# The Crucial Role of Proof: A Classical Defense Against Mathematical Empiricism

by

Catherine Allen Womack

Submitted to the Department of Linguistics and Philosophy in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1993



Author ..........................................................
Department of Linguistics and Philosophy
March 10, 1993

Certified by ..........................................................
James Higginbotham
Professor
Thesis Supervisor

Read by ..........................................................
George Boolos
Professor
Thesis Reader

Accepted by ..........................................................
George Boolos
Chairman, Departmental Committee on Graduate Students

# The Crucial Role of Proof: A Classical Defense Against Mathematical Empiricism

by

Catherine Allen Womack

Submitted to the Department of Linguistics and Philosophy
on March 10, 1993, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

## Abstract

Mathematical knowledge seems to enjoy special status not accorded to scientific knowledge: it is considered a priori and necessary. We attribute this status to mathematics largely because of the way we come to know it—through following proofs. Mathematics has come under attack from sceptics who reject the idea that mathematical knowledge is a priori. Many sceptics consider it to be a posteriori knowledge, subject to possible empirical refutation. In a series of three papers I defend the a priori status of mathematical knowledge by showing that rigorous methods of proof are sufficient to convey a priori knowledge of the theorem proved.

My first paper addresses Philip Kitcher's argument in his book *The Nature of Mathematical Knowledge* that mathematics is empirical. Kitcher develops a view of a priori knowledge according to which mathematics is not a priori. I show that his requirements for knowledge in general as well as a priori knowledge in particular are far too strong. On Kitcher's view, some correct proofs may not even convey *knowledge*, much less a priori knowledge. This consequence suggests that Kitcher's conception of the a priori does not respond to properties of mathematics that have been responsible for the view that it is non-empirical.

In my second paper I examine Imre Lakatos' fallibilism in the philosophy of mathematics. Lakatos argued that some mathematical propositions are subject to what he calls "refutations", by which he means to include falsification on extra-logical grounds. Lakatos cites Kalmar's scepticism about Church's Thesis as a case in point. I examine this case in detail, concluding that the failure of Lakatos' thesis in this prima facie favorable case casts doubt upon the thesis generally.

My third paper is a defense of the classical conception of proof against Thomas Tymoczko's thesis that only arguments that are surveyable by us can count as proofs. Tymoczko concluded from his thesis that the computer-assisted proof of the Four

Color Theorem involves an extension of the concept of proof hitherto available in mathematics. The classical conception regards the computer-assisted proof as a real proof, which we are unable to survey. Tymoczko recognizes that formalizability is a criterion for whether an argument is a proof, but he does not, in published work, note that formalizability and surveyability are often conflicting ideals. The classical theory recognizes both ideals because it regards the question whether something *is* a proof as distinct from the question of whether we can recognize it as such, or how confident we can be that it is one.

Thesis Supervisor: James Higginbotham
Title: Professor

# Acknowledgments

Completing this dissertation required only slightly less personnel, strategic planning and financial and emotional resources than the Allied landing at Normandy. Given this, it is unsurprising that I have many people to thank.

First of all, I want to thank my advisors Jim Higginbotham and George Boolos for providing a challenging and rigorous intellectual environment in which I could witness active philosophical work and engage in asking the tough questions of analytic philosophy. George's obvious joy in his work made me enthusiastic about philosophy of mathematics right from the start of my tenure at MIT; Jim's sense of humor helped make the monumental task of writing a thesis seem just a bit lighter.

Hilary Putnam encouraged me and gave me positive feedback on my ideas when I was feeling most discouraged. I feel priviledged to know him and have benefitted greatly from his brilliance and his kindness. Our weekly chats while I was his TA at Harvard helped motivate me to think seriously and work hard; I felt rejuvenated and excited about philosophy after talking with him.

Sylvain Bromberger has always been supportive, funny, sympathetic, stern when necessary, and right much of the time. I have been able to count on him when I needed an honest opinion.

I would be remiss not to thank the people who were responsible for helping get me into this business in the first place. The philosophy department at the University of South Carolina nurtured and guided me, warned me about the perils of going into professional philosophy, and then helped me prepare for it. A few people deserve special mention: Barry Loewer, who now teaches at Rutgers, was my mentor and is still my friend. Bob Mulvaney remains one of my role models and is the best teacher I have ever had. Davis Baird introduced me to works in philosophy that later turned into part of my dissertation; years later he invited me to give a talk on my work which helped focus my ideas. Ferdy Schoeman combined a dedication to the philosophical life with social activism and was a role model for everyone who knew him. I saw Ferdy two weeks before he died of leukemia in June of 1992; his last words to me were, "We are so proud of you, Catherine—we think you have a very promising future philosophically and otherwise".

Crucial though it may be, intellectual inquiry does not always pay the bills. I wish to thank Gary Dryfoos of IS/CSS (neé Project Athena) for hiring me as a minicourse instructor despite the fact that, at the time, my relationship with computers was fraught with fear and ignorance. Amazingly, Carla Fermann also gave me a job, this time as a consultant. They, along with Jeanne Cavanaugh, Tawney Wray and others, have been helpful, supportive, understanding and most indulgent of me.

Formatting this dissertation would have been an onerous task had it not been for the expertise of Jeff Tang; Thanks, Jeff, for helping make my bibliography the fanciest one on the block.

I also worked at Bentley College, and the faculty and staff there have been personally and philosophically helpful. Special thanks go to Michael Hoffman, Bob Frederick, and Sally Lydon.

While pursuing the degree that would mark the beginning of my life's work, I found (quite by accident) another activity that has become a source of creative satisfaction and income—yes, I am talking about tap dancing. I want to thank the folks at the Leon Collins Tap Dance Studio for being a second family to me; they accepted me regardless of how my thesis work was going. They tolerated (or ignored) my constant complaining, and taught me to express myself artistically. In particular I want to thank Josh Hilberman, Pam Raff, Julia Boynton, Sue Ronson, and Dianne Walker, who have taught me much about music and myself. I made many friends there who have been fun and supportive, among them Linda Pompura, Pat Merritt, Rose Giovanetti, Eve Agush, Josh Hlberman, and Charlie Borden.

My friends have been my surrogate family for a long time. I could not have completed this dissertation without them. I want to thank Norah Mulvaney for being there for me year in and year out, always honest and loving. Marin Farach was my thesis enforcer, a thankless job that only a long-time friend would be willing to undertake. Thanks for your support and encouragement, Martin; I look forward to working and playing together for years to come. Thanks to Mike Wolfson for being generous with his time and his car (I will always have a soft spot for that red Honda). His firm conviction that everything would turn out right kept me going. Deborah Savage has been loyal, reliable, a great roomate (even though I will never figure out just how she managed to blacken that stainless steel bowl of mine), and a wonderful friend- thanks, Deb. Eric Chivian needs his own category, but I will just say thanks to him for seeing me through such a difficult process.

Since all of my friends, both casual and close, have borne the burden of helping Catherine get her thesis done, I will just say a big thanks to them all, and know that I owe a serious karmic debt to the world.

I left my biggest debt of gratitude to the end. My family has never lost faith that I would succeed (even when I did), and has backed up that faith with support, love, constant reassurance, and a substantial amount of currency. Thanks and I love you Mom, Dad, Nanny, Papa, Elizabeth, John, Clare, Billy, Cathy, Evans, Pat, Winifred, Will, Sam, Pierce, and Xina. And that is not even my extended family!

That leaves one task to dispatch: I am dedicating my dissertation to my sister Elizabeth. She is my best friend, my constant ally, a fair critic, and source of fun, silliness, and joy. I look forward to sharing successes, failures, joys and disappointments with you for the rest of our lives. I dedicate this to you with love and respect.

# Contents

**1 Is Mathematical Knowledge A Priori: Responses to Kitcher's Skepticism** **11**

1.1 Introduction . . . . . . 11

1.1.1 Epistemological Questions . . . . . . 13

1.1.2 A Preliminary Psychologistic Account of Knowledge . . . . . . 16

1.1.3 A Psychologistic Account of A Priori Knowledge . . . . . . 17

1.1.4 Kitcher's Account of A Priori Warrants . . . . . . 21

1.1.5 Kitcher's Challenge and the Apriorist Response . . . . . . 22

1.2 Kitcher's Attack on Mathematical Apriorism . . . . . . 23

1.2.1 The Role of Proofs in Mathematical Knowledge . . . . . . 23

1.2.2 Kitcher's Account of Proof . . . . . . 29

1.2.3 More Challenges to A Priori Knowledge . . . . . . 36

1.3 A Case Against Kitcher's Views on Challenges to Knowledge . . . . . . 41

1.3.1 Introduction . . . . . . 41

1.3.2 The Study . . . . . . 42

1.3.3 Applying Asch's Study to Kitcher . . . . . . 47

1.4 Final Comments . . . . . . 49

**2 Church's Thesis: a Case Study for Lakatos' Philosophy of Mathematics** **54**

2.1 Introduction . . . . . . 54

2.2 Lakatos' View of Mathematics as Fallible . . . . . . 57

2.2.1 Some Preliminaries: Terminology, Taxonomy . . . . . . 57

2.2.2 Mathematics is Quasi-Empiricist . . . . . . 63

2.2.3 Fallibilism as a Philosophy of Mathematics . . . . . . 72

2.3 Church's Thesis– A Case Study for Fallibilism . . . . . . 74

2.3.1 Introduction . . . . . . 74

2.3.2 Church's Thesis . . . . . . 76

2.3.3 Philosophical Arguments in Favor of Church's Thesis . . . . . . 80

2.3.4 An Argument Against Church's Thesis . . . . . . 89

2.4 Concluding Remarks: The Plausibility of Fallibilism as a Working Philosophy of Mathematics . . . 97

**3 Surveyability and the Four Color Theorem** **99**
3.1 Introduction . . . 99
3.2 History of the Four Color Theorem . . . 101
3.2.1 Kempe's Attempted Proof . . . 102
3.2.2 20th Century Developments on the Four Color Theorem . . . 104
3.3 Computer Facts about the Four Color Theorem . . . 106
3.4 How the Four Color Theorem Challenges the Classical Conception of Proof . . . 106
3.5 Thomas Tymoczko on the Four Color Theorem . . . 108
3.6 Objections to Tymoczko's View . . . 114
3.6.1 Teller's Comments on Surveyability . . . 114
3.6.2 Experiment and Mathematical Proof . . . 117
3.6.3 More on Surveyability—Has the Proof Really Been Surveyed? 122
3.6.4 Another Classical Defense of A Priori Proof . . . 123
3.6.5 A Computer Proof Predating the Four Color Theorem . . . 124
3.7 What is the Epistemological Status of Computer Proofs in General? . 128
3.7.1 Probabilistic Methods in Computer Proofs . . . 128
3.7.2 Does the Use of Probabilistic Methods Alter What Counts as a Proof? . . . 130
3.8 Closing Comments . . . 132

# List of Figures

3-1 Kempe's unavoidable set of configurations . . . . . . . . . . . . . . . . 103
3-2 a sample 6-ring . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 105
3-3 geometric illustration of the Theorem of Pappus . . . . . . . . . . . . 125

# Chapter 1

# Is Mathematical Knowledge A Priori: Responses to Kitcher's Skepticism

## 1.1 Introduction

What is a priori knowledge? Immanuel Kant was responsible for providing philosophers with an account that has turned out to be both a guiding principle and philosophical conundrum for hundreds of years; he wrote "we shall understand by a priori knowledge, not knowledge which is independent of this or that experience, but knowledge absolutely independent of all experience".[1] Turning these words into a plausible account of a priori knowledge has proved an arduous task.

