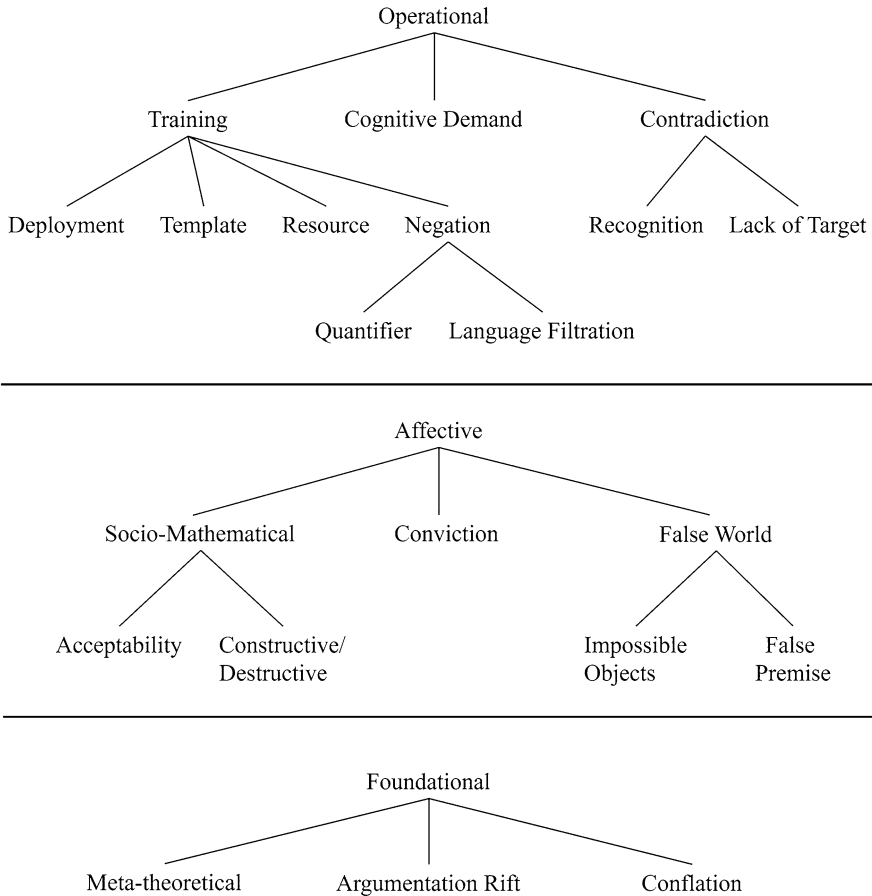


Fig. 1 The Hypothesis Framework for (Students' Difficulties with) Proof By Contradiction (HFPBC)

1. Recognition Hypothesis. Students have difficulty recognizing what constitutes a contradiction in the strict logical or mathematical sense: some proposition r and its negation $(r \wedge \neg r)$. It is not sufficient to reach a statement that is strange, implausible, or unfamiliar, or that is contrary to something learned but never proved in high school.
2. False World Hypothesis. PBC requires students to accept, for the sake of argument, premises that will eventually prove to be false, and may already be known to be false. This sort of counterfactual reasoning may be more challenging in the mathematical than the everyday context. For example, one can readily imagine that Hillary Clinton won the 2016 U.S. Presidential election, but how does one envision a counterfactual world in which there are only finitely many prime numbers? How does one suspend disbelief for the purpose of reasoning, and what sorts of reasoning can be considered reliable in such a world?

3. Lack of Target Hypothesis. In DP problems both the hypothesis and the conclusion are known at the outset, so one can aim at the conclusion or even work backward from it. In PBC, by contrast, the target is “a contradiction” which might appear at any stage and might contradict any piece of the student’s mathematical knowledge, or any prior assertion in the proof. This removes a potentially important piece of cognitive structure or scaffolding from the proof construction process.
4. Resource Hypothesis. Perhaps the difficulty of PBC is not due to the characteristic logical structure of this proof technique, but to the types of mathematical resources that must be drawn upon during the reasoning. For example, proofs of irrationality draw upon prior knowledge of rational and irrational numbers and divisibility. Students may activate unproductive resources during their proof construction, or those resources may be activated differently in the contexts of PBC and DP.

The Resource Hypothesis is naturally part of a theoretical framework that has been called Knowledge In Pieces (diSessa, 2013) or simply the Resource Framework (Hammer et al., 2005). It has been prominent in physics education research, and more generally in the study of conceptual change in science. It stands in opposition to an alternative “coherence” viewpoint that sees conceptual change in learning as similar to the process of theory change in science. In that viewpoint, students’ knowledge prior to instruction is seen as a coherent theory of their experience that happens to be false, rather like the flat Earth theory. Through learning they replace it with a normative theory aligned with the current scientific consensus. The Resource Framework holds instead that students’ prior knowledge consists of multiple relatively independent and finer-grained resources, not systematically linked by coherence or cross-referencing, that are activated in a highly context-dependent way and so may result in mutually inconsistent assertions being made in differing situations. Proofs, or explanations, provided by students are constructed from available resources in real time, not retrieved fully-formed from an organized store of knowledge, in response to cues in the problem situation. Within this framework, for example, one would not ask whether a student “really believes” that the quotient of two irrational numbers is necessarily irrational, or that $(a + b)^2 = a^2 + b^2$, but rather *in what situations* these resources are activated. The construct of resources is broader than, but related to those of concept image (Tall & Vinner, 1981) and example space (Goldenberg & Mason, 2008). A student’s rich concept image for a central concept in a problem will make a wide range of resources available to them, but resources are not necessarily linked to a specific concept and the student may not be thinking in terms of that (or any) concept in a given problem. Similarly a broad example space may include or suggest useful resources, or examples against which resources can be evaluated. A student’s resources also include epistemological resources or frames, that is, their understandings of what ways of justifying and using knowledge are appropriate and expected in a given educational context, for example whether examples or diagrams are acceptable forms of proof.

Before proceeding, it is important to note that our work does *not* provide a direct comparison of the relative difficulty of PBC versus direct proof. Designing such a

comparative study seems like a major challenge. It would presumably need to control for the “intrinsic difficulty” of the proof tasks selected, the mathematical content areas involved, the needed background knowledge, and perhaps the level of experience of the students with both proof techniques. Instead, we identify factors that underlie student difficulties with PBC that are visible in their written work or expressed to us in their interviews, and determine which of our hypotheses these factors are linked to. To some extent we can see whether these factors also occur in their work on DP or are unique to PBC. Our sample sizes are larger than in prior research on PBC, in terms of both numbers of student participants and number and variety of PBC tasks.

This paper is organized as follows. “[Literature Review](#)” contains a brief review of the literature on PBC and on the main hypotheses we study here, adapted from the comprehensive survey and framework for all the hypotheses proposed in our companion paper (blinded in press). “[Participants and Methods](#)” details the design and methods of our study, and “[Results](#)” presents our results. “[Discussion and Future Directions](#)” contains discussion of the results and future directions, including recommendations for both research and pedagogy.

Literature Review

Despite general claims that PBC is more difficult or less convincing for students than DP (Tall, 1979), little empirical evidence supports these claims, and this evidence is often contradictory. The most careful studies investigating the sense of conviction students derive from PBC versus DP were conducted by Brown (2012, 2013, 2018) and gave decidedly ambiguous results. When given side-by-side comparisons of DP and IP for the same theorem, students preferred a direct approach for some theorems and an indirect approach for other theorems. In summarizing her findings, Brown (2018) wrote: “it seems that length, complexity, and familiarity are criteria students bring to bear on proofs before considerations of proof type when selecting the most convincing proof” (p. 17). There are no comparable studies in the areas of proof production or comprehension, but one might expect the results to be similarly nuanced.

The literature does contain case studies of small numbers of students given specific PBC tasks, reflections by teachers of PBC on their students’ difficulties, and theoretical proposals for the origins of these difficulties. In our companion paper (Quarfoot & Rabin, [in press](#)), we exhaustively review the PBC literature, extracting all the hypotheses that have been seriously proposed to account for students’ difficulties with it, and organize them into a Hypothesis Framework for (Students’ Difficulties with) Proof by Contradiction (HFPBC), shown in Fig. 1.

We structure the space of hypotheses in three categories: Operational (those dealing with specific subtasks, steps, or demands in the process of producing a PBC), Affective (the emotional and attitudinal views held by students and communities related to PBC), and Foundational (the theoretical and logical issues that underpin PBC). Since we analyze students’ written work as they produce PBCs, most of the hypotheses our data can address are in the Operational category. This includes three

of our principal hypotheses: Recognition, Lack of Target, and Resource. We will also briefly refer to the Quantifier, Cognitive Demand, and Template Hypotheses. One of our design hypotheses, False World, is in the Affective category, and we rely primarily on our interview data to address it. We do not have evidence specifically addressing the Foundational hypotheses, and a different study design would probably be required to do so. Here, we briefly summarize the hypotheses addressed in this paper, with a few citations to their origins or prior studies.

Recognition Hypothesis

The Recognition Hypothesis posits that students may not recognize contradictions in the setting of mathematical proof because they do not appreciate the strictness of the requirement that some proposition and its negation have both been deduced. Recognizing a contradiction presupposes sufficient facility with mathematical logic, as well as metacognitive monitoring of each deduction reached during the proof in relation to potentially contradictory background knowledge (Chamberlain, 2017). Evidence for this hypothesis may include students' claims to have reached a contradiction when in fact they have not, as well as their overlooking contradictions that were actually reached. We are not aware of previous *empirical* work on this hypothesis.

False World Hypothesis

Each PBC necessarily begins by making an assumption for the sake of argument that will eventually prove to be false. Thus students must engage in counterfactual reasoning, about objects which cannot exist or supposed properties that those objects cannot actually have. Imagining, and reasoning within, such a counterfactual world is likely to be more challenging in the mathematical than the everyday context, because everyday counterfactual assertions are false as a matter of fact, but mathematical ones are *necessarily* false. That is, everyday counterfactuals are merely contrary to the actual state of the world, and could be true if that state were different, but mathematical ones are logically self-contradictory and could never be true. Our Framework subdivides this hypothesis further into Impossible Objects and False Premises, but for the present analysis we do not make this distinction.

Antonini and Mariotti (2006, 2008) showed that some students are unsure of what can be accepted for the sake of argument, and even whether standard logical reasoning can be applied in such "impossible worlds". Baccaglini-Frank et al. (2013) describe students' attempts to make sense of "pseudo-objects," whose assumed properties are impossible in Euclidean geometry, when applying PBC in geometric proofs. Students' prior conceptions of logic and proof may take the form of "Greek axiomatics" (Harel & Sowder, 2007), wherein axioms are self-evident truths and logical deduction preserves truth. In contrast, PBC may require the more flexible "modern axiomatics," in which axioms are arbitrary assumptions and the validity of an argument is independent of the truth of the statements comprising it.

Lack of Target Hypothesis

In a DP, the hypothesis and the conclusion are known at the outset, and the goal is to construct a chain of logical reasoning connecting them. One can work forward from the hypothesis as well as backward from the conclusion and meet somewhere in the middle. In PBC, the “conclusion” of the proof is “a contradiction,” which is initially unknown, so one cannot aim at it or work backward from it. A proposition r forming half of the contradiction “ r and not- r ” may appear unexpectedly at any stage of the proof, and “not- r ” may be the explicit hypothesis of the theorem, a conventional assumption (for example, that some fraction encountered was written in lowest terms), or any piece of background mathematical knowledge (Jourdan & Yevdokimov, 2016). Students are likely to find this lack of direction disorienting. However, we do not know of prior empirical studies exploring this hypothesis.

Resource Hypothesis

The Resource Framework has been used extensively in physics education research to study students’ problem-solving processes (Sabella & Redish, 2007; Tuminaro & Redish, 2007). It posits that problem-solving depends on the activation of cognitive resources (background knowledge, didactical contracts, heuristic strategies) that are assembled as needed from distinct and uncorrelated elements and are activated in a highly context-dependent fashion. As we apply this hypothesis to PBC, it states that how and which resources are activated accounts for more of students’ difficulties with PBC than does the logical distinction between DP and PBC. Dawkins and Karunakaran (2016) make a similar point in suggesting that studies of mathematical proof should not treat it as a stand-alone phenomenon but should contextualize it in the mathematical content area involved (number theory, geometry, analysis, etc.).

On the one hand, the Resource Hypothesis suggests that DP and PBC are not so different in terms of the sources of students’ difficulties, which owe more to how students access and activate resources than to the logical nature of the proof technique. On the other hand, resource activation is highly context-dependent and so similar resources are likely to be used differently in different types of proof and in different content areas. Further research is needed to disentangle these aspects of the Resource Hypothesis.

In addition to the above hypotheses, which informed the design of our study, some of our evidence is relevant to other hypotheses in our Framework, which we sketch here. More thorough discussions can be found in (Quarfoot & Rabin, [in press](#)).

Quantifier Hypothesis

There is abundant evidence that students struggle with quantified statements in mathematical logic, particularly when multiple and/or implicit quantifiers are

present (Dubinsky & Yiparaki, 2000; Selden, 2012; Shipman, 2016). Since PBC requires correctly negating such statements, and recognizing when they are contradictory, this is a plausible explanation for students' difficulties with PBC.

Cognitive Demand Hypothesis

The Cognitive Demand Hypothesis arises from the framework of cognitive load theory (CLT) in the information processing perspective (Centre for Education Statistics and Evaluation, 2017; Sweller, 1988, 1994). All proof tasks impose significant cognitive demands on learners, who find that they must now attend to the logical basis of algebraic manipulations they may have performed automatically in prior mathematics courses. However, there is reason to expect that PBC imposes especially heavy demands. In addition to the burden of counterfactual reasoning (False World Hypothesis), the prover must continually reflect on prior deductions in the proof and compare them with each other and with relevant background knowledge for possible contradictions. This hypothesis can be traced to Leron's (1985) important paper:

The moment the negative assumption is declared, along with the intention of falsifying it by means of a future contradiction, a cognitive strain is set up in the mind of the learner, perhaps because of the difficulty of living in a false world, still operating as if it were real. This cognitive strain grows (linearly?) with the time spent living in this world, i.e. with the distance between the negative assumption and the terminal contradiction. Perhaps the feeling of frustration and incomprehensibility is proportional to the length of the 'negative stretch' of the proof. (p. 324)

We are not aware of empirical studies testing this hypothesis.

Template Hypothesis

Certain types of proof follow characteristic patterns, or templates, that students may learn to recognize and emulate. Examples include proofs of certain summation formulas by mathematical induction, or proofs that two sets are equal by showing that each is a subset of the other. The Template Hypothesis suggests that such structured templates are less prevalent in PBC than in DP, or perhaps are less recognizable by students. While various authors (Antonini & Mariotti, 2008; Tall, 1979; Thompson, 1996) have touched on this idea, it does not appear that any have tried to explore it systematically.

Participants and Methods

The research question investigated in this paper is,

Which student difficulties seen in our data (students' written proof construction work on class homework and exam problems, as well as their reflective

thinking in interviews) can be associated with our four original hypotheses, and others in the HFPBC?

To explore this question, we planned a mixed methods study investigating students' performance, work, and thinking about PBC. We stress that the study was designed around only three hypotheses initially, with the fourth added in the course of our early data analysis. Implications of our data for other hypotheses in our Framework were considered at a later stage.

The participants were students in a ten-week (one quarter) Introduction to Proof course taught by one of the authors at a large public university in the southwestern United States. The class is required for all mathematics majors and is normally taken by sophomores or juniors following the two-year calculus/linear algebra sequence. The course textbook (Chartrand et al., 2015) is the most widely used book for such courses in the United States according to David and Zazkis (2017). There were 106 enrolled students, who were all invited to participate in the study, and 72 agreed to do so. The author of this paper who was not the course instructor made a brief in-class presentation about the study (in the instructor's absence) and asked students to submit consent forms. The majority of participants, who merely agreed to make their written coursework available to us, were not compensated for their participation, but those students who volunteered for interviews (see below) were paid for their time. 48 of the 72 participants were mathematics majors, and the others had various other STEM majors. Although we did not ask for gender identification, it seemed there were roughly equal numbers of male and female participants.

All course homework assignments and exams were graded using Gradescope (www.gradescope.com), which preserved the students' graded coursework for our later analysis. PBC was covered about halfway through the course and roughly a week of class time was spent introducing this technique and illustrating it with example proofs. It was then available as a known proof method in the second half of the course. The justification presented for PBC was based on the formal logic of conditional statements: if we are able to deduce a contradiction C from some statement P then we know that $P \implies C$ is true while C is false, and this means that P must be false. The first example presented was the standard proof that $\sqrt{2}$ is irrational. We selected the homework and exam problems on PBC by collecting a wide variety of such problems from the course textbook, other textbooks, and the prior research literature, along with some of our own design. These included proofs of irrationality of various numbers and a range of other problems in algebra, geometry, and number theory. We decided which problems to assign after thorough discussions of students' likely proof approaches, the difficulty of the problems, and which problems might provide evidence for specific hypotheses. The majority of assigned problems were from the textbook, and all were appropriate for a course at this level. An undergraduate grader scored about half of the problems on each homework assignment, while graduate teaching assistants scored the exam problems. However, our analysis is based on our own examination of student work, not that of these graders.