Nonetheless, we do have some intuitions about what kinds of knowledge should be a priori on any reasonable account of a priori knowledge. On standard accounts, mathematical knowledge is held to be a priori, necessary, certain. Part of the explanation

[1] [Kant 1965], B2–3.

for this view is that the processes by which we come to know truths of mathematics—following proofs—offer special guarantees that other types of processes (in particular, perceptual ones) do not offer. But what is it about these processes that entitles us to claim these guarantees? Exactly what *are* these guarantees?

Philip Kitcher, in his book *The Nature of Mathematical Knowledge*[2] lays out what he sees as two ways of doing epistemology; he calls them *apsychologistic* and *psychologistic* approaches to epistemology. He points out various problems with the apsychologistic view, and maintains that doing epistemology psychologistically gives us the best chance for explaining the nature of knowledge in general, and a priori knowledge in particular. Once he has laid out a conception of a priori knowledge that fits his constraints on an adequate theory, he proceeds to give us reasons to think that mathematical knowledge might not be a priori after all. He seems to think that the guarantees we need for a process to qualify as a *warrant* for a priori knowledge are blocked by a number of challenges in cases of mathematical knowledge.

Proponents of a classical view of mathematical proof should take his charges seriously; if we agree with his general approach to epistemology, then we must examine closely the processes by which we come to follow proofs. It is with that project in mind that we will come to see that in fact Kitcher's requirements for a priori warrants are far too stringent; while it is important that the process of say, following a proof, be immune to certain recalcitrant experiences, we must distinguish between recalcitrant experiences which offer *reasons* and experiences which merely undermine my confidence in the theorem proved. It is to be hoped (by this author) that in fending off Kitcher's attack on the a priori nature of mathematical knowledge, we will illuminate Kant's famous words, and uncover some assumptions about the a priori so that it follows that (at least most) mathematical knowledge is a priori.

[2] [Kitcher 1984]

### 1.1.1 Epistemological Questions

Kitcher says that there have been two approaches used by philosophers to characterize knowledge. Before the end of the nineteenth century, he says that many philosophers used what he calls a *psychologistic* approach to epistemology[3]. For them, whether a belief state was a state of knowledge depended on how the belief was produced. Of course they supposed that knowledge was a state of true belief; what made it knowledge was that the processes engendering belief, consisting of events both internal and external to the subject, were of the appropriate kind. They saw their work in epistemology as specifying what kinds of processes were the right ones for engendering knowledge. Kitcher seems to be describing a version of reliabilism—the view that reliably generated true beliefs constitute knowledge, even though the believer may be ignorant of the process engendering the belief.

According to Kitcher, the twentieth century ushered in a new view about what constitutes knowledge, a view which denies that psychological processes have any relevance for whether a state is a state of knowledge. Kitcher calls this approach *apsychologistic* because its proponents consider "knowledge [to be] differentiated from true belief in ways which are independent of the causal antecedents of a subject's states"[4] What is important is the logical connections among a subject's beliefs; If a subject's belief that p is "connected in the right way" to certain other beliefs, the subject knows that p[5].

Kitcher appears to be giving an account of *foundationalism* in epistemology. According to this view, certain beliefs—foundational ones—are justified because of some intrinsic quality (e.g. being analytic) of the belief itself, even though the subject may

[3] Kitcher does not mention any philosophers specifically, but this description could apply to Locke, Hume, among others.

[4] [Kitcher 1984], p.41.

[5] Accounts varied, but in particular, p had to be "connected" to its logical consequences; if, say, p implies q and I know that p, then I should also know that q.

be ignorant of the existence of this intrinsic quality of the belief.[6]

Kitcher objects to apsychologistic epistemology because he believes it ignores some fundamental questions about mathematical knowledge, like "how do we know the axioms of mathematics?" He says that apsychologistic epistemologists often attribute to axioms a special status; they are called 'self-evident', 'a priori', 'analytic'. To attach these labels to mathematical propositions does not, in Kitcher's view, answer his question. His opponents would argue that these distinctions *do* help by separating the epistemological status of the axioms from the ways we come to know them. We shall consider this issue in greater detail later. However, we should note here that both foundationalism and reliabilism share the feature that whether a belief counts as knowledge relies on facts about which the believer may be ignorant.

Kitcher points out that other philosophers (notably Gilbert Harman and Alvin Goldman[7]) share his dissatisfaction with apsychologistic epistemology. They hold the view that knowledge depends crucially on having the right kind of process producing belief. Apsychologistic accounts of knowledge are flawed in that we can cite cases in which a subject may have a true belief, backed up by excellent reasons and the correct logical connections to other beliefs, but the circumstances under which the subject acquired his belief were defective in some significant way, thus precluding knowledge. So, even if a belief is say,a necessary truth, if it is arrived at by some unreliable method, then according to the reliabilists, it would not count as knowledge.

Kitcher says that Gettier examples show how the foundational approaches to knowledge are flawed:

Suppose that X comes to believe that p, p is true, but X's reason for believing that p is not the "right" kind of reason. Consider the following case. Jane sees Joan driving a black car, and comes to believe that Joan owns a black car. Joan *does* own a black car, but she happened to be driving Janet's black car when Jane saw her. So,

[6] [Clay and Lehrer 1989], p.xi.
[7] [Kitcher 1984], p.41.

Jane's reason for believing that Joan owns a black car *justifies* her belief, but it is not sufficient for knowledge.

The literature on this topic is well-known and suggests that justified true belief is not constitutive of knowledge. Kitcher is using Gettier problems to attack *internalist* theories of knowledge here; internalism attributes knowledge based on the internal features of beliefs (how they are connected to each other) rather than "the relationship between the belief and what makes it true".[8] He uses this set of problems to reject the apsychologistic view and instead focuses on how beliefs are acquired.

Kitcher suggests that their lack of attention to the psychological processes engendering belief created problems for the apsychologistic epistemologists, especially when they tried to give a characterization of a priori knowledge. He points out the problems in one account, given by A.J. Ayer[9], who suggested a way to define a priori knowledge:

X knows a priori that p iff X believes that p and p is analytically true.

Kitcher says that it follows from the above account that if we can show that mathematics is analytic (which is no small task), then we can say that we know a priori all statements of mathematics we believe. But of course this is not a correct conclusion. I could come to believe a mathematical statement in an unacceptable way– suppose I come to believe the Pythagorean theorem by dreaming about it, or hearing it from an unreliable source. My belief would not count as knowledge, let alone a priori knowledge.

Of course Ayer could respond to Kitcher's charge by saying that a priori is not the basic notion here, rather analyticity is. It could be that there is a class of propositions, all of which I know a priori just by coming to believe them. In this case Ayer is distinguishing between one's reasons for believing something and the evidence for its truth.

---

[8] [Clay and Lehrer 1989], p.xi.
[9] [Ayer 1946]

Kitcher notes that apsychologistic epistemologists tried to improve their formulations—to make more sophisticated versions of justified true belief—but he says all of them failed for similar reasons: "Our success [in defeating apsychologistic proposals for knowledge] results from the fact that the mere presence in a subject of a particular belief or a set of beliefs is always compatible with peculiar stories about causal antecedents"[10]. Kitcher seems completely convinced that Gettier examples preclude the possibility of *any* correct apsychologistic characterizations of knowledge. The only adequate approach for him is to use completely psychologistic principles which count as knowledge beliefs produced only by certain kinds of processes.

### 1.1.2 A Preliminary Psychologistic Account of Knowledge

Kitcher introduces a simple psychologistic account of knowledge. He uses the term 'warrant' to refer to those processes which produce belief "in the right way".[11] His analysis follows:

X knows that p iff p is true and X believes that p and X's belief that p was caused by a process which is a warrant for it[12].

Filling out the theory requires specifying conditions on warrants. Most important to showing that a process is a warrant is showing that, given that some process caused a belief, it also functioned to warrant that belief. Background conditions- features of the world both external and internal to the subject's psychology- can affect the warranting power of a process. Given background conditions, a process may not qualify as a warrant for some belief.

Kitcher offers an example from perception.[13] Suppose I am looking at some flowers

[10] [Kitcher 1984], p.16.

[11] [Kitcher 1984],p.16.

[12] Kitcher adds that 'process' refers to a token process, a specific sequence of events- not a process type. Two processes can both belong to the same type but not both warrant belief that p, given different background condition.

[13] [Kitcher 1984], p.20.

on a table under normal conditions. I come to believe that there are flowers on the table. According to any reasonable theory of knowledge, perception counts as a process that can potentially warrant belief. Now suppose that on some other occasion, the flowers on the table are surrounded by high quality fake flowers which I cannot distinguish from real flowers. It is possible that I underwent the exact same process in both cases (I saw them the same way in similar light, etc.). But, because I cannot tell the real flowers from the fake ones I cannot now be said to have knowledge that there are real flowers on the table, whereas in the former case I could. So even processes which are potential basic warrants do not function independently of other beliefs; background conditions affect the warranting power of a process.

This example is a standard one in epistemological literature; it is used to point out the contextually relative and sensitive nature of processes; this assumes a strongly externalist view of knowledge, as facts external to the believer of which she may be ignorant may influence the warranting power of the process by which she comes to hold a belief.

Kitcher says that the *same* process can be a warrant at some but not all times, depending on background conditions. He will need to fill in the details of how the warranting process works, what background conditions affect the warranting process, and how that process is affected. We will see later that in his account of priori warrants, he maintains that sufficiently many background conditions interfere with the warranting processes involved in acquiring mathematical knowledge so as to preclude its being a priori.

### 1.1.3 A Psychologistic Account of A Priori Knowledge

Now that Kitcher has outlined a general psychologistic approach to knowledge, he turns to the special case of a priori knowledge. 'A priori' applies to an item of knowledge. To say that I know a priori that p is to say that a certain kind of process caused my belief that p. So, to say that mathematics is a priori is to say *how* we come to know mathematical statements. But what kinds of processes are a priori processes? In particular, what kinds of processes are a priori warrants for mathematical knowledge?

Kant' well-known explication of a priori knowledge, (given at the beginning of this paper) leaves a lot unclear. Especially vague is the phrase "independent of all experience". It is ambiguous– it could mean "independent of *all* knowledge" or "independent of any particular item of knowledge".

For purposes of making this more clear, Kitcher uses a standard interpretation of Kant: an item of knowledge is a priori if any experience which would enable us to acquire the concepts involved would enable us to have that knowledge. To make this explicit, Kitcher introduces some terminology.

Let X's experience at t be her sensory state at t. X's sequence of experiences she has had up to t is X's life at t. A life is sufficient for X for p iff X could've had that life and gained sufficient understanding to believe that p[14].

Kitcher uses this terminology to give the following definition of a priori knowledge:

X knows a priori that p iff X knows that p and, given any life sufficient for X for p, X could've had that life and still have known that p.

Kitcher note that this account will not work; it is far too weak a formulation. A lot hinges on how we interpret the modality "could've". Does it mean that X does not actually have to *have* a life in which X acquires the appropriate concepts? This account still does not seem to guard sufficiently against defective ways of belief acquisition. Furthermore, it would seem to follow from Kitcher's first formulation that I could know a priori e.g. that violet is darker than blue—that statement could be analytic (on some accounts), and if so, then in any life in which I acquired the relevant concepts, I would have come to believe it. This argument would also work

[14] [Kitcher 1984], p.22.

on universal empirical knowledge– that there are bodies, etc.

Also, a proper formulation of a priori knowledge should distinguish between empirical knowledge of propositions that can be known a priori and true a priori knowledge; we have to count as two different processes the cases in which e.g. I come to know 2+2=4 by proving it and by counting small piles of rocks. What we need to do in order to characterize true a priori knowledge is to specify the ways we actually come to know a proposition a priori.

Kitcher gives an improved version:

X knows a priori that p iff X believes that p, p is true, and p was produced by a process which is an a priori warrant for it.