We analyzed both homework and exam problems, since these data sources are complementary in some respects. Students are under less time pressure when completing homework assignments, so their responses may be more thoughtful and their

Table 1 The PBC problems, in order of assignment. Problem 8 was not graded, but was included in our analysis. Problem 11 was also solved by students using DP. HW = Homework, MID = Midterm Exam, FINAL = Final Exam

1	HW 5	Prove that the product of an irrational number and a nonzero rational number is irrational.
2	HW 5	Prove that the sum of the squares of two odd integers cannot be the square of an integer.
3	HW 5	Show that no positive integers m, n satisfy $m^2 + m + 1 = n^2$.
4	HW 5	Consider the equation $a^2 + 1 = 2^n$, where $a \geq 2$ and $n \geq 1$ are integers. (a) Show that a must be odd. (b) Show that in fact no such integers exist.
5	HW 5	Show that no consecutive positive integers a, b, c satisfy $a^3 + b^3 = c^3$.
6	HW 5	In class we proved that there exist irrational numbers a, b such that a^b is rational. Rewrite this argument in the form of a proof by contradiction.
7	HW 6	Show that there are no positive integers m, n such that $7/17 = 1/m + 1/n$.
8	HW 6	Show that the lines bisecting two angles in a triangle cannot be perpendicular to one another.
9	HW 6	Show that no prime numbers a, b, c satisfy $a^3 + b^3 = c^3$.
10	MID 2	Prove that $\sqrt{2/5}$ is irrational.
11	MID 2	If x, y are positive real numbers, prove that either $\frac{x}{x+2y} \geq \frac{1}{3}$ or $\frac{y}{y+2x} \geq \frac{1}{3}$.
12	HW 8	If p, q are distinct primes, prove that \sqrt{pq} is irrational.
13	HW 8	Show that any positive integer of the form $3k + 2$ must have a prime factor of the same form.
14	FINAL	Prove that $\sqrt{5/7}$ is irrational.

errors less likely to be simply the result of carelessness. On the other hand, students have more opportunities to obtain assistance on homework (from classmates, their instructor, or online sources), while exam solutions are (presumably!) their own work. Table 1 lists all the graded PBC problems, and one additional ungraded homework problem (#8) that we included in our analysis because of its prior appearance in the research literature (see Baccaglioni-Frank et al., 2013). Table 2 provides the other problems that will be discussed in this paper.

All the PBC problems in Table 1 come from the textbook, with the following exceptions. We added Problem 7, from D’Angelo and West (2000), page 137, because it requires little formal mathematical background, can be solved in many ways, and is unusual enough that students are unlikely to know a standard solution method. As noted, Problem 8 has been considered in the research literature before

Table 2 Other problems discussed in this paper

1	HW 7	A relation R is defined on the set $A = \{a + b\sqrt{2} : a, b \in \mathbb{Q}, a + b\sqrt{2} \neq 0\}$ by xRy iff $x/y \in \mathbb{Q}$. Prove that R is an equivalence relation, and describe the distinct equivalence classes.
2	HW 8	A function F is defined from the natural numbers to the set of odd natural numbers by $F(n) = k$, where k is the odd natural number such that $3n + 1 = 2^m k$ for some nonnegative integer m . That is, k is what’s left after factoring out as many 2’s as possible from $3n + 1$. Determine whether F is one-to-one and/or onto.

(Baccaglioni-Frank et al., 2013). Geometry problems are otherwise rare in our course, and this one involves an “impossible object” or pseudo-object. Students’ attempts to draw a diagram and reason about this object might provide evidence for their comfort level with counterfactual thinking, as envisioned by the False World Hypothesis. Problem 10 is a variation on one discussed by Tall (1979). Since it involves the square root of a fraction rather than an integer, it requires some originality compared with standard textbook examples. Unlike the similar Problem 14 from the final exam, it does not require Euclid’s Lemma (which states that if a prime divides the product of two integers, it must divide one of them), as this had not yet been covered in the course. We chose Problem 11, from Vandervelde (2010), page 64, for the midterm exam to provide challenges in correctly negating the statement and working with inequalities. Although we did not anticipate this, it turned out to be one of the few problems that could be solved equally well by DP or by PBC, and indeed about half our students approached it in each way. With this one exception, all the problems that we considered to be PBC problems were indeed approached that way by the vast majority, usually all, of our students. The direct proof problems assigned for homework in the course all came from the textbook and were not specifically matched to the PBC problems in terms of content or difficulty.

Some aspects of students’ challenges in proof construction, particularly affective ones, are unlikely to be evident from their written work. These include their understanding of why PBC works, how they feel about counterfactual reasoning, and how easy or hard they find it to seek and identify contradictions. Therefore, to complement and triangulate our analysis of student work, we also solicited student volunteers for interviews, but only obtained six volunteers, all of whom were accepted. Nevertheless, there were three male and three female interview subjects, spanning a wide range of achievement levels in the course (final course grades A through C). The interviews, which took place just after the second midterm, were semi-structured, stimulated-recall interviews (Schubert & Meredith, 2015). Students were shown their own prior work on certain PBC problems, usually Problems 3, 7, 8, and 11 in Table 1, and sometimes others for which that student’s work was particularly interesting. They were asked to identify the contradiction they had reached and explain why it was a contradiction, how they searched for and then recognized the contradiction, why they chose a particular approach, and what was the hardest part of the problem for them. Sometimes they were shown the work of another student and asked to locate the contradiction or compare this solution with their own. (We did *not* confront students with specific errors they had made, so we do not have data on their explanations for those errors or their reactions to having them pointed out.) After this, they were asked some more general questions, such as what makes a PBC work, how they feel about reasoning from counterfactual assumptions, and whether they prefer PBC or DP for any reason. We had been concerned that students might not recall their recent written work or their thought processes while doing it. Hence, during the interviews we provided each student with a copy of their written work, and on homework assignments, we occasionally inserted questions in which students were asked to reflect on the totality of their solving process (for our use in the interviews or analysis). While students were familiar with their work and did not express uncertainty about their prior thought processes, the reader is reminded that

Table 3 Interview protocol**Questions About Specific Proofs:**

Where is the contradiction in your proof? What makes it a contradiction?
 How did you search for and then recognize the contradiction?
 At what point in working out the proof could you see that a contradiction was going to emerge?
 How did you negate the theorem statement?
 Did you already know, or believe, the result before proving it?
 Did that affect your proving process?
 What was the hardest part of the proof for you?

General Questions About PBC:

How do you begin or approach a PBC?
 What features of a problem indicate to you that PBC is appropriate?
 Do you have strategies for seeking a contradiction?
 Can you anticipate where the contradiction might come from?
 How easy or difficult is it for you to recognize the contradiction when it appears?
 Do you sometimes think you've found a contradiction, and then realize it isn't one?
 How does it feel to work with an assumption that is not true, e.g., that there are only finitely many prime numbers?
 What makes a PBC work?
 Do you find PBC similar or different from DP?
 Do you have any preference between PBC and DP?
 Is either one easier to do or to understand?
 Does either one give you a better sense for why the theorem is true?

interviews have inherent drawbacks. These include verbal overshadowing, the erosion of ecological validity, the malleable nature of memory/self-reporting, and the coercive influence of interviewers (disessa, 2007; Ericsson & Simon, 1993; Ryan & Schooler, 1998). A summary of the interview protocol appears in Table 3.

Students' written work was analyzed in the following way. We read each student's solution to each problem in Table 1 and wrote a brief summary of their solution method, judging its correctness and noting any errors. We attended in particular to whether the proof began with the correct negation of the claim, correct use of quantifiers and logic, the type of contradiction reached and whether it was indeed a valid contradiction, other contradictions the student may have reached but not noticed, unjustified assertions or circular reasoning, any unnecessary steps or assumptions, and the ways that resources were used. We then compared these summaries to find repeated or common themes and errors, which were coded. Some errors or approaches unique to individual students were also noted if they related to our hypotheses.

The interviews were videotaped and subsequently transcribed¹. We then explored these data using Thematic Analysis (Braun & Clarke, 2006). This qualitative tool looks for *patterned* responses *across* data sources (here, different interviews).

¹ We thank Amelia Stone-Johnstone for her assistance in conducting the interviews.

Specifically, we identified themes related to seeking contradictions, (dis)comfort with counterfactual reasoning, (un)certainty about the logical basis of PBC, ability to identify and explain contradictions in interviewees' work and that of other students, and preferences between PBC and DP. Interview results were used to triangulate our analysis of written work when possible. However, our analysis of written work took place after the interviews, and most of the interview data did not directly address the reasoning patterns we found in the written work. Rather, the interview data were mostly complementary, bearing on students' affective reactions to the problems and to PBC generally that were not visible in their written solutions.

Results

Written Work

The aspects of student thinking that are visible in their written work bear most directly on the Recognition and Resource Hypotheses: students' ability to recognize contradictions and how they select and employ background knowledge. In our analysis, we first explore three problems in depth, and then offer additional findings that span multiple data sources. One reason we have chosen to look at three problems in great detail is to address a deficiency in the PBC literature. While conducting our literature review, we found that most articles focused on one or two common PBCs (often, the irrationality of $\sqrt{2}$, or a variant of this). In an effort to add to the resources available to future researchers, we wanted to offer a thorough review of three problems that, to our knowledge, have not been studied previously.

Problem 7, Table 1: Development of the Resource Hypothesis

In our analysis, we encountered several errors reflecting unproductive resources in areas like properties of rational and irrational numbers, factorization, and divisibility. Problem 7 in Table 1 revealed many of these. Students had to prove that no positive integers m and n satisfy the equation $\frac{7}{17} = \frac{1}{m} + \frac{1}{n}$ (that is, $7/17$ cannot be written as a two-term Egyptian fraction). There are multiple approaches to this problem, but only eight of the 64 submitted solutions were correct. Most of these correct solutions were variations on the following (which is not the simplest possible approach): Assuming that there are such integers, write $\frac{7}{17} = \frac{m+n}{mn}$. Since 17 must divide mn , we assume without loss of generality that 17 divides m and thus $m = 17k$ for some positive integer k . Then

$$\frac{7}{17} = \frac{1}{17k} + \frac{1}{n} \leq \frac{1}{17} + \frac{1}{n},$$

implying that $\frac{6}{17} \leq \frac{1}{n}$. Thus, n can only be 1 or 2, neither of which leads to an integer value for m , contradicting the assumption.

The most common (and incorrect) approach was to write $\frac{7}{17} = \frac{m+n}{mn}$, or equivalently $7mn = 17(m+n)$, and conclude that $m+n=7$ and $mn=17$. Students then

verified that no positive integers satisfy both conditions. Twenty-four of the 64 solutions were of this type. We coded this way of thinking based on $\frac{7}{17} = \frac{m+n}{mn}$ as Strong Fraction Equivalence (SFE): the view that equal fractions must have identical numerators and denominators. Drawing the same conclusion from $7mn = 17(m+n)$ was coded as Strong Unique Factorization (SUF): the view that if $ab = cd$ then necessarily a equals c or d , and b equals the other. Neither of these views seems related to the logic of PBC as such, and indeed we observed them in DP problems as well. This and similar observations drew our attention to the Resource Hypothesis, that students' "difficulties" with PBC may reflect the types of mathematical resources they activate while constructing such proofs. Since this problem came from homework, not an exam, students' use of SFE and SUF are unlikely to reflect simple carelessness due to time pressure. On the other hand, our subjects are STEM majors who are unlikely to "really believe" in the truth of SFE/SUF and could presumably provide numerical counterexamples if asked to do so. This is consistent with the resource framework, which postulates that student knowledge is not coherent, highly structured, or "cross-referenced" for consistency, but rather consists of relatively isolated resources that can be activated in response to specific tasks, or cues in those tasks (diSessa, 2013; Hammer et al., 2005).

SFE and SUF are examples of a more general pattern of thinking, or use of resources, that we observed frequently in our data and coded as Appearance Trumps Possibility (ATP). In both SFE and SUF, students attended to what was algebraically visible in a given expression rather than what numerical possibilities might be consistent with it. In doing so, they were able to "progress" to a "solution" using opportunistic logic that they would likely disagree with in another setting. This idea has similarities to Harel and Sowder's (1998) "ritual proof scheme" and Vinner's (1997) "pseudo-conceptual reasoning".

Another example of ATP was provided by a student who derived $m = \frac{17n}{7n-17}$ and claimed the contradiction that an integer cannot equal a fraction. Indeed, the right side appears as an algebraic fraction, but its value could be an integer: it is necessary to rule out the possibility that $7n - 17$ might divide $17n$ for some particular value of n . Another student obtained the same equation and considered separately the possibilities that n might be even or odd. For n odd, the student noted correctly that a fraction of the form odd/even cannot be an integer, but claimed incorrectly for n even that even/odd cannot be an integer either. One student claimed that $7 = 17(\frac{m+n}{mn})$ leads to the contradiction that 17 divides 7, supporting this with the explicit statement that $\frac{m+n}{mn}$ is an integer. Here, the student seems to be using the appearance $7 = 17 \cdot$ (something) to invoke divisibility and is opportunistically hoping/claiming that $\frac{m+n}{mn}$ is an integer. Another asserted that the only possibilities consistent with $m+n = 7k$ are that m and n are k and $6k$, $2k$ and $5k$, or $3k$ and $4k$. We consider this another example of ATP, possibly related to the phenomenon of integer bias (Christou, 2015).

We have listed the variety of solutions in detail to make the point that none of the errors seems directly related to the foundational/logical structure of PBC. Rather, they seem to reflect the activation of inappropriate resources concerning fractions, divisibility, and so forth, many falling under our broad category of ATP. Not all

students presented their solutions formally as PBC, with an explicit negation of the hypothesis at the start. Some may conceptualize their process instead as simply seeking solutions of an equation and finding none, which seems cognitively and affectively different from formal PBC reasoning.

The wide variety of solution methods and the small number of correct solutions suggest that the solutions are the students' own, rather than the result of significant cheating. However, after the class had concluded and our analysis was completed, we learned that the online "tutoring" site Chegg had received three queries about this problem during the week that it was assigned as homework. One of three "expert solutions" posted there was an example of SFE, one claimed the incorrect "integer cannot equal fraction" contradiction, and one incorrectly applied Euclid's Lemma. We have no evidence as to which or how many of our students viewed these solutions. The fact that all three solutions by "experts" (who are likely to be students themselves) incorrectly applied relevant resources could be seen as additional evidence supporting the Resource Hypothesis.

Problem 10, Table 1: Resource Errors and Beyond

Next we analyze Problem 10 in Table 1, from the second midterm, asking for a proof that $\sqrt{2/5}$ is irrational. This is a variation on the $\sqrt{5/8}$ problem used by Tall (1979). At this point in the course, students had done homework problems requiring proofs of irrationality of $\sqrt{3}$ and $\sqrt{2} + \sqrt{3}$. A correct solution, given by 28 of 65 students, assumes that $\sqrt{2/5} = a/b$ for integers a and b having no common factor, so that $5a^2 = 2b^2$, and then deduces sequentially that a^2 , a , b^2 , and b must be even, which is a contradiction. Euclid's Lemma had not been covered yet, so students could not use it to conclude from $5|b^2$ that $5|b$ also.

Students made a variety of errors on this problem. Some used SFE, for example concluding from $\frac{a}{b} = \frac{\sqrt{2}}{\sqrt{5}}$ that $a = \sqrt{2}$ and $b = \sqrt{5}$, which are not integers. Some relied on incorrect claims about irrational numbers, for example that the quotient of two irrational numbers must be irrational. Some simply made assertions that are equivalent to the claim being proved, for example that $b = a\sqrt{5/2}$ which is "clearly" not an integer. Some also applied the concept of divisibility outside the set of integers, for example claiming from $a^2 = (2/5)b^2$ that $2/5$ divides a^2 . Another claimed that if a^2/b^2 is even, then a/b must also be even. All these errors seem to reflect the activation of inappropriate resources (usually via forms of ATP) rather than the structure of PBC as such, again supporting the Resource Hypothesis. Some of these types of errors have been observed before (e.g., Barnard & Tall, 1997).

There were some logical errors that might plausibly be related to student understanding of the PBC technique itself. These include instances of circular reasoning, assuming within the proof that $\sqrt{2/5}$ is irrational, or concluding the proof with, "but this contradicts that $\sqrt{2/5}$ is irrational". While these issues could support the Foundational Hypotheses, there is prior evidence of students' tendency to assume the truth of the statement being proved within the proof itself in cases of DP as well (Stavrou, 2014). Interestingly, one student viewed a restriction on generality to be a contradiction. Having observed that a^2 is even, so that $a^2 = 2k$ for some integer k ,

the student noted that an even square is always a multiple of 4, whereas $2k$ is even but not necessarily a multiple of 4, and claimed this as a contradiction. This might reflect a misunderstanding of what constitutes a contradiction (Recognition Hypothesis), or it might be an instance of ATP (Resource Hypothesis) since in this situation $2k$ will always be a multiple of 4, although only a factor of 2 is algebraically visible in this expression.