Kitcher has shifted the burden of defining a priori knowledge to the definition of an a priori warrant. Recall that warrants are psychological processes resulting in beliefs. But what kinds of processes count as a priori warrants? Clearly, perception is ruled out, but what does qualify? Kitcher gives no examples of his own but offers Kant's use of pure intuition (with respect to geometry) as a candidate. He does not explain what he thinks our intuition *is*, but assumes it works roughly in the following way. Using pure intuition, we (roughly) create a mental picture of, say, a triangle, inspect it, and make judgments about its qualities. What is important to isolate is exactly what makes that process an a priori warrant.

Kitcher says there are three conditions on a process which purports to serve as an a priori warrant:

1. it must produce warranted belief independent of experience.

2. it must produce true belief independent of experience.

3. the same type of process must be available independent of experience.

It is unclear what Kitcher wants from 3. A number of things are left vague. Does he mean the same type of process to be available independent of all experience of just any particular experience? Surely he does not mean the former. As for the latter option, he is obligated to provide us with a coherent explication.

Kitcher does not want to confine a priori knowledge to necessary truths; he would like to maintain the possibility of contingent a priori knowledge. He does not give any examples here, but presumably he not not want the apriority of mathematical knowledge to hinge on its necessity.

If all a priori truths were necessary, then 1. would follow, says Kitcher. No matter what experiences we had, our mathematical beliefs would be warranted. Why is this so, if the warranting power of a process is affected by contextual information or background conditions? The answer goes roughly as follows:

Everything necessary that is known a priori has the following property: every process that is a warrant for it is an a priori warrant. Why? Because from its necessity we know that there are no possible worlds in which it is false, so there is not a counterfactual situation in which something which *was* a warrant for a belief ceases to be one. The only ways that a warrant could lose its warranting powers would be if 1) the contextual information changes; or 2) there are worlds in which the belief is false. So, if 2+2=4 is a necessary truth, then it seems to follow that *any* process I went through to arrive at that belief would be an a priori warrant for it if it were a warrant at all.

Kitcher wants very strong requirements on a priori knowledge. He says a priori warrants should be "ultra-reliable— they never lead us astray".[15] He adds that it should follow from his account that "in a counterfactual situation in which an a priori warrant produces the belief that p, then p".[16]

[15] [Kitcher 1984], p.24.
[16] [Kitcher 1984],p.24.

### 1.1.4 Kitcher's Account of A Priori Warrants

From these considerations Kitcher gives the following analysis of a priori knowledge:

2. X knows a priori that p iff X knows that p and X's belief that p was produced by a process which is an a priori warrant for it.

3. A is an a priori warrant for X's belief that p iff A is a process such that, given any life e, sufficient for X for p

   (a) some process of the same type could produce in X a belief that p;

   (b) if a process of the same type were to produce in X a belief that p, then it would warrant X in believing that p;

   (c) if a process were to produce in X a belief that p, then p.

In the above account, Kitcher often refers to types of processes. To understand what he means here, we need to know how to specify what a process is and how to divide them into types. Kitcher defines a process as the terminal segment of the causal ancestry of a belief, restricted to states and events internal to the believer. Otherwise, he says, the process would not be available independent of experience.

In the interests of neutrality, Kitcher does not give a specific taxonomy for type-identification of processes. He does say, though, that our intuitions provide some guidance for dividing them. It is obvious to us that some ways of acquiring beliefs are different from others. For example, hearing a statement of the Pythagorean theorem from one's grandmother and following a proof of it clearly should count as different ways of coming to believe that the sum of the square of the hypotenuse of a right triangle equals the sum of the squares of the lengths of the two shorter legs. So these two processes should count as belonging to different types.

Naturally, how fine-grained a distinction we make between types will vary, depending on the context. However, Kitcher warns that some type-division proposals would flout any of our principles of taxonomy. Any taxonomy which counts e.g. both the process of following a proof of a theorem and also hearing it from your grandmother as being of the same type should be disallowed. Even though Kitcher claims the theory is neutral, it must not violate our intuitive principles for what count as dissimilar ways of forming beliefs. The principles he has in mind may turn out to influence how we type-identify processes.

### 1.1.5 Kitcher's Challenge and the Apriorist Response

Kitcher sees his psychologistic framework as constraining the apriorist program in important ways. If the apriorist philosopher is to succeed in making a priori knowledge a useful notion for epistemology, then she must follow the form he has specified, and then fill in the details. She must specify processes according to the restrictions in 3) and give type-identity conditions which conform to some principles of classification which are he says are standardly used in dividing processes of belief-formation.[17] If her account of a priori warrants has satisfied 3), she succeeds; Otherwise her case for the existence of a priori knowledge has failed.

I intend to meet this challenge, not by satisfying Kitcher's requirements, but rather by showing that his analysis places unrealistic constraints on what counts as a priori knowledge. I will show that the conditions on a priori warrants in 3) result in *nothing* being a priori, which is a problem. While there are plenty of reasons to object to a classical notion of the a priori, Kitcher's approach tries to provide for an account of a priori knowledge; failing to show that anything satisfies his account makes his entire epistemological approach less plausible. Also, I will show that there are reasons to believe that within Kitcher's framework his arguments against a priori knowledge work equally well against all knowledge. So if he is successful, he will have eliminated the possibility of any kind of knowledge of mathematics, not just a priori knowledge

[17] [Kitcher 1984], p.26.

of it.

## 1.2 Kitcher's Attack on Mathematical Apriorism

Kitcher sets up his psychologistic framework to include the notion of a priori warrant so he can attack what he calls "mathematical apriorism". According to him, proponents of mathematical apriorism consider mathematical knowledge to be a priori knowledge. Since most statements in mathematics are justified by use of proofs, Kitcher focuses on what he thinks is the traditional notion of proof. He tries to show that the process (or processes) of following a proof does not meet the requirements for a priori warrants.

### 1.2.1 The Role of Proofs in Mathematical Knowledge

Kitcher begins his examination by looking at how we standardly characterize proofs. He objects to what he calls a *structural* conception of proofs– the view that a proof in a system is a sequence of sentences in the language of the system such that each member of the sequence is either an axiom of the system or a sentence which results from previous members of the sequence in accordance with some rule of the system. He thinks that it is presumptuous to think that proofs in standard formal systems are the only acceptable kind of proofs.

Other criteria enter into our decisions as well, like acceptance by the mathematical community. Kitcher's point is well-taken but there are reasons to think that the mathematical community accepts proofs at least in part *because* they are of a standard form. Without standards of formal rigor, it would be much more difficult to tell whether a proof was acceptable. Also, it is possible that without such standards mathematicians would have more disputes over whether something was a proof.

Kitcher adds that the notion of proof evolves through time; in the past 100 years, we have become more rigorous and advanced. Our proofs are not written in the language of first-order logic; they are abbreviations of formal proofs. What counts as a formal, rigorous proof or an informal, abbreviated proof is dependent on the community. Given the fact that the community changes constantly, what makes out current proofs "genuine" proofs?

Given that acceptable proofs are informal abbreviations, dependent on the audience, and in a constant state of change, what makes them "genuine" proofs?

Kitcher says that apsychologistic epistemologists answer Kitcher's question in the following way: genuine proofs are those whose axioms are "basic a priori principles" and whose rules of inference are just those that are "elementary a priori rules of inference".[18] But to say this is not to give a complete explanation, Kitcher responds, unless accompanying it is a thesis about how these principles can be known, and how we can use these rules of inference to extend our knowledge[p.37], a thesis which must be detailed and well-argued.

A better way, says Kitcher, of characterizing proofs within the framework of an adequate epistemology is to give a functional definition. Proofs are sequences of sentences that serve a certain purpose for us. But what purpose? Kitcher says the apriorist would have to explain the purpose as follows:[19]

> proofs codify psychological processes which can produce a priori knowledge of the theorem proved. Similarly, to follow a proof is to engage in a particular kind of psychological process which results in the acquisition of a priori knowledge.

What does it mean to say that a psychological process *can* produce a priori knowledge? A number of psychological processes can result in e. g. believing that 2+2=4, but only certain ones count as following a proof. Kitcher's characterization must

[18] [Kitcher 1984], p.37.
[19] [Kitcher 1984], p.37.

specify what kinds of processes are the right ones in order to separate a priori from a posteriori knowledge.

To clarify what he means here Kitcher introduces some terminology. A statement is a basic priori statement "if it can be known a priori by following a process which is a basic warrant for belief in it".[20] Recall that basic warrants are processes which involve no other beliefs, according to Kitcher. So far he has offered no examples of processes that might qualify.

Kitcher says that proofs must begin from basic a priori statements. Further statements result from applications of apriority-preserving rules of inference. A rule is apriority-preserving just in case "there is a type of psychological process, consisting in transition from beliefs in instances of the premise forms to the corresponding instances of the conclusion form, unmediated by other beliefs, such that, if the instances of the premise forms are known a priori, then the transition generates a priori knowledge of the instance of the conclusion form."[21] We are still left not knowing exactly what he has in mind. For example, it is unclear how his analysis would explain how we use e. g. modus ponens.

Using these psychologistically defined terms, we can now define proof as Kitcher thinks the apriorist should:[22]

> To follow a proof is to undergo a process in which, sequentially, one comes to know the statements which occur in the proof, by undergoing basic a priori warrants in the case of basic a priori statements and, in the case of those statements which are inferred, by undergoing a transition of the type which corresponds to the rule of inference in question.

Characterizing proofs functionally as well as structurally gives us insight into what purposes proofs serve in mathematics. While it is true that what we mean by

[20] [Kitcher 1984], p.38.
[21] [Kitcher 1984], p.38.
[22] [Kitcher 1984], p.38.

'proof' is 'proof in a standard formal system with a certain form...', that does not completely explain why we consider those particular sequences to be proofs. What makes them proofs is that they do a certain job– they convince us of the truth of the theorem proved, using clear, explicit, accepted reasoning. Proofs serve a prescriptive, normative function. If I have followed a proof of the Pythagorean theorem, then I can conclude with impunity that whenever I do computations involving right triangles, if I add the squares of the lengths of the two shorter legs, the sum will equal the square of the hypotenuse. Following a proof of a theorem gives me good reasons to believe that it is true, and these reasons justify my belief in the theorem. In fact, following a proof *compels* my belief in the theorem.

The mathematical apriorist would readily agree that proofs are distinguished by the fact that they increase our mathematical knowledge. Since proofs begin with basic a priori principles and proceed using apriority-preserving rules of inference, one can follow a proof and extend his knowledge without adverting to experience; that is, once he understands the concepts involved, his resulting knowledge is warranted or justified, matter what kind of experiences he has had.

For Kitcher, "Psychological processes" refers to internal causal processes of the subject. They must be internal since their production warrants a priori knowledge. For a proof to "codify" or pattern psychological processes there should be kind of correspondence between the steps in the proof and the steps in the processes.

On standard accounts, The activity of following a proof of a theorem involves engaging in a process that results in acquisition of a priori knowledge. Kitcher is right to point out that we are owed an account of what that process *is*. The apriorist could respond that the notion "following a proof" is a primitive notion, but it is in fact a complicated process. Consider the following example: I come to believe p based on my belief of p and q. If we examine the individual steps, we see that there is a transition in my belief state that is somehow brought about by the previous step.

Whatever makes me make the conclusion *is* the process; it is the relation between the step and the transition.

For Kitcher's analysis to be successful, he must also explain how proofs codify psychological processes. Roughly speaking, we do engage in certain mental activities when we follow the steps in a proof, but there is no reason to believe that there is a 1-1 correspondence between the steps in a proof and the psychological processes we undergo. Kitcher offers no suggestions about how to translate steps in a proof into psychological processes. And we do not have any intuitions about how many discrete psychological processes we undergo in applying the rule of modus ponens, for example. Whereas Kitcher rightly points out that we do undergo some psychological transitions when we follow proofs, his analysis leaves the details of how this works unexplained. Of course, so does the apriorist, but that means that his account does not provide more explanatory power than the standard view.