Since this problem appeared on an exam, some student errors may reflect the influence of time pressure. In addition, since PBC was a relatively new topic at the time of the exam, students had limited experience with it. Proofs of irrationality of numbers of the form \sqrt{x} are a common and standard subclass of PBC problems in most textbooks, and it is interesting to note that by the time of the final exam, students seemed to have mastered the “script” or template for solving them. A similar problem concerning $\sqrt{5/7}$ on the final exam (Problem 14 in Table 1) was solved correctly, following a standard template, by 47 of 58 students. This is contrary to the Template Hypothesis, and speaks to the importance of including students’ experience levels and time as explanatory variables in studies of PBC. For at least this subclass of PBC problems (simple irrationality proofs), students can and do learn a reliable template for solving them over time.

Problem 11, Table 1: A Mixture of Issues

We analyze one more problem in depth, Problem 11 from Midterm 2 in Table 1. This problem was unique in that it was intended as a PBC problem, but about half the students gave direct proofs, revealing the challenge in attempting to compare DP and PBC. More specifically, 25 of 52 solutions used DP (of which 13 were correct) while 27 used PBC (of which 17 were correct). The intended PBC solution was to assume the negation, that there exist positive real numbers x and y such that $\frac{x}{x+2y} < \frac{1}{3}$ and $\frac{y}{y+2x} < \frac{1}{3}$, and to deduce that $x < y$ and $y < x$, a contradiction. A DP can be given essentially by reversing these steps, starting from the fact that either $x \geq y$ or $y \geq x$ and deriving the target inequalities from these. Some students tried proof by contrapositive, in the form of the claim that if the two target inequalities are not satisfied then x and y are not both positive. However, in manipulating the inequalities (e.g., in clearing denominators), they invariably behaved as if they were multiplying by positive quantities, which is no longer part of the hypothesis. The most common error in the DPs was “backward logic,” starting from the desired conclusion and deriving a true statement from it, without attention to whether the steps were reversible.

Students who used PBC did make some errors that suggest operational issues with the proof technique itself. Many correctly negated the disjunction of inequalities, but omitted any quantifiers, so that it was not clear whether the resulting conjunction was true for all, or (correctly) for only some x and y (Quantifier Hypothesis). Some then exhibited numerical values for x , y that made the negated inequalities false, and claimed this as a contradiction. The idea that a specific true instance of a theorem can be used to contradict its negation in a PBC is interesting, and occurred in other problems also. It is a PBC version of the “proof by example”

that students sometimes propose in DP, and has been reported previously in the literature (Shipman, 2016). One could also view it as a misunderstanding of what constitutes a contradiction (Recognition Hypothesis), but traceable to students' lack of attention to quantifiers when doing algebra (Quantifier Hypothesis).

In addition to these challenges, students' background knowledge about inequalities tended to be weak, since inequalities are typically neglected in high school relative to equations. Common errors include thinking that inequalities may be combined as if they were equations, or multiplying by quantities that need not be positive. Students also may not understand that when two inequalities are combined, the resulting single inequality contains less information than the original pair. This problem activated such resources in both the DP and PBC approaches (Resource Hypothesis).

Additional Observations on Various Hypotheses

We found further instances of SFE, SUF, and, more generally, ATP in many of the problems we examined (Resource Hypothesis). In Problems 5 and 9 from Table 1, which both deal with integer solutions of $a^3 + b^3 = c^3$, it was common for students to write $c^3 = (a + b)(a^2 - ab + b^2)$ and then conclude by SUF that one factor on the right must be c and the other c^2 . There were additional instances of claiming the contradiction that an integer equals a fraction based on a fractional algebraic form, without considering whether the fraction could reduce to an integer in particular cases. Another feature of equations that is usually not algebraically visible is the domain of their variables, and indeed, students were often uncertain about what it meant to say that a number belonged to \mathbb{N} , \mathbb{Q} , or \mathbb{R} , and how the properties of these sets of numbers differ. As we have noted, students sometimes apply the notion of divisibility outside of \mathbb{N} . A student working on Problem 14 from Table 1 wrote $\sqrt{5/7} = a/b$ with integers a, b and then proceeded to consider two cases: either a is rational, or a is irrational (!). Direct proof Problem 1 from Table 2 revealed substantial confusion between the roles of the sets A and \mathbb{Q} in the problem. In checking the reflexive, symmetric, and transitive properties, students would often check that $x/y \in A$ rather than $x/y \in \mathbb{Q}$. That has to do with whether A is closed under division, not whether R is an equivalence relation. Sometimes they would *check* that $x/y \in A$ but *write* that $x/y \in \mathbb{Q}$. One interesting and subtle error was to claim that $x/y \in \mathbb{Q}$ is only possible if $x, y \in \mathbb{Z}$, because rational numbers are *defined* as ratios of integers, after all. These errors partly reveal the great diversity of operational challenges students face when trying to construct DPs and PBCs (Operational Hypotheses).

Other problems also provided support for the Resource Hypothesis. For example, Problem 2 in Table 2 asks whether a certain function $F(n)$ on the natural numbers is one-to-one and/or onto. (It is neither.) The language, "That is, k is what's left after factoring out as many 2's as possible from $3n + 1$ " was included deliberately to clarify the problem for students who might have difficulty interpreting the formal definition alone. A common approach was to formally invert the function definition to obtain $n = (k \cdot 2^m - 1)/3$, from which some students concluded that because it was invertible it was both one-to-one and onto. The two ATP issues these students did not attend to were whether this value for n is in fact a natural number, and the

fact that the function value only gives k and not also m , so one cannot recover n from the function value alone in this way. A possible difference in how such resources are activated in PBC versus DP problems is the fact that students here did *not* make the claim that an integer cannot equal a fraction, but rather hopefully assumed the reverse: that the apparent fraction would reduce to an integer, since that is what a successful solution requires. Thus, it appears that the activation of resources has an *opportunistic quality* similar to confirmation bias in that it is oriented toward supporting the desired conclusion.

There are also examples from our data that support the Negation and Quantifier Hypotheses. As the literature has previously documented (Shipman, 2016), students do have difficulty negating mathematical statements, especially when quantifiers are present or implicit. Some students negated the claim in Problem 1 of Table 1, “The product of an irrational number and a nonzero rational number is irrational,” as “The product of an irrational number and a nonzero rational number is rational”. The implicit quantifiers in the claim are universal ($\forall x \in \overline{\mathbb{Q}}, \forall y \in \mathbb{Q} \setminus \{0\}$), so the quantifiers in its negation should be existential. The negation produced above is ambiguous in this regard, but would normally be interpreted by mathematicians as universally quantified, which would be incorrect. Interestingly, this did not usually affect the students’ subsequent reasoning. Either they interpreted their negation as existentially quantified, or their reasoning was so formal that they never addressed the distinction. This sort of incorrect negation also occurred in Problem 12 of Table 1, in the form of the negation, “if p and q are distinct primes, then \sqrt{pq} is rational.”

Students’ lack of attention to quantifiers caused difficulties also in Problem 13 of Table 1. The PBC would begin with the assumption that some positive integer of the form $3k + 2$ has no prime factors of that form, and hence, the prime factors must be of the form $3k$ or $3k + 1$. In students’ written work, it would often be ambiguous whether all, or only some, of its prime factors must have the form $3k + 1$. In this case, clarity on this distinction is essential to successfully complete the proof.

Interview Findings

Student interviews also provided evidence concerning our original hypotheses. The False World Hypothesis was one of the three principal hypotheses motivating our study, and the interviews explicitly asked about it. Antonini and Mariotti (2008) found that students were uncomfortable working in a false or impossible world and uncertain as to what reasoning could be trusted in that world (False World Hypothesis). In our interviews, students did *not* express this kind of discomfort, either spontaneously or when asked explicitly about it. Indeed, they had a fairly solid understanding that PBC works by deducing logical consequences from provisional assumptions that may turn out to be untrue (contrary to the False World Hypothesis). We hoped that Problem 8 from Table 1 might help us explore this issue since it provides a geometric situation involving a pseudo-object. This problem asks students to show that two angle bisectors in a triangle ABC cannot be perpendicular to one another. Students had a high success rate on this problem (49 of 61 correct) and most included a diagram of the pseudo-object from which

they reasoned with no evident confusion. A few students used a DP approach by calculating the angle between the bisectors and verifying that it is obtuse. We asked about any feelings of confusion or discomfort from reasoning in this counterfactual situation, and one student (Kevin, a pseudonym) told us on the contrary that:

“I think it might just be from experience of knowing that hand-drawn pictures can be inaccurate, and then there are a lot of stuff like optical illusions where some things look perpendicular when they’re not... It wasn’t necessarily to reassure myself that the statement was true, because I knew the statement was true, and it was more so to visualize the relationship of that new angle and how it relates to A , B , and C .”

This seems to show a solid understanding of the role of such diagrams in PBC. Compare the following quote from the student Maria in Antonini and Mariotti (2008), which exemplifies the sort of confusion described by the False World Hypothesis:

“Moreover, so as $ab = 0$ with a different from zero and b different from zero, that is against my common beliefs and I must pretend to be true. I do not know if I can consider that $0/b = 0$. I mean, I do not know what is true and what I pretend to be true.”

Another of our students, Jeff, responding to a general question as to how PBC differs from DP, said,

“I feel like direct proof is from A direct to B , and then PBC is you assume this is the path and you find something in between the path, something is wrong, so this cannot be the path.” (The student drew an arrow from location A to location B with an obstacle in between.)

Based on these and similar comments, our students seemed to understand the logical basis of PBC rather well, and did not exhibit the sort of disorientation or discomfort that the False World Hypothesis would predict.

We had hoped that asking whether students believed a claim to be true before attempting to prove it via PBC would provide evidence as to whether they were more troubled by counterfactual assumptions when they already knew them to be false than when they were genuinely uncertain or uncommitted. However, their response was generally that they “believed” all claims proposed for proof by their instructor or textbook, since a proof would not be asked for unless the claim was true. We could not draw a useful inference from this.

The interviews also provided an opportunity to explore other hypotheses. For example, when asked to express a preference between PBC and DP, Max commented:

“Direct proof is much easier for me because always for me [when] the question can use the direct proof, I can just see the relationship directly from the question, but if I cannot figure out such a relationship then I will just try to use the indirect proof.”

That is, DP is indicated when the student can see a relationship between the hypothesis and the conclusion, a path from one to the other. If not, then one can use PBC and hope to obtain a contradiction. Sarah agreed:

“... usually I would use DP if I can know in advance how I would prove it. If I cannot see clearly how I would go through my whole proof – which direction I should go – maybe I would choose contradiction.”

However, Carol seemed to prefer the less constrained format of PBC:

“It’s because I know that I’m setting up the statement into like a negation, or setting it up as a contrapositive, or setting it up as induction, that there’s kind of like these steps to these problems. So if I know what the steps are I can do it, but then with direct proofs I feel like sometimes there’s only one way, and if I don’t find that one way I’m just at a loss.”

As mentioned earlier, for certain classes of PBC problems, such as proofs of irrationality of square roots, there was a fairly clear template that students had largely mastered by the end of the course (Template Hypothesis). Some of our interview subjects also had clear ideas as to what might signal the use of PBC and what sorts of contradictions (contrary to the Lack of Target Hypothesis) one should be looking for in a PBC. As for signals, Max offered this thought:

“Sometimes the question is like ‘prove that there is no blah-blah-blah’: so I think this one is the most obvious one for me to use proof by contradiction because I can just negate that in the way that there is (sic) some actual numbers that can give that relationship. So I will just directly use proof by contradiction.”

As for the conclusion of a PBC, two students offered these thoughts on the common contradictions one might expect to see:

“Common contradictions? You mean odd equal to even? Something small is equal to something big. You assume some number is a natural number or integer and then it turns out that it’s not an integer, or you assume something is rational and it turns out to be irrational.” (Jeff)

“Yeah, so a lot of times contradictions would be if you are looking for an integer and you get a fraction, or you are looking, I guess, in another case for an integer [and] you get a number between two numbers that are right next to each other, 15 and 16, and you know that there are no integers between 15 and 16. Or, ... if you are supposed to get an odd number and you get an even number. Or if you are supposed to get a prime number and you get a number that is divisible by 2.” (Kevin)

We also saw some evidence for the Cognitive Load Hypothesis. When discussing the challenges of PBC, Sarah noted that:

“[C]ontradiction requires you to keep looking back at what you think is right, and what you want to contradict, so sometimes ... is harder [than direct proof], and the proof is long when you do the PBC ...”

Max said similarly that

“...when I do the proof step by step because the step is so long and when I write to the end of the question I will have forgot what I assumed before and how this can contradict to that.”

Here, it appears that both students are aware of the burden of continually scanning one's work for potential contradictions and the added length of a PBC as challenging aspects of this form of proof.

Discussion and Future Directions

Our mixed methods study was designed to obtain evidence bearing on three major hypotheses about the origins of student difficulties in producing PBC. The data came from interviews and analysis of student work in an Introduction to Proof course. We extended our analysis to include a fourth hypothesis supported by our data, and have also tried to draw conclusions about some of the other hypotheses in the literature, as summarized in our HFPBC. Here, we review our findings and offer recommendations for the research community moving forward, as well as pedagogical suggestions.

The Resource Hypothesis

By far, the clearest evidence in our study for students' difficulties with PBC was tied to the Resource Hypothesis. The influence of students' background knowledge in particular areas of mathematics (e.g., divisibility, factorization, properties of common number domains) and the idiosyncratic way this knowledge was drawn upon dominated our analyses of student errors. Indeed, such errors were so common that we developed the labels Appearance Trumps Possibility, Strong Fraction Equivalence, and Strong Unique Factorization to help categorize them.

On one hand, the predominance of resource issues in our data from both PBC and DP indicates that many student difficulties have a common source independent of the proof method. On the other, the opportunistic quality of students' use of resources shows that resources are deployed differently when pursuing different goals. Further research should explore both facets of this hypothesis.

The Recognition Hypothesis

The Recognition Hypothesis posits that students might fail to recognize a contradiction, or mistakenly claim to have reached one, because they don't appreciate the strictness of the requirement to deduce a pair of propositions of the forms P and $\text{not-}P$. We saw clear evidence that students incorrectly claim to have reached contradictions, but such claims were almost always traceable to specific resource or operational issues rather than to the logical notion of contradiction itself. The

resource issues often fell under our codes of SFE, SUF, and more generally ATP. Students attended to the formal appearance of an algebraic expression rather than its meaning in terms of the numerical values of the variables that might satisfy it. The most prevalent operational issue was lack of attention to quantifiers, especially implicit ones, which often resulted in incorrect negations and, therefore, incorrect identification of contradictory propositions.

The False World Hypothesis

We did not find evidence supporting the False World Hypothesis. When questioned about the possibility of pseudo-objects, students rationalized such objects using the imprecise, sketch-like nature of pictures. Furthermore, none of the interviewees expressed confusion or discomfort about reasoning from a false assumption, either spontaneously or when asked about it directly. Our findings do not necessarily contradict those of authors like Antonini and Mariotti (2008), whose student populations differed from ours, but they point to the importance of contextual factors such as school grade level, experience with PBC, and so forth.

The Lack of Target Hypothesis

We also did not find strong evidence for the Lack of Target Hypothesis. By the end of the class some students had formed clear expectations of the most likely contradictions to seek in common PBC problems, and pursued them confidently. Although some did experience a lack of direction while carrying out a PBC, others appreciated the flexibility of PBC relative to DP, the latter feeling overly constrained with only a single correct approach.

Other Hypotheses

Through the interviews and the progression of student work over the quarter, we found little support for the Template Hypothesis. Indeed, by the end of the quarter, students seemed to have mastered the most common PBC problem types (e.g., proving a number is irrational, proving there are no integers with some property). Furthermore, they could articulate the types of prompts that would suggest using PBC and the types of contradictions they might arrive at by the end of the proof.

Our data confirm that students incorrectly negate quantified statements (Quantifier Hypothesis), especially when the quantifiers are implicit. More generally, though, they simply do not attend to the quantification of the variables in their algebraic expressions, so that it is unclear whether these variables have specific values or denote arbitrary elements of some set. Sometimes a proof can be completed formally without clarifying this, and in other cases it leads to an incorrect claim of reaching a contradiction, thus contributing to the Recognition Hypothesis.

Two students we interviewed did allude to the cognitive demand of continually scanning one's work for potential contradictions (Cognitive Demand Hypothesis) but this evidence by itself is only suggestive.

Advice for Researchers

Our study did not directly address the relative difficulty for students of PBC versus DP. However, the common claim in the literature that PBC is more difficult lacks clear quantitative supporting evidence. If this is a claim about proof construction, one presumably needs to compare student performance on a set of PBC and DP problems having “equal intrinsic difficulty”, since few individual proofs are equally approachable using both techniques. Designing such a set of tasks and an equitable scoring rubric for them seems quite challenging.