If the psychologistic epistemologists are right, then the best way to answer the question "what job do proofs do?" is to be had by looking at what we do when we follow proofs, and how proofs reflect mental processes we undergo.

This investigation could prove helpful for answering questions in epistemology. For example, if psychologists discovered that the processes corresponding to steps in a proof all required perceptual mechanisms, then a case could be made that mathematical knowledge is empirical. On the other hand, if experiments determined that mental processes in the practice of doing mathematics were just the instantiations of logical principles, starting from logical axioms, then that would be evidence that mathematical knowledge is knowledge of logic. Or if psychological data show only non-empirical mechanisms at work in calculation, that we would be more inclined to consider mathematical knowledge a priori.

Kitcher's analysis of proof interpreted charitably might, with the appropriate accompanying data, yield some hypotheses about how our mathematical practices work.

But, it fails to take into account the job proofs do– they are arguments, giving us sufficiently good reasons to believe that a theorem is true. Kitcher never links the notion of proof to the justification of the truth of a statement; or, more important for his project, to justification of one's knowledge of mathematics.

Kitcher clearly states that he is rejecting the apsychologistic rendering of knowledge as justified true belief for the reason that the account fails to disqualify as knowledge cases of true belief acquired in some epistemically defective way. Gettier-type cases give ample intuitive evidence for the need for an appropriate causal story if a true belief is to count as a state of knowledge. But in the case of mathematical knowledge, there are two stories to be told: first, some account of how we acquire knowledge of mathematics; second, an explanation of how proofs serve to justify our beliefs that many mathematical statements are true, how proofs reveal the deductive structure of mathematics, and how we can use them to extend our knowledge of mathematics.

Frege was interested in working out the details of the second story. He thought epistemology concerning mathematics should definitely be apsychologistic.[23] Frege was disturbed that some mathematicians "confuse the grounds of proof with the mental or physical conditions to be satisfied if the proof is to be given".[24] He cites one of his favorite examples from the literature of his time Schroeder's "Axiom of Symbolic Stability. It guarantees us that throughout all our arguments and deductions the symbols remain constant in our memory- or preferably on paper".[25] That psychology could affect the foundations of mathematics to the extent that we needed safeguards against mysteriously changing variable letters seemed absurd to Frege. What he thought affected the foundations of mathematics was the degree of rigor with which many results were formulated.

[23]In "Frege's Epistemology", Kitcher defends the view that his psychologism is not the kind to which Frege would have objected.

[24]Grundlagen, p.VIII

[25]Grundlagen, pp.VIII-IX.

Frege acknowledges that much of mathematics seems self-evident. To require a proof of $2 + 2 = 4$ is "almost ridiculous". But proofs for Frege do more than just establish the truth of a theorem: "the aim of proof is, in fact, not merely to place the truth of a proposition beyond all doubt, but also to afford us insight into the dependence of truths upon one another".[26] Proofs hold the key to mathematical advancement. By doing them we learn the limits of application of techniques and concepts. Proofs uncover part of the deductive structure of mathematics. Knowing where a theorem fits within this structure helps us decide where and when to look for new theorems.[27]

Kitcher's account also does not provide an explanation of how proofs lead us to mathematical discoveries. The apriorist has no reason to accept his notion of proof, for it fails to capture key aspects of proof– its use in justification, its use in extending our knowledge. She can concede the benefits of a causal account in explanation of the origins of mathematical knowledge. But Kitcher's story may turn out to be insufficient for purposes of doing work in foundations of mathematics.

### 1.2.2 Kitcher's Account of Proof

Now Kitcher is ready to provide the following thesis about the form of apriorist proof:[28]

4)there is a class of statements A and a class of rules of inference R such that:

a) each member of A is a basic a priori statement; b) each member of R is an apriority-preserving rule;

c) each statement of standard mathematics occurs as the last member of a sequence, all of whose members either belong to A or come from previous members in

[26]Grundlagen, p.2.

[27]By "structure", I do not mean logical structure. I merely use the term to refer to whatever organisation exists in various fields of mathematics. No logicistic assumptions are intended.

[28] [Kitcher 1984], p.39.

accordance with some rule in R.

He says that whereas 4c) has been traditionally regarded as controversial, 4a) and 4b) have been accepted without question. Since the goal of his program is to uncover apriorist assumptions about mathematical knowledge, Kitcher considers 4a) and 4b) just as suspect as 4c). Kitcher makes two criticisms of mathematical apriorism. The first attacks claims which are instances of 4a); the second examines the apparent incompatibility between the supposition that many theorems of mathematics can be known a priori and the fact that some of these theorems can be proved only by demonstrations of great length.

Kitcher attacks 4a) first, saying that apriorists have committed themselves to the existence of basic a priori statements, which he thinks is a terrible mistake; processes traditionally regarded as a priori warrants are in fact not, so they cannot call instances of 4a) a priori.[29] I will first consider a worry directed at the general form of his argument.

Here is the general structure of his argument against 4a):

1')Traditionally, we have regarded many statements as being basic a priori.

2')A necessary condition for basic apriority is being caused by a process which is an a priori warrant for it.

3')Many statements which have been traditionally regarded as basic a priori are not caused by processes which serve as a priori warrant for them.

4')Therefore, such statements are not basic a priori.

Kitcher's argument initially looks reasonable, but it depends crucially on the plausibility of 2'). Unless Kitcher's case for 2') is quite persuasive, there is as much reason to conclude not 2') as there is to conclude 4') Why? Well, we have strong reasons to think that that 1') is true– intuitions and philosophical traditions support the fact that many statements are a priori. If Kitcher wants to hold 2'), then he must give

[29] [Kitcher 1984], p.39. Kitcher devotes Chapters 3 and 4 to the task of discrediting several theories of knowledge which are committed to the existence of basic a priori statements.

evidence for it that either appeals to our views about a priori knowledge or shows us how our intuitions were mistaken. Otherwise, the more sensible solution is to conclude that a priori warrants are not necessary for a priori knowledge. Later in this section I will argue that this is just what we should conclude.

Kitcher's second criticism questions the correctness of 4) as a characterization of a priori proof. Consider his inductive argument:[30] Let S be any true mathematical statement. By 4c) there is a sequence of sentences, [which is the proof of S], all of whose members belong to A or come from previous members by one of the rules in R. We can show by induction, using 4a) and 4b) that every statement in the sequence is knowable a priori. A fortiori, S is knowable a priori. Hence every truth of standard mathematics is knowable a priori.

It follows from the analysis above that if S is a proof, (consisting of basic a priori principles or following from one by use of an apriority-preserving rule of inference) then we should be able to come to know S by following a proof of it. That process (following the proof) will then serve as an a priori warrant for the belief that S.

Kitcher says that the existence of very long proofs of mathematical statements may threaten the inductive argument. There are theorems whose proofs are so long that one person could take years to go through one of them. Since I am fallible, it is possible that such a proof contained some errors that I overlooked. I may not be completely certain that I followed the proof correctly, and my knowledge of the statement is therefore not a priori.

Kitcher gives three possible resolutions of this conflict:[31]

i) we can accept the inductive argument and the point about long proofs, concluding that no version of 4) can be correct.

ii) we can accept the inductive argument and reject the point about long proofs, thereby concluding that 4) is sufficient to establish apriorism.

---

[30] [Kitcher 1984], p.40.
[31] [Kitcher 1984], p.40.

iii) we can reject the inductive argument, concluding that 4) does not suffice to establish apriorism.

We know that, for familiar reasons, inductive arguments involving vague predicates are not always valid. Kitcher points out that the worry about long proofs could be because the term 'a priori' is vague. It is possible that the statements encountered early in a proof have a high degree of certainty, but inferences resulting in later conclusions do not preserve certainty. After some point it would not be correct to ascribe knowledge of the conclusion, just as at some point in the process of adding 1 to a number it would not be correct to call that number 'small'. iii) is certainly a possible explanation of the problem long proofs present, but saying that 'a priori' is vague does not tell us whether we should reject apriorism; it is unclear what conclusion we should draw if it turns out that apriority is a vague notion.

Apriorists, Kitcher says, will oppose iii) and defend a priori knowledge of the conclusions of long proofs, adopting ii). Kitcher acknowledges Hume's observation that as we review proofs and others agree that they are correct, we become more convinced of their truth. But the fact that our certainty increases with agreement of our peers is not relevant to the truth of the theorems. Kitcher considers uncertainty about a proof to be incompatible with a priori knowledge of it.

The apriorist will want to separate the psychological feelings of certainty about proofs from the epistemological status of the theorems proved. He suggests two ways to defend ii). We could say that uncertainty stems from the fact that most proofs are informally structured, and formalization would remove any doubt. Another possibility is to propose that we can know a proposition without knowing it for certain.

The first suggestion clearly will not work. Quite the contrary– presenting a proof in formal notation will increase its length enormously, exacerbating the problem. Kitcher correctly notes that some theorems never receive rigorous proofs, even by informal standards. Also, the activity of formalizing proofs is just as subject to errors, so formalization leaves us in a worse state.

What about the second option? Does rational uncertainty preclude a priori knowledge? Kripke thinks not:[32]

> Something can be known, or at least rationally believed, a priori, without being quite certain. You've read a proof in a math book; and, though you think it's correct, maybe you've made a mistake. You do often make mistakes of this kind. You've made a computation, perhaps with an error.

Kripke thinks it is a mistake to conflate apriority and certainty. Kitcher does acknowledge a distinction: "One can go easily astray here, by conflating a priori knowledge with knowledge obtained by following a non-empirical process".[33] But Kitcher disagrees with the view that rational uncertainty is compatible with a priori knowledge. A priori knowledge for him is, in Mark Steiner's words, "...incorrigible– we could never be justified in giving it up once it is warranted. And perhaps it is even unrevisable– meaning that nothing at all could shake our conviction"[34].

But Kant never had such stringent requirements for a priori knowledge- for him it was nonempirical and necessary.[35] So it looks like there are at least two notions of apriority, but Kitcher chooses to attack the more stringent one[36]. Mathematical knowledge probably does not meet the requirements for Kitcher's notion of apriority, but that is not surprising. Nor is it disturbing; no one would expect that it should. What is important for the apriorist is that mathematical knowledge be a priori in Kant's sense, an issue we will examine later.

Kitcher thinks that uncertainty interferes with the a priori warranting process because experiences could yield situations in which e. g. the book in which the theorem

[32]Kripke, 1972,p.39
[33] [Kitcher 1984], p.43
[34]Steiner, p.452.
[35]Steiner, p.452.
[36]Steiner, p.452.

was proved is discredited, or the mathematical community decides to reject the proof. Kitcher says that when I have doubts arising from following a complicated proof, then if there are also circumstances under which experiences suggested the falsity of the theorem (e. g. the book I read was discredited by mathematicians), then I cannot conclude that I know the theorem a priori. However, this is not the case with regular warranting processes "because of the kindly nature of background experience".[37] So, rational uncertainty does not preclude knowledge, but it does preclude a priori knowledge, which leads Kitcher to conclude i).

Why does Kitcher believe that non-apriori warrants are not also undermined by rational uncertainty? it seems as if he is saying that even if I am warranted in believing that p (either a priori or a posteriori), I do not *know* that p unless background conditions are "right". Non-a priori beliefs must then be less affected by background conditions, whereas a priori beliefs are more likely to be affected. This is exactly the opposite of my intuitions about knowledge.