Even if PBC is not *more* difficult than DP, to explore why it *is* challenging, larger and better-designed studies are needed to focus specifically on *one or two hypotheses at a time*; hopefully the HFPBC will be of use in narrowing researchers' foci. In addition, researchers must strive to move beyond the small, canonical set of PBC problems (“Prove $\sqrt{2}$ is irrational”). Care is required in designing tasks/questions, for as Brown (2018) found, task features that are most salient to students may not be those intended or even anticipated by researchers. We believe that researchers must *carefully define and operationalize* vague language like “more difficult than DP”. As the structure of the HFPBC suggests, researchers might care about operational (Can I produce a PBC?), affective (What psychological forces are at play as I produce a PBC?), or foundational issues (What logical issues support a PBC?). Finally, care must be taken not to generalize findings beyond the specific tasks and subject populations studied. One finding might be true for pre-service teachers in one country, but not another. Similarly, care must be taken to consider and report the age and experience level of students. Finding that students have trouble knowing when to deploy PBC is expected if students have only recently been introduced to the technique, but would be surprising for senior undergraduate math majors.

Another issue to consider is which types of data can best reveal the importance of which hypotheses. For example, by exploring student work over the course of an entire term, we were better able to test the Template Hypothesis; students grew comfortable with PBC given enough time and practice. Longitudinal data might also be useful when exploring the Acceptability Hypothesis (the idea that PBC is somehow less acceptable or palatable than DP). In student interviews, some favored DP simply because they were introduced to it first and had more exposure. Such views might erode after using PBC for several years, or seeing examples of PBCs in later courses that feel quite natural. In contrast, we found that interviews were the best tool for exploring the Affective and Foundational Hypotheses because the possible influence of these factors was not apparent in students' work. In addition, we often needed to ask several questions and follow-ups to get a clear idea of how students were thinking about PBC. We did not directly confront students with errors they had made, which limited our ability to explore their justification of resources they used. Having students construct proofs as a group, or explain their reasoning to a peer, might make this more visible.

Our study did not touch on the meta-theoretical issues underlying the validity of the PBC technique (Meta-theoretical Hypothesis). If these are genuine sources of student difficulties in constructing such proofs, it would be profitable to contrast them with similar issues arising in (direct) proof by mathematical induction (Brown, 2003). There too, the validity of the proof technique is not self-evident but depends on a possibly unfamiliar logical or axiomatic foundation, in this case the well-ordering property of the natural numbers or the notion of inductive sets.

We would urge that more attention be paid to the Resource Hypothesis as an origin of student difficulties with proof. We saw common issues in the activation of resources in both DP and PBC when their content areas were similar, suggesting that student behavior in the two proof domains is not so dissimilar after all. On the other hand, the opportunistic quality of resource activation leads to differences in the two contexts. It is quite possible that the goal of deriving a contradiction leads to characteristic differences in how resources are accessed and deployed. Both directions should be explored further to achieve a more nuanced understanding of PBC. Students bring strategies, understandings, and incomplete knowledge from their earlier mathematical experiences to their encounters with proof that shape their learning of it.

Finally, there may be synergies or interaction effects between multiple hypotheses that could be revealed by suitable research designs. For example, Leron (1985) suggested that cognitive load might increase linearly with the duration of the search for a contradiction in a PBC. One could imagine that this would lead to increasing affective discomfort living in the corresponding false world, and an increasing tendency to activate inappropriate resources in a more urgent search for a contradiction. Similarly, the opportunistic way that resources are deployed in PBC and DP suggests that the Resource Hypothesis must interact with other complementary hypotheses in interesting ways that should be explored.

Advice for Teachers

The first decision for teachers presenting PBC is how the technique will be introduced to students: what will be the first example proof and how will the technique's validity be justified? Will PBC be justified as a technique via truth tables, formal logic, links with counterfactual thinking from everyday life, or some combination of these ideas? Will the first example be the proof that $\sqrt{2}$, or some other number, is irrational? Teachers can invite discussion of the lack of conviction that PBC may provide for students, and the "false world" disorientation that it may entail. Students should encounter a wide variety of PBC problems not limited to proofs of irrationality or nonexistence. They need opportunities to reflect on the patterns that recur in PBC proofs. What are the common contradictions? What signals suggest its use? Where are the hidden quantifiers and how does this affect the argument?

Our results could support increased pedagogical attention to three areas: quantifiers and logic, the use of resources, and the development of students' understanding of PBC over time.

Quantifiers

When learning to operate in any new cognitive or physical space, learners need a framing that tells them what features need to be attended to. Many students have very limited framings for proof. Before learning to work with quantifiers appropriately, they need to become aware that the notion of quantification exists and requires attention. Apart from making some distinction between identities (to prove) and equations (to solve), the importance of quantifiers in mathematics is underemphasized in high school. Students' attention should be drawn to their importance in formulating claims precisely, and the language that expresses them (often implicitly) such as *any*, *every*, *each*, *some*, *all*, *no*, *an*, and *unique*. Mathematical claims should be interpreted as general even when the literal statement may seem not to be, as in *A square is a rectangle* or *A prime greater than 2 is odd*.

In addition, the roles of examples and counterexamples to claims should be explored. Educators should help students make the transition between formally defined statements ($\forall x, y \in \mathbb{Q}, x + y \in \mathbb{Q}$) and language-based statements ("the sum of rationals is rational"). This is particularly important when the language-based version hides the nature of the quantifiers. The negation of a claim can be understood as a description of the set of all (potential) counterexamples even when there are none (Dawkins, 2017; Yopp, 2017). Students should negate many everyday and mathematical quantified claims and understand why the quantifiers "flip" between universal and existential.

Finally, attention should be given to how variables are used to prove quantified claims (what is called universal or existential instantiation in logic): when does x represent a specific number, when is it an arbitrary element of some set, when is it a hypothetical solution to an equation that may or may not actually exist, and how does this influence the reasoning? Interpreting the meaning of an algebraic calculation that ends with $0 = 1$ is an excellent warm-up for PBC and an indication that algebra is a mode of deduction rather than computation. Attempting to solve an equation that, in fact, has no solution can introduce the idea of counterfactual reasoning. The ordering of multiple quantifiers has been established in the literature as a source of students' difficulties, but this level of complexity was absent from most of our tasks. Once the fundamentals are in place, it can be discussed in terms of the implied functional dependence allowed between the quantified variables.

Resources

In an Introduction to Proof course based on an explicit set of axioms (for the real numbers, or in group theory, say) the instructor can limit the allowed resources to the axioms themselves and previously proved statements. When the course is more wide-ranging or less formal, like ours, students will access a wide variety of resources as they search for contradictions, sometimes inappropriately. A critical, even skeptical, attitude toward resources employed should be encouraged. Claims that seem plausible may not be true.

Many of the inappropriate uses of resources in our data fell under our code Appearance Trumps Possibility, in which students attend to the form of an algebraic expression rather than to the numerical values that could satisfy it. We saw claims

that an algebraic fraction (e.g., a rational function) could not equal an integer (e.g., a linear function) even though this can certainly happen for specific values of the variable. Students need opportunities to confront counterexamples to such claims. Similar incorrect claims would include that an expression that is not an algebraic perfect square cannot be a numerical perfect square for any value of its variable, or that an algebraic factorization of an expression gives the only possible numerical factorization for any value of its variable (our code SUF). We also saw that students may not realize the importance of attending to the domain of a variable, which is not directly represented in the appearance of algebraic expressions. This is another framing issue that may be new to them. In addition to the formal properties of the domains of natural numbers, integers, rational and real numbers, they need to be aware that concepts like divisibility, smallest element (of a set), and closure under some operations apply to only some of the domains.

Time

Although the initial presentation of PBC is important, this is only an entry point, and students need time to build a schema for this type of proof. A sense of when PBC is the indicated proof method, what type of contradiction might be the goal of a particular proof, or what assumptions might be important (for example, assuming fractions are expressed in lowest terms) takes time to develop. The treatment of PBC should not be rushed or limited to a small number of examples, and reflection on the technique should be encouraged over time. Instructors often assign tasks in the hope that students will notice the patterns that appear within and between these tasks, but this is not automatic for students and requires that their attention be explicitly directed toward those patterns. Expectations for students should be linked to the amount of experience they have had.

Finally, we would encourage educators to approach proof with a greater sense of balance. It is rarely the case that a particular theorem can *only* be proved by PBC (or via DP), or that a particular PBC argument (e.g., the typical textbook proof of the irrationality of $\sqrt{2}$) is uniquely the “most elegant argument” or a “proof from The Book”, a phrase inspired by Erdős, who claimed that God maintained a record of the “perfect” proof for each theorem. Indeed, elegant direct arguments exist for the irrationality of $\sqrt{2}$ (Goodstein, 1948; Square root of 2 n.d.; Direct proof of irrationality? n.d.). Also, it might be helpful for students to *frequently* see both direct and indirect arguments for problems. These side-by-side comparisons allow them to appreciate the value added by each approach. Indeed, in the typical PBC of the infinitude of primes, Leron (1985) noted that the beautiful idea is to take a finite list of primes and build a new number having a prime factor not in this list. This idea can easily be framed in a more direct way that maintains the elegance, hence decoupling the elegance and the need for using a PBC.