Kitcher does not argue for this point other than to say that if the quality if our lives were different, that rational uncertainty would preclude knowledge; since many of our ordinary beliefs are not undermined by experience (unlike some of our mathematical beliefs), ordinary knowledge is not blocked. I find this argument puzzling; we have lots of experiences that undermine our perceptual judgments, but few that undermine e. g. the belief that 2+2=4. Optical illusions, perceptual infirmities, and poor lighting are all common examples of how experience can lead us astray. But, we do count very many of our perceptual beliefs as knowledge. If Kitcher allows rational uncertainty to interfere with a priori knowledge, which we intuitively count as equally or even more certain than a posteriori knowledge, then I do not see how he can stop rational uncertainty from precluding knowledge. Certainly any obvious attempts to solve this problem would strike me as ad hoc solutions, unless they gave compelling reasons to

[37] [Kitcher 1984], p.43

explain away and overcome our intuitions.

Kitcher considers the problem with uncertainty undermining knowledge of long proofs to be the same one Descartes encountered with deductions in the Regulae. Since extended deductions "exceeded the scope of what we can simultaneously present to ourselves"[38], they are uncertain. Descartes' solution involved practicing following the deduction to ourselves so often that eventually we can apprehend the entire proof in one mental act. Although Kitcher, like others, views Descartes' solution as infeasible because of our physical limitations, he agrees with Descartes' picture of the psychological process of following a proof.

Kitcher gives a modern version of Descartes' view.[39] A proof begins with an axiom, which I intuit, and from it I infer (using, perhaps, one-premise rules) the next statement, and from that the next one, and so on. Suppose also that I can present to mind only one axiom and three inferential steps in the proof. Then I store the results, recalling them in order to go on following the proof. In Kitcher's terminology, I undergo a process which is a basic warrant for belief in the axiom. The problem arises when I no longer believe the axiom on the basis of the original warrant, but only because I remember having undergone a warranting process. But that process– the recollection– is not itself an a priori warrant for my belief in the axiom, so my belief is uncertain. So we must conclude that no version of 4) can be correct.

It is true that we cannot represent long proofs to ourselves; few mathematicians would ever consider that a requirement for knowledge of a theorem. In fact, some psychological studies have shown that we cannot in general represent more than 7 symbols in short-term memory.[40] If this is the case, then Kitcher's skeptical worries about long proofs are ill-founded. Furthermore, his skepticism leaves open the possibility for other, even more extreme, skeptical worries. For example, maybe during

[38] [Kitcher 1984], p.43.
[39] [Kitcher 1984], pp.44-45.
[40]find reference for this.

the process of following a proof I have forgotten what the words mean. Clearly this is not a reasonable worry, but it is unclear how Kitcher's skepticism is more moderate than the above worry.

What Kitcher's analysis shows is that he has constructed a notion of following a proof such that no one can have a priori knowledge of proofs. We are not required to conclude that the process of following a proof is not sufficient for a priori knowledge, but can instead reply that Kitcher places unreasonable restrictions on what constitutes following a proof. It is also unclear whether he has come up with any *sufficient* conditions on a priori knowledge.

### 1.2.3 More Challenges to A Priori Knowledge

Kitcher continues his attack on mathematical apriorism from anther perspective. In the context of challenging the status as a warrant of Kant's process of pure intuition, he outlines some standards required of all processes which purport to serve as a priori warrants. Not only must a process be nonsensuous to qualify, it must be *infallible.* Kitcher claims that pure intuition fails as an a priori warrant on the grounds that it is not infallible. In order to make his claim effective, he will have to 1) give an account of pure intuition and show that it is fallible; and 2) show that its fallibility undermines its status as an a priori warrant. He will also claim that his criticisms of pure intuition apply to all putative a priori warranting processes.

Some terminology is required to formulate Kitcher's thesis. A type of process which generates belief is called *dubitable* "if there is a life given which it would be reasonable to believe that some processes of the type engender false beliefs."[41] Consider some process $a$, a candidate for an a priori warrant for belief that p. We assume that we know that $a$ belongs to a type of process, called the *availability type* of $a$, such that a process of that type would be available given any sufficient experience. If $a$'s

[41] [Kitcher 1984], p.54.

availability type is dubitable, and if I come to believe that p via $a$, then on Kitcher's account it is possible that I could come to believe falsehoods via $a$. Kitcher's thesis goes as follows:

5) If the availability type of $a$ is dubitable and, if there are lives which would suggest the falsity of p, then there are sufficient lives given which the available processes of the same type as $a$ would not warrant belief that p.

The idea here is that if I can have grounds for questioning the reliability of a given type of process for generating warranted beliefs, then if there are also circumstances under which I have experiences suggesting the falsity of the belief that p, then I would not be warranted in the belief that p.[42] I would not be so warranted because although $a$ produces p, there could be a process of type $a$ which produced p, but would not warrant belief that p. This situation violates condition b) on warrants, that is if a process of the same type were to produce a belief that p, it would warrant belief that p.

Kitcher acknowledges that he is taking it for granted that there are experiences which could suggest the falsity of (in this case) geometrical axioms. Does he have the right to this assumption? After all, he points out that for Kant not only are mathematical truths necessarily true, but they must necessarily appear to be true; we cannot even imagine the falsity of mathematical statements.[43] It was also Kant who appreciated the fact that we cannot even imagine the falsity of true mathematical statements. But Kitcher firmly maintains that even if there are not direct experiences of the falsity of mathematical statements, there are various *indirectt* ways of suggesting their falsity. He uses a sample statement from geometry– that the sum of the angles of a triangle equals 180 degrees, which is a priori according to Kant.

Now, Kitcher considers three kinds of misleading experience which could challenge

[42] [Kitcher 1984], p.54.
[43] [Kitcher 1984], p.55.

our belief in the statement:[44]

1. direct challenge– a perceptual experience of a figure which, judged by our very best criteria, appears to contradict the statement.

2. theoretical challenge– a sequence of experiences which suggest that a physics-cum-geometry which does not include this statement will provide a simpler total description of the phenomena than a physics-cum-geometry which does.

3. social challenge– a sequence of experiences in which apparently reliable experts deny the statement, offer hypotheses about errors we have made in coming to believe it, and so forth.

Kitcher does not entertain the possibility of veridical challenges, in which our experiences *correctly* suggest the falsity of the statement; he agrees to concede that mathematical truths are necessary, thus excluding the possibility of such a challenge.

Let us examine each of these challenges in turn. Since Kant holds that our psychological constitution dictates the general structure of experience[p.55], direct challenges are ruled out. It is certainly hard, on any view, to imagine a perceptual experience which would suggest the falsity of a mathematical statement. If we allow direct challenges to be a threat to a priori knowledge then they seem to be a threat to knowledge as well. Again, the reason for the threat to all knowledge is that *all* warrants are affected by background conditions. A perceptual experience suggesting the falsity of a mathematical statement would seem to affect the warranting power of *any* process by which we come to know mathematical statements, not just the a priori processes. Surely Kitcher does not want direct challenges to threaten the status of our mathematical beliefs as knowledge; but if they threaten the apriority of our beliefs, it looks like they threaten their status as knowledge also.

---

[44] [Kitcher 1984], p.55.

Theoretical challenges in geometry are much easier to imagine. Discoveries of non-Euclidean geometries gave new interpretations to many theorems and caused others to be rejected. But despite this drastic revision in the status of Euclidean geometry, it is nonetheless true that the sum of the angles of a *Euclidean* triangle equals 180 degrees, and my knowledge of that theorem is a priori knowledge in Kant's sense, that is it is nonempirical and necessary. The existence of non-Euclidean geometries does mean that there is no unique description of the structure of space, but if I relativize all of my beliefs about Euclidean geometry by prefacing them with "in a Euclidean system...", then their a priori status remains. Of course, I have in effect replaced my former geometric beliefs with new ones, but they are still a priori knowledge.

Furthermore, it hard to imagine what kind of problems a theoretical challenge could present for, say, my arithmetical knowledge. I cannot give an example of any sequence of experiences which would suggest that a theory without the statement $2 + 2 = 4$ would be simpler than a theory with the statement. Perhaps Kitcher does not expect theoretical challenges to threaten arithmetic knowledge.

So, if worse comes to worst, Kitcher concludes, he can always use social challenges to make his case. Although it seems unlikely that experts would deny a statement we accept, we can imagine them producing theorems containing well-hidden flaws which we cannot detect, theorems that we do not believe but which they argue for convincingly. Kitcher says that such experiences would suggest the falsity of a mathematical statement, which is sufficient to preclude our a priori knowledge of it.

Hilary Putnam gives an example of the kind of scenario Kitcher must have in mind. He describes circumstances under which it would be rational to believe that Peano arithmetic was inconsistent even though it was not:[45]

> Thus, suppose I am caused to hallucinate by some marvelous process(say, by making me a 'brain in a vat' without my knowing it, and controlling

[45] [Putnam 1979], pp.97-98.

all of my sensory inputs superscientifically), and the content of the hallucination is that the whole logical community learns of a contradiction in Peano arithmetic (Saul Kripke discovers it). The proof is checked by famous logicians and even by machine, and it holds up. Perhaps I do not have time to check the proof myself but I would believe, and rationally so, I contend, that Peano arithmetic inconsistent on such evidence.

Kant would agree that his conception of a priori knowledge is open to social challenges. He never claimed such privileged epistemological status for pure intuition– that it be immune from *any* kind of doubt or peer pressure. Descartes' requirements for knowledge are closer to Kitcher's; still, Descartes would claim that social challenges apply only to *memories* of a priori warrants, not the warrants themselves. Steiner notes that "Descartes invokes the Deity to bolster only knowledge based upon past 'clear and distinct ideas'"[46]. Neither Plato nor Descartes would consider present a priori knowledge so vulnerable. Kitcher, on the other hand, expects a priori warrants to guarantee the elimation of all doubt– including the doubt that one is rational. If a candidate process does not result in indubitable true belief, then that process does not qualify as an a priori warrant.

Surviving social challenges is far too stringent a requirement for a priori knowledge. It is odd that Kitcher says whereas these experiences do not rule out knowledge, they do rule out a priori knowledge. He considers only indubitable types of processes to be sufficient to count as a priori warrants. However, he never offers an explanation of why he prefers a strict notion of apriority to the Kantian one. Limiting processes to nonsensuous ones would be reasonable; perhaps other constraints should apply. Indeed, other constraints should apply. He does too little to explain what count as sufficient conditions for a priori warrants. But Kitcher stacks the deck against mathematical apriorism by placing such severe requirements on a priori knowledge,

[46]Steiner, p.452.

requirements that mathematical apriorists need not accept.

## 1.3 A Case Against Kitcher's Views on Challenges to Knowledge

### 1.3.1 Introduction

Let us now look at Kitcher's allegations against apriorism from a different viewpoint. In the last section we saw the social challenge presented as a problem for Kantian intuition. Recall that a social challenge is any situation suggesting the falsity of (in this case) our mathematical beliefs due to e. g. apparently reliable experts denying the statement, offer explanations of how we erred, and so forth. According to Kitcher, a process does not count as an a priori warrant unless it can withstand social challenges. I offered objections to Kitcher's requirement of resisting social challenges; in this section we will see that Kitcher's requirements for a priori knowledge reveal crucial flaws in his views on mathematical knowledge.

Perhaps there are experiences suggesting the falsity of some of our mathematical beliefs. There are also experiences suggesting the falsity of some of our non a priori beliefs . Why doesn't the social challenge apply equally well to say, empirical knowledge? Kitcher says that rational uncertainty is compatible with non-a priori knowledge because of the kindly nature of background experience. However, if background experiences were sufficiently recalcitrant, rational uncertainty could create the same problems for empirical knowledge as it does for a priori knowledge. I will give an example of such a situation, which will force Kitcher to adopt one of the following positions:

1) social challenges are irrelevant to a priori knowledge, because no one requires immunity from all doubt for any kind of knowledge, no matter what process produces

it.

2) social challenges undermine a priori knowledge, but also non-a priori knowledge. The social challenge, if it applies, applies to beliefs regardless of the processes that produced them.

In order for Kitcher to avoid being committed to one of these positions, he must explain why social challenges apply to a priori knowledge only. I will show his explanation to be unsatisfactory. We can have experiences suggesting the falsity of many of our beliefs, but this is more a fact about our psychology than about the processes through which we acquire beliefs. I will present a psychological study that suggests we are easily subject to coercion about knowledge of things that are supposed to be certain; we are less inclined to be swayed about matters which accommodate dissent. If Kitcher's social challenge is successful, he has shown that processes leading to even ordinary beliefs fail to qualify as warrants, undermining knowledge itself, not merely its a priori nature. If he fails, then we can conclude that the social challenge is just a phenomenon resulting from facts about us and how we rely on others to bolster our confidence about many matters.

### 1.3.2 The Study

In his *Studies of Independence and Conformity of a Minority Against a Unanimous Majority*, Solomon Asch[47] presents a situation in which a social challenge seems to arise.The experiment tested people to determine the conditions of independence and lack of independence by a minority of one in the face of unanimous group pressure. Asch did an experiment in white male college student groups by setting up a disagreement between a single person and a group concerning a simple and clear matter of fact in the immediate environment. The group that disagreed judged the facts wrongly; the way the experiment was set up, the data to be judged couldn't reason-

[47] [Asch 1956]

ably be judged incorrectly by a person. The judgments of each person were to be stated publicly. The object of Asch's experiment was to use the response pattern of the subjects in order to state conditions responsible for independence and failure of independence in the face of unanimous opposition by the majority.

### Description of the Experiment

Asch's experiment was set up as follows. 7-9 white male college students were instructed to gather in a classroom to take part in what appeared to be a simple experiment in visual discrimination. They were instructed by an examiner to match the length of a given line – the standard – with one of three other lines. One of the three comparison lines was equal to the standard. The other two lines differed from the standard (and from each other) by considerable amounts– 3/4– 1 1/4 inches. The entire task consisted of 18 such comparisons. Individuals were instructed to announce their judgments publicly in the order in which they were seated. The comparison lines were numbered 1, 2, and 3 from left to right. The subjects stated their judgments by calling out the appropriate number.

This experiment would be just another innocuous test in visual perception if it were not for one vital fact – all but one member of the experimental group had met previously with the experimenter and were instructed to respond on certain trials with incorrect and unanimous judgments. The subject, who was unaware of this arrangement, heard the majority respond unanimously from time to time with estimates that clearly contradicted his own observation. The majority sometimes matched the standard to lines that departed from the standard by amounts of 3/4 inch to 1 3/4 inches. The differences in the lines were perceptually obvious; under control conditions - with subjects judging individually - their estimates were more than 99% accurate (of 7 control subjects, only two erred– one with one error, one with two errors) for an average of 0.8% error. So the unsuspecting subject, called the critical subject, was put in the position of a minority of one against a wrong and unanimous majority.

It should be noted that the majority was instructed to announce the judgments clearly and firmly, but not to take issue with the critical subject. They and the examiner were advised to act passively and impersonally toward the critical subject and not to act surprised at his answers.

The examiner read instructions from a card, explaining that the test involved visual discrimination of the lengths of lines. He instructed the subjects to announce their judgments aloud and as accurately as possible. The order in which the members of the group gave their judgments was always arranged (unbeknownst to the critical subject) so that the critical subject would answer next to last.

The task was 18 comparisons, consisting of a set of nine comparisons shown twice without a pause. There were also six neutral trials in which the majority responded correctly. In the interest of establishing some degree of trustworthiness in the majority, Asch made the first two trials neutral. The six neutral trials were numbers 1, 2, 5, 9, 10, and 13. That left twelve critical trials, ones in which the majority responded incorrectly.

### Quantitative Results of Asch's Study

The results of the experiment were surprising. Out of 123 critical subjects, only 29(approximately 25%) made errorless judgments, as compared to the control group, in which 35 of 37(95%) performances were errorless. The mean number of errors was 4.41 in the experimental group as opposed to 0.8 in the control group. The mean percentage of error in the experimental group was 36.8 The action of the majority brought about distortion of 1/3 of the reported estimates. 27% of the subjects made 8-12 errors as determined by the majority while only 24% gave errorless performances.

The experimenters noted that it is significant that the majority elicited widely differing reactions from the critical subjects. Also significant, though, was the fact that *most* (2/3) of the critical subjects' estimates were correct, which for them showed that the facts to be judged, not just the majority effect, were influential on the subjects' decisions. The experimenters presupposed that the stimulus conditions exert a fundamental effect on the character and course of the majority influence. That is why they choose as the object of judgement "facts or relations that possessed an independent status... Group action necessarily derives its significance from the reference it has to the facts, real or alleged".[48] It is Asch's emphasis on challenging the subjects' knowledge of an obvious and independent fact that makes this experiment such a useful example as a candidate social challenge, despite the fact that the knowledge in question is a posteriori, not a priori knowledge.[49]

## Qualitative Results of the Study

In addition to computing quantitative results, Asch's group interviewed critical subjects after the experiment. The interviews consisted of a series of questions designed to uncover the subject's feelings about his answers, e. g. whether he thought they were right, if he ever answered with the majority against his own choice. Then, after full disclosure of the purpose of the experiment, the subject was questioned as to his suspicions about the experiment. Any subjects who definitely suspected the purpose were eliminated from the study. The rest were questioned about their reactions to the situation.

The most common reaction was one of puzzlement. They reported having felt that during the experiment something was wrong, but they could figure out the source of

[48] [Asch 1956], p.13.

[49] Recall from earlier discussion that if there are possible experiences suggesting the falsity of the belief, then if the candidate process is dubitable, the social challenge succeeds in undermining the status of the process as an a priori warrant. But, if we can show that there are cases in which background experiences interfere with processes which are non-apriori warrants, then it follows that Kitcher has undermined non-a priori knowledge as well.

the problem. Later in the interview, when asked who they thought was right and who was wrong "most subjects, including the staunchest independents, at some time felt doubt about their accuracy, while the most pliable subjects at times felt the majority to be wrong".[50] What separated the independent subjects(0-2 errors) from the yielding subjects(3-12 errors) was not so much their immunity from doubt as their ability to free themselves of it. Virtually all of the subjects experienced conflict, but their manners of coping with it differed widely. Independents tended to say either that they felt they were right or, even though they doubted themselves, they felt obligated to report what they saw.

Yielders seemed to find being different intolerable. They thought others were following the leader, or they doubted their own judgement and gave the majority the benefit of the doubt. Other yielders denied that they went along with the majority and underestimated the number of errors they made. Some reported that if the question had been of a different sort, particularly one which allowed for dissent, they would have felt more comfortable with answering truthfully – "If it had been a political question, I don't think I would have agreed if I had a different feeling".[51] Asch explains the possible reasons for compliance or independence in the following way:[52]

> Independence requires the capacity to accept the fact of opposition without a lowered sense of personal worth. The independent person has to organize his overt action on the basis of experience for which he finds no support; this he can do only if he respects his experiences and is capable of claiming respect for them. The compliant person cannot face this ordeal because he translates social opposition into a reflection of his personal worth. Because he does so the social conflict plunges him into pervasive and incapacitating doubt.

[50] [Asch 1956], p.28
[51] [Asch 1956], p.42.
[52] [Asch 1956], p.42.

Asch's account of the differences between the yielding and the independent subjects postulates that the processes they undergo to arrive at a judgement diverge crucially. It seems that both groups receive the same raw data regarding the lines, but for important psychological and sociological reasons, one group is unable to report what is clearly seen in the trials. The explanation for this phenomenon does not make reference to differences between the perceptual mechanisms of the compliant and those of the independent subjects. Each subject experienced a situation which suggested the falsity of his (in this case) perceptual beliefs.

### 1.3.3 Applying Asch's Study to Kitcher

There are a number of ways to apply this case to Kitcher. Asch's experiment does appear to present a social challenge to perceptual knowledge. If the social challenge is successful against ordinary knowledge as well as a priori knowledge, then Kitcher wins; but as Mark Steiner says, he wins too much.[53] Kitcher is interested in attacking the a priori status of knowledge, not the status of a belief as knowledge simpliciter. If he does the latter, then the result is a reductio ad absurdum of his thesis, for he will have undermined the possibility of knowledge at all, a position he cannot reasonably hold.

Kitcher does try to ward off social challenges to non-a priori knowledge when he makes the point that reasonable uncertainty is typically compatible with knowledge because of the kindly nature of background experience.[54] But Kitcher does emphasize, in his discussion of social challenges, that for processes to count as warrants, background beliefs must also support the belief in question. "If you have reason to believe that your senses sometimes play tricks on you, then if you also have reason to think that the perceptual belief which you are inclined to form is false, your percep-

[53]Mark Steiner, J. of PHilosophy, find other info.

[54] [Kitcher 1984], p.43

tual process (which may, in fact, be perfectly normal) does not warrant the belief"[55]. Here Kitcher seems to be saying that situations like social challenges do succeed in undermining knowledge.

The above conclusion, however, does not fit with Asch's experimental data, particularly with the subjects' reports. Although 75% of the subjects made at least one error, the explanation for this phenomenon had nothing to do with a fault with their perceptual mechanisms' ability to warrant belief in their judgments; it had nothing to do with an inability to arrive at a correct judgement. The critical subjects that answered incorrectly did so because of subtle influences of peer pressure, because of fear of being conspicuous, or fear of causing an aberration in the experimenter's statistics.[56]

The obvious conclusion to draw from the Asch experiment is not that perceptual mechanisms are not sufficient to generate knowledge; rather, we should see that people vary in their ability to rise above conflict and self-doubt in order to report their beliefs accurately. Although Kitcher tries to maintain that social challenges undermine only the a priori nature of knowledge, he does not explain how he can do so and leave knowledge intact.

The Asch experiment likely could be modified to involve not a perceptual judgement but one involving some non-empirical process, say, doing simple addition and reporting the sum. Suppose that the subjects behaved similarly(not an obvious result, but a possible one if the task assigned were sufficiently simple and the subjects sufficiently , but not too, mathematically competent). We would no more conclude that the process the subjects followed in doing addition was faulty than we would conclude from the Asch experiment that the subjects' visual perception faculties were faulty.

My proposed modification is meant to suggest that what the existence of the

---

[55] [Kitcher 1984], p.56
[56] [Asch 1956], p.47

social challenge comes to is this: People can have experiences which suggest the falsity of their a priori beliefs; but then again, people can have all sorts of misleading experiences. That this is possible is a fact about our psychology, not a fact about the mechanisms we use to arrive at knowledge, for it can occur regardless of the mechanism used. That I can be deceived about a proof I have followed does not mean that I do not know it a priori. If I have engaged in a non-empirical process resulting in knowledge of the theorem proved, then I can be said to know it a priori. In my last section I will discuss what I see as Kitcher's misunderstanding of the notion of proof which has led to his stringent requirements on a priori knowledge.

## 1.4 Final Comments

We have seen from earlier discussion that Kitcher has very strict requirements on what constitutes a proof. In this section I would like to show how his conception of proof overlooks precisely what we think is important about proofs, what sets them apart from other types of reasons for having beliefs. I believe that the apriorist can acknowledge the importance of Kitcher's epistemology without having to sacrifice classical views on proof.

Recall what Kitcher's notion of proof entails. A proof is a sequence of sentences, each of which is either a basic a priori statement or results from above sentences through the application of an apriority-preserving rule of inference. For Kitcher, following a proof is undergoing a series of transitions which generate knowledge of the theorem proved.