Declarations

Conflicts of Interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Antonini, S., & Mariotti, M. A. (2006). Reasoning in an absurd world: Difficulties with proof by contradiction. In J. Novotná, H. Moraová, M. Krátká, & N. Stehlíková (Eds) *Proceedings of the 30th PME Conference*, 2, 65–72, Prague, Czech Republic.
- Antonini, S., & Mariotti, M. A. (2008). Indirect proof: What is specific to this way of proving? *ZDM Mathematics Education*, 40(3), 401–412.
- Baccaglini-Frank, A., Antonini, S., Leung, A., & Mariotti, M. A. (2013). Reasoning by contradiction in dynamic geometry. *PNA*, 7(2), 63–73.
- Barnard, T., & Tall, D. (1997). Cognitive units, connections and mathematical proof. In E. Pehkonen (Ed.) *Proceedings of PME*, 21(2), 41–48, Lahti, Helsinki.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101.
- Brown, S. (2003). The evolution of students' understanding of mathematical induction: a teaching experiment. (Doctoral dissertation). Available from ProQuest database. (3090458).
- Brown, S. (2012). Making jumps: An exploration of students' difficulties interpreting indirect proofs. In S. Brown, S. Larsen, K. Marrongelle, and M. Oehrtmann, (Eds.), *Proceedings of the 15th Annual Conference on Research in Undergraduate Mathematics Education*, 2, 7–16.
- Brown, S. (2013). Partial unpacking and indirect proofs: A study of students' productive use of the symbolic proof scheme. In *Proceedings of the 16th Annual Conference on Research in Undergraduate Mathematics Education*, 2, 47–54, Denver, CO.
- Brown, S. (2018). Are indirect proofs less convincing? A study of students' comparative assessments. *The Journal of Mathematical Behavior*, 49, 1–23.
- Centre for Education Statistics and Evaluation. Cognitive load theory: Research that teachers really need to understand. Technical report, Sydney NSW, Australia, 2017.
- Chamberlain Jr, D. J. (2017). *Investigating the development of proof comprehension: The case of proof by contradiction*. PhD thesis.
- Chartrand, G., Polimeni, A. D., & Zhang, P. (2015). *Mathematical proofs: a transition to advanced mathematics*. Pearson, 3rd edition.
- Christou, K. P. (2015). Natural number bias in operations with missing numbers. *ZDM Mathematics Education*, 47, 747–758.
- D'Angelo, J. P., & West, D. B. (2000). *Mathematical thinking: Problem-solving and proofs*. Prentice Hall, 2nd edition.
- David, E. J., & Zazkis, D. (2017). Characterizing the nature of introduction to proof courses: A survey of R1 and R2 institutions across the U.S. In A. Weinberg, C. Rasmussen, J. M. Rabin, M. Wawro, and S. Brown, (Eds.), *Proceedings of the conference on research in undergraduate mathematics education*. MAA.
- Dawkins, P. C. (2017). On the importance of set-based meanings for categories and connectives in mathematical logic. *IJRUME*, 3(3), 496–522.
- Dawkins, P. C., & Karunakaran, S. S. (2016). Why research on proof-oriented mathematical behavior should attend to the role of particular mathematical content. *The Journal of Mathematical Behavior*, 44, 65–75.
- de Villiers, M. (1990). The role and function of proof in mathematics. *Pythagoras*, 24, 17–24.
- Direct proof of irrationality? <https://mathoverflow.net/questions/32011/direct-proof-of-irrationality>. Accessed Jul 19 2019.
- disessa, A. A. (2007). An interactional analysis of clinical interviewing. *Cognition and Instruction*, 25(4), 523–565.

- diSessa, A. A. (2013). A bird's-eye view of the "pieces" vs "coherence" controversy (from the "pieces" side of the fence). In S. Vosniadou, (Ed.) *International handbook of research on conceptual change*, volume 2nd, 31–48. Routledge, New York & London.
- Dubinsky, E., & Yiparaki, O. (2000). On student understanding of AE and EA quantification. In E. Dubinsky, A. Schoenfeld, and J. Kaput, (Eds.) *Research in collegiate mathematics education, IV*, CBMS Series Issues in Mathematics Education, 239–289. American Mathematical Society, Providence, RI.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. MA, rev. edition: MIT press, Cambridge.
- Goldenberg, P., & Mason, J. (2008). Shedding light on and with example spaces. *Educational Studies in Mathematics*, 69(2), 183–194.
- Goodstein, R. L. (1948). Proof by reductio ad absurdum. *The Mathematical Gazette*, 32, 198–204.
- Hammer, D., Elby, A., Scherr, R. E., & Redish, E. F. (2005). Resources, framing, and transfer. In J. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 89–119). Greenwich, CT: Information Age Publishing.
- Hanna, G., & de Villiers, M. (2012). *Proof and proving in mathematics education: The 19th ICMI study*, 15. Springer Science & Business Media.
- Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A. H. Schoenfeld, J. Kaput, and E. Dubinsky, (Eds.) *Research in collegiate mathematics. III*, 7, 234–283. Providence, RI: American Mathematical Society and Washington, DC: Mathematical Association of America.
- Harel, G., & Sowder, L. (2007). Toward comprehensive perspectives on the learning and teaching of proof. In F. K. Lester, (Ed.), *Second handbook of research on mathematics teaching and learning*, 805–842. Information Age Publishing.
- Jourdan, N., & Yevdokimov, O. (2016). On the analysis of indirect proofs: Contradiction and contraposition. *Australian Senior Mathematics Journal*, 30(1), 55–64.
- Leron, U. (1985). A direct approach to indirect proofs. *Educational Studies in Mathematics*, 16(3), 321–325.
- Quarfoot, D., & Rabin, J. M. (in press). A hypothesis framework for students' difficulties with proof by contradiction. IJRUME.
- Reid, D., & Dobbin, J. (1998). Why is proof by contradiction difficult? In A. Olivier and K. Newstead, editors, *Proceedings of the 22th Conference of the International Group for the Psychology of Mathematics Education*, 4, 41–48, Stellenbosch, South Africa.
- Ryan, R. S., & Schooler, J. W. (1998). Whom do words hurt? Individual differences in susceptibility to verbal overshadowing. *Applied Cognitive Psychology*, 12(7), S105–S125.
- Sabella, M. S., & Redish, E. F. (2007). Knowledge organization and activation in physics problem solving. *American Journal of Physics*, 75(11), 1017–1029.
- Schubert, C. W., & Meredith, D. C. (2015). Stimulated recall interviews for describing pragmatic epistemology. *Physical Review Special Topics-Physics Education Research*, 11(2), 020138.
- Selden, A. (2012). Transitions and proof and proving at tertiary level. In G. Hanna & M. de Villiers (Eds.), *Proof and Proving in Mathematics Education: The 19th ICMI Study* (pp. 391–420). Netherlands, Dordrecht: Springer.
- Shipman, B. A. (2016). Subtleties of hidden quantifiers in implication. *Teaching Mathematics and Its Applications: International Journal of the IMA*, 35(1), 41–49.
- Square root of 2. https://en.wikipedia.org/wiki/Square_root_of_2. Accessed Jul 19 2019.
- Stavrou, S. G. (2014). Common errors and misconceptions in mathematics proving by education undergraduates. *Issues in the Undergraduate Mathematical Preparation of School Teachers: The Journal*, 1.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4(4), 295–312.
- Tall, D. (1979). Cognitive aspects of proof, with special reference to the irrationality of $\sqrt{2}$. In *Proceedings of the Third International Conference for the Psychology of Mathematics Education*, 203–205, Warwick, UK.
- Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12(2), 151–169.
- Thompson, D. R. (1996). Learning and teaching indirect proof. *The Mathematics Teacher*, 89(6), 474–482.

- Tuminaro, J., & Redish, E. F. (2007). Elements of a cognitive model of physics problem solving: Epistemic games. *Physical Review Special Topics-Physics Education Research*, 3(2), 020101.
- Vandervelde, S. (2010). *Bridge to higher mathematics*. Lulu.com.
- Vinner, S. (1997). The pseudo-conceptual and the pseudo-analytical thought processes in mathematics learning. *Educational Studies in Mathematics*, 34(2), 97–129.
- Yopp, D. A. (2017). Eliminating counterexamples: A Grade 8 student's learning trajectory for contrapositive proving. *The Journal of Mathematical Behavior*, 45, 150–166.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.