Obviously Kitcher is not defining a notion of actual proof, the kind found in mathematics texts and taught in mathematics classes; he is talking about an ideal notion of proof. There is certainly a distinction in mathematics between proofs that we in practice see and do and the notion of rigorous formal proof. The former are sketches of the latter, abbreviations which give us the idea of how to go from one step to the next and so on to the conclusion. Many steps are left out for the sake of brevity. If we had to formalize every proof, we would not be able to do much mathematics, for the process of making a proof rigorous is arduous and time-consuming. However, we do know how to formalize a proof to make it rigorous. We also have criteria of correctness for formal proofs.

Mathematician Saunders Maclane echoes this view about proof:[57]

> An absolutely rigorous proof is rarely given explicitly. Most verbal or written mathematical proofs are simply sketches which give enough detail to indicate how a full rigorous might be constructed. Because of the conviction that comes from sketchy proofs, many mathematicians think that mathematics does not need the notion of absolute rigor and that the real understanding is not achieved by rigor.

He goes on to say that, despite some dissenting views, the notion of absolute rigor plays an important role in mathematics. Kitcher acknowledges that proofs are almost never written out formally; he says that some theorems in analysis "never receive general proofs which are rigorous even by the standards of informal rigor which mathematicians accept."[58] He admits that formal proofs would make the process of following a proof enormously difficult, making it much harder to generate a priori knowledge. So Kitcher is not advocating a strictly formal notion of proof for two reasons: 1) it is not in keeping with our standard mathematical practice; and 2) in the interests of charity to his opposition, he does not promote a notion of proof even more prone to the pitfalls he sees on the way to a priori knowledge of theorems. Kitcher rightly characterizes proofs as serving some function. The problem lies in his explication of what function they do serve. Not only is a proof a proof in a formal theory, it is an argument designed to convince us of the truth of the theorem proved.

[57] [Maclane 19xx]
[58] Kitcher, p.26.

If you understand and believe the premises, then your belief in the conclusion is justified, it is warranted– that is, your reasons are sufficient to allow you to draw the conclusion. And your belief is justified because you arrived at it in the *right* way.

Kitcher says that following a proof is engaging in a series of transitions which generate knowledge. There are a number of problems with this view. We can say, indeed we should say something much stronger than that the above transitions generate knowledge. The transitions should actually compel belief of the theorem proved. If you understand and believe the premises, you ought to believe the conclusion. Proofs have considerable normative power - they allow us to infer ought from is . Proofs confer entitlement; they allow us to infer statements if we go about it using the appropriate rules of inference and the right axioms. Also, the transitions have to generate knowledge on certain ways; not just *any* situation in which I draw the conclusion is a case in which I acquire a priori knowledge of the conclusion.

Not only does Kitcher misconstrue the job proofs do, he attacks apriorism in a peculiar way. His strategy is as follows. He says that following a proof is engaging in a psychological process which generates a priori knowledge of the theorem proved. But, he adds, if the psychological process does not warrant belief against a backdrop of misleading experience, then our knowledge of mathematics is not a priori. Since the process does not generate knowledge under averse conditions(under social, possibly theoretical challenges) then our knowledge of mathematics is not a priori. Kitcher selects a candidate process, pure intuition, which purportedly warrants knowledge, but falls prey to such challenges. He concludes that no psychological process will generate a priori knowledge of theorems of mathematics.

Here is where Kitcher has made a serious error. He concludes that since he has found a process which is not an a priori warrant, no process can serve as an a priori warrant. The real problem at work here is that Kitcher has failed to specify sufficient conditions on a priori warrants. The apriorist, however, has an account of proof which suggests which processes would qualify as warrants. It is not a specific list of conditions, but rather a line of argument she can follow to reject Kitcher's views, since the burden of proof is on him to show her how traditional processes fail to warrant a priori knowledge.

Consider some rule of inference, say, modus ponens. If any rule of inference is apriority-preserving, surely modus ponens is. What happens when you apply the rule is this: you come to believe $A$, you come to believe $A \rightarrow B$. Then, modus ponens allows you to move from these two premises to conclude $B$. Is it the case that believing $A$ and $A \rightarrow B$ always generates knowledge of $B$? Probably not. It could happen that I believe $A$, $A \rightarrow B$, and not conclude $B$(suppose I get hit by a bus before I get the chance to make the inference). Or, I believe $A$, $A \rightarrow B$, and then something happens(lightning strikes, or there is an earthquake) and I come to believe $B$. In neither of these cases did the psychological processes I engaged in result in knowledge of $B$; in the first case it was blocked by my untimely demise, and in the second case I was distracted from following the process by a natural disaster.[59] But do these cases show that modus ponens is not apriority-preserving? Of course not. All they show is that not all psychological processes generate knowledge. The apriorist says that there are processes which work, and what we do when we use modus ponens to go from $A$, $A \rightarrow B$ to $B$ is a prime example.

Kitcher has made the situation look bleak by never constructing even a prima facie plausible example of a candidate process. In our examples above, if some process does not allow us to go from $A$, $A \rightarrow B$ to $B$, then maybe we have picked a bad process. We should not conclude that no process allows us to make the inference; we now that using modus ponens is exactly the right way to do it. While it is true that a process must generate knowledge by some kind of transition, not just any kind of transition will do. The process must generate knowledge in the right way. And it

[59] I am indebted to George Boolos for discussion of these examples

is up to Kitcher, not the apriorist, to provide such an account. Since Kitcher has not only failed to provide necessary conditions for knowledge, but also lacks sufficient conditions, the apriorist may reject his attack on the apriori nature of mathematical knowledge.

# Chapter 2

# Church's Thesis: a Case Study for Lakatos' Philosophy of Mathematics

## 2.1 Introduction

The late Imre Lakatos put forth a view in *Proofs and Refutations* that rejects standard views about mathematics; in particular, he rejects standard accounts of proof and how proof conveys mathematical knowledge. According to him, classical accounts of epistemology and foundations of mathematics do not capture what is special about mathematical practice, and what we come to know by doing mathematics. Standard views attribute to mathematical axioms special status, e. g. a priori, analytic, necessary. He says those are the wrong kinds of classifications; we should look at mathematical knowledge, mathematical proof, and mathematical theories in a completely different way. The traditional classifications above tell us little about how we actually come to know mathematical statements. What Lakatos in interested in is shifting our focus from the structure of proofs and theories to the processes by which mathematical statements come to be accepted or rejected.

Lakatos motivates his alternative view by citing what he sees as the failures of formalism, logicism, and "inductivism" in the foundations of mathematics. Various attempts to lay the foundations of mathematics in logic, or in inductive proofs and definitions, or in provably consistent formal systems all met with serious and sometimes insurmountable problems. His reaction to these problems is to jettison the standard distinctions that are thought to set mathematics apart from the sciences. Lakatos thinks that we should distinguish mathematical theories from scientific theories by looking at how they are verified or falsified. The difference between them will be in the nature of their falsifiers.[1]

Lakatos sets up his alternative taxonomy and attempts to show that mathematics is not verifiable, but rather is conjectural, falsifiable, and subject to refutations. Therefore, for him, certainty in mathematical knowledge is impossible. If this is the case, then how does he explain what proofs do? After all, following proofs is how we standardly acquire mathematical knowledge, knowledge which is considered to be a priori and certain. Lakatos responds by distinguishing between what he calls "informal" proofs and formal proofs. For him, the real work of mathematicians is properly done within the realm of *informal* mathematics, where theories are tested, refuted, refined, expanded, and applied to new areas. While formal proofs cannot be refuted, they also do not expand our knowledge of mathematics by pointing to new areas of research. Informal proofs, which do not have the standards of rigor required for formal proofs, may contain assumptions that point to "hitherto unthought of possibilities"[2] and new insights.

In order to make his case, Lakatos will have to explain what appears to be strong evidence against his claim that formalization does not increase fields of mathematical

[1]As a student of Popper, Lakatos was heavily influenced by the notions of verification and falsification of theories. However, we will focus discussion on Lakatos' use of those notions, which diverges from Popperian classifications.

[2][Currie and Worrall 1978], p.69.

inquiry. For example, the systematization of axiomatic set theory had an enormous impact on modern mathematics. Transfinite induction arose out of this program of systematization. Lakatos does not think that formalization serves merely hygenic purposes, as he recognizes that formal proofs have a degree of certainty that informal proofs lack; rather he considers the program of formalization to be the least fruitful of mathematical enterprises, as it gives rise to no refutations.

Since it is in the informal theory where we find refutations, the plausibility of his philosophy of mathematics rests on his account of what kinds of refutations or falsifiers mathematics is subject. We will need to examine Lakatos' account of falsifiers to see if mathematics is conjectural.

Lakatos' charges are serious ones, deserving of a thorough response. He concludes his attack on traditional accounts by suggesting a view which undermines both the entire epistemological and metaphysical structure and the methodologies behind standard philosophies of mathematics. Since the cost of making these changes is so high, the burden of proof is on Lakatos to provide an alternative explanation of mathematical knowledge. What kind of explanation he provides, how plausible it is, and how well it applies to actual cases in mathematics will determine how much of a threat Lakatos' view poses to the apriorist philosopher of mathematics.

Once the stage is set, then, as good Popperians we should test his theory. We will look at a well-known thesis in mathematics—Church's Thesis—and consider what would count as refutations of it. Lakatos mentions that Laszlo Kalmar's criticism of Church's Thesis is a rare and notable case of someone taking seriously the possibility of refutations in mathematics.[3] One might ask *why* such cases are rare; if there are few cases which conform to Lakatos' picture of the epistemological structure of mathematics, then he should either say why that is the case or give other evidence for his view.

[3][Currie and Worrall 1978], p.42.

Lakatos' view certainly does not require that he explain all phenomena in mathematics. But, if his theory yields quite counterintuitive results in the intended cases, or fails to give a coherent account of how actual results in mathematics are known or are revised, then the apriorist has not been given adequate reason to give up the position that proof conveys knowledge which is certain.

## 2.2 Lakatos' View of Mathematics as Fallible

### 2.2.1 Some Preliminaries: Terminology, Taxonomy

Lakatos begins an article on the foundations of mathematics[4] with a discussion of scepticism as it applies to the philosophy of mathematics. He says that skeptics use the question "how do you know?" to try to show that there is no foundation for knowledge. They keep asking "how do you know?" to establish that there is an infinite regress in all knowledge claims, that "any rational effort to obtain knowledge is powerless."[5]

In mathematics, we come to know statements by following proofs. But how do we know that the proofs actually prove anything? Lakatos says that to prove that a proposition is true, "foundationalists" must establish that *something* (e. g. an axiom) is true, and must also establish some way to transfer truth from proposition to proposition (e. g. rules of inference). A way to answer the sceptic is to construct a system with true axioms and rules of inference that take us only from true propositions to true propositions.

According to Lakatos' rational reconstruction of the history of epistemology, the foundationalists developed three ways to try to fight scepticism and establish a firm foundation for knowledge. All three ways involved developing what he calls "deductive

[4] *Infinite Regress and the Foundations of Mathematics, in [Currie and Worrall* 1978*], pp.3-23.*
[5][Currie and Worrall 1978], p.4.

systems". He describes the basic characteristics of deductive systems below:[6]

> Deductive theories [are distinguished by] a *principle of retransmission of falsity* from the 'bottom'[(conclusions)] to the 'top'[(premises)]; a counterexample to the conclusion will be a counterexample to at least one of the premises... [Also] a *principle of transmission of truth* holds from premises to conclusions. We do not demand, however, from a deductive system that it should transmit falsehood or retransmit truth.

Lakatos does not explain how transmission of truth or retransmission of falsity are supposed to work in general. It is likely that different systems will transmit truth or retransmit falsity in different ways, depending on the subject matter. For example, the rule modus ponens presumably would transmit truth in an axiomatic system of mathematics, although how it does so would have to be spelled out. Retransmission of falsity is harder to characterize– we can imagine observations which cause us to question generalizations in some theory, but how to characterize that process in general is far from obvious.

Lakatos identifies what considers he considers to be three major programs designed to create a foundation for knowledge. They are not exhaustive, but are mutually exclusive and represent three ways of organizing knowledge into deductive systems:[7]: 1) Euclidean (henceforth referred to as EUCL); 2) Empiricist (also called quasi-empirical, henceforth referred to as QE); and 3) Inductivist (henceforth referred to as IND). Any mathematical or scientific theory could fall under one of these three categories.

Lakatos says that EUCL theories are distinguished by the existence of (generally well-known) axioms at the top, with *infallible truth-value injections* of the truth value *True* , which flows downwards through deductive channels (generally via proof).

[6] [Currie and Worrall 1978], p.4.
[7] [Currie and Worrall 1978], p.4.

The channels do not admit proof of anything other than true theorems: "Since an [EUCL] theory contains only indubitably true propositions, it operates neither with conjectures nor refutations".[8]

Lakatos does not provide an example of a EUCL theory or how truth "flows" downwards through the system. However, it would seem to follow that first-order logic is at least a good candidate EUCL theory, with modus ponens and other rules of inference providing a way to transmit truth through so-called deductive channels. ZF set theory would seem to be another possible EUCL theory; however, he has not given us enough information to be able to identify particular mathematical theories as EUCL or not. Also, he offers no explanations of how a sample truth-value injection works in a specific EUCL theory.

For Lakatos, QE theories consist of propositions at the bottom (which he calls basic statements, a term borrowed from Popper) from which "there is a possibility of *infallible truth-value injection*... which, if the truth-value is False, flows upward through the deductive channels (explanations) and inundates the whole system."[9] QE theories in general contain either conjectural or demonstrably false propositions.

Again, Lakatos does not give an example of a QE theory, nor does he explain how a sample "injection" of falsity affects propositions further up in the system. We can imagine a situation in which the falsity of some key observation statement might cause us to reject some generalization in a scientific theory. But, we do not know how in general the falsity of a proposition is supposed to "inundate the whole system."

Lakatos spends less time discussing IND theories, as it is the contrast between EUCL and QE theories that he considers most relevant for philosophy of mathematics. He says that IND theories differ from the other two in the following way. They are distinguished by a truth-value injection of *True* which flows *upwards* from the basic statements. This pattern of flow is called the *principle of retransmission of truth*,

[8] [Currie and Worrall 1978], p.4.
[9] [Currie and Worrall 1978], p.5.

through which the system is inundated. His explanation of IND theories suffers from the same problems as do his accounts of EUCL and QE theories.

While it is true that we can construct cases e. g. in which some true observation statement contributes to the confirmation of a generalization, we are left not knowing what Lakatos considers to be the canonical examples of this kind of theory or how he thinks the truth of a statement can "inundate" a system with truth. His characterizations of all three sorts of theories lack the concreteness required to apply them in specific cases. Lakatos plans to use his distinctions to show that mathematics is QE, but so far we have little information with which to classify actual theories in this way. The metaphor of flowing and inundation of truth or falsity does not *explain* how propositions in an actual theory are proved or disproved.

Lakatos claims that the patterns of transmission of truth values are independent of *how* the truth-values are determined in a particular theory. For example, a QE theory is not necessarily an empirical one– that is determined by the nature of the basic statements of the theory. The patterns of flow are even independent of what flows through the system, e. g. truth or falsehood, probability or improbability.[10] Of course Lakatos must qualify this claim, for it is not the case that a "pattern of flow" could transmit say, axiomhood from axioms to theorems in a system.[11]

QE and EUCL theories differ markedly in how they develop over time. Lakatos describes three stages of EUCL theory development:

1. the prescientific stage, a period of trial and error;
2. the foundational period, which serves to demarcate the boundaries of the theory;
3. the application period, during which problems inside the system are solved.

Lakatos contrasts the above pattern with that of a QE theory, which

[10] [Currie and Worrall 1978], p.29.
[11] I am indebted to George Boolos for this observation.

starts with problems followed by daring solutions, then by severe tests, refutations. The vehicle of progress is bold speculations, criticism, controversy between rival theories, problemshifts. Attention is always focussed on the obscure borders. The slogans are growth and permanent revolution, not foundations and accumulation of eternal truths.[12]

Lakatos engages in polemics against EUCL methodology, calling it "puritanical and antispeculative"[13]. Lakatos also considers the IND program a failure, but it is EUCL theories that are his real target. A large part of the motivation behind his fallibilism is his interest in showing that attempts to preserve the Euclidean status of mathematics are wrong-headed and futile. Lakatos takes careful aim at logicism and formalism, citing what he sees as the failures of Frege, Russell and Hilbert to shore up the foundations of mathematics with logic or with satisfactory consistency proofs.

The problems of what Lakatos calls the *Frege-Russell* approach are with the axioms of the system– they are not indubitably true, and in the case of Frege's system they were not even consistent. Lakatos describes a controversy during which various proponents of the EUCL program fought to establish axioms via various methods, e. g. set-theoretical, constructivist, or logical intuition. Using logical intuition seemed a promising method for setting up a EUCL mathematical theory, for as Lakatos points out, "whoever wins the battle for the axioms, logical intuition has to be relied upon to carry truth from the top to the remote parts of the system."[14] He says that this move will satisfy skeptics, for even they have to rely on logic to criticize the foundationalists. "...to show that all mathematics does not need any other *but* logical intuition will certainly be a huge gain: there will be only one single source of certainty both for the axioms and for the truth-transmission."[15]

[12] [Currie and Worrall 1978], p.30.
[13] [Currie and Worrall 1978], p. 29.
[14][Currie and Worrall 1978], p.12.
[15] [Currie and Worrall 1978], p.13.

Lakatos reports that Russell's attempt to "trivialize"(Lakatos' term) mathematics by reducing it to logic failed. Type theory contained axioms that were perhaps true, but certainly not indubitably true, e. g. those of choice, reducibility, infinity. According to Lakatos, the axioms *explain* the theorems, but they do not *prove* them.[16]

Given that Russell's program failed, what are we entitled to infer about all EUCL programs? Lakatos would have us believe that all such programs are epistemologically bankrupt. However, maybe we just lack the satisfactory axioms for capturing the truths of mathematics. It is possible that Russell's system is a EUCL one, but his particular attempt failed.

So, if Russell's system is in fact a QE system, it should be testable. If this is the case, then how do we test it? Lakatos notes that in Russell's system, all the theorems are derivable in it, so there do not seem to be any potential falsifiers. He suggests that we test the system for consistency. If Russell's system turns out to be inconsistent, then it would definitely be a QE theory. If there is no way to show that it *is* consistent, then Lakatos' case for the QE status of mathematics might be strengthened. This concern leads him to look at the another major attempt to provide a foundation for mathematics, formalism.

Lakatos says that Hilbert's program was designed to end skeptical worries about foundations by showing the following:[17]

1. all arithmetical propositions which are formally proved (the arithmetical theorems) will certainly be true if the formal system is consistent, in the sense that $A$ and $\bar{A}$ are not both theorems;

2. all arithmetical truths can be formally proved;

3. meta-mathematics, [a] new branch of mathematics set up to prove the consis-

[16] [Currie and Worrall 1978], p.19. Of course Russell was aware of this fact. Axioms do not show theorems to be true; they are no more evident than the theorems they prove.

[17] [Currie and Worrall 1978], p.20.

tency and completeness of formal systems, will be a particular brand of Euclidean theory: a finitary theory, with trivially true axioms, containing only perfectly well-known terms, and with trivially safe inferences.

Lakatos quotes Hilbert, who contends that "arithmetical truth—and, because of the already accomplished arithmetization of of mathematics, all sorts of mathematical truths—will rest on a firm, trivial, 'global' intuition, and thus, on 'absolute truth' ".[18] Lakatos concludes that a consistency proof will thus show that mathematics will have *no* falsifiers[19].

It should be noted here that Lakatos' description of Hilbert's program is misleading. He states that the goal of the program was a consistency proof for arithmetic. In fact, this was *not* Hilbert's goal, but rather a by-product of his program. What is important to show is not that arithmetic is consistent, but rather that the use of set theory to prove facts about arithmetic does not result in new theorems; what can be proved with the use of set theory should be provable *without* it.

The complete story of the failure of Hilbert's program is an interesting and complex one. What is important to note here is that Gödel's second incompleteness theorem showed that that it is impossible to prove the consistency of arithmetic using only finitary methods. Many important results came out of reactions to that failure, in particular Gentzens' non-finitary proof of the consistency of arithmetic up to $\epsilon_0$, using transfinite induction. He draws the conclusion from all this historical evidence that mathematics is undeniably quasi-empiricist in nature.

### 2.2.2 Mathematics is Quasi-Empiricist

Lakatos' arguments for the quasi-empiricist nature of mathematics are mainly negative ones. He maintains that major attempts to show that it is either EUCL or

[18] [Currie and Worrall 1978], p.20.
[19] [Currie and Worrall 1978], p.32.

IND failed. So by process of elimination, it must be QE. But what does it mean for mathematics to be QE? Lakatos says that QE theories are conjectural and falsifiable. So far, we do not have a good idea of what it means for a theory to be falsifiable– how does the falsity of some statement in the theory affect the status of statements further up? We have not seen examples of QE theories; we have not seen any explanations of how particular statements can "falsify" particular theories. Lakatos' view needs to be filled in with examples from specific cases if we are to see how falsification works.

On standard views, formal mathematical theories, if consistent, are not falsifiable. It would follow that nothing would count as a potential falsifier for a consistent mathematical theory. We can imagine the existence of potential falsifiers in science; basic statements, observations like "the reading on the meter was 3.5", may undermine some hypothesis. However, in mathematics there do not seem to be any obvious candidates. Lakatos says that both mathematics and science are QE; the difference between them is in the nature of their falsifiers.

There are, according to Lakatos, two kinds of falsifiers in mathematics– logical and heuristic. He states that logical falsifiers are statements of the form $p\&\neg p$. He does not explain how they work, but he does mention an example of a logical falsifier. He claims that Frege's system was 'refuted' by Russell's discovery of a logical falsifier. Systems can reveal inconsistencies that stem from incorrect axioms or faulty rules of inference. It is true that most people would agree that mathematics (Lakatos speaks of comprehensive axiomatic set theories in particular) is subject to logical falsification. A theory has been logically falsified if one finds a contradiction that can be proved in the theory.

Lakatos argues that logical falsifiers do not seem to capture the kind of falsification done by what he calls the 'hard facts'. He seems to be looking for some way to do in mathematics what we do in science when we use observation statements to test hypotheses. He says if we limit our scrutiny to formal theories, then we will find only logical falsifiers.

Lakatos proposes another way to falsify mathematical theories: "if we insist that a formal theory should be the formalization of some informal theory, then a formal theory may be said to be 'refuted' if one of its theorems is negated by the corresponding theorem of the informal theory. One could call such an informal theorem a *heuristic* falsifier of the formal theory."[20]

While the notion of a logical falsifier is a familiar one in classical mathematics, the definition of 'heuristic falsifier' relies on an unfamiliar distinction—the difference between formal and what Lakatos calls 'informal' mathematical theories. We will examine this distinction later. First, let us examine a scenario in which he describes a potential heuristic falsifier.

Take set theory as a sample mathematical theory. It is testable if it is a QE theory, but *how* can it be tested? Lakatos suggests two ways to criticize a set theory. One way is to test the axioms for consistency, looking for logical falsifiers. Another (more subtle) way is to test the definitions for "'correctness' of their translation into branches of mathematics like arithmetic."[21] It is unclear what Lakatos means here, but we will see that the latter test is a search for a heuristic falsifier. Consider the following scenario:[22]

Suppose that we have a formal proof in formal set theory whose intended interpretation is that there exists a non-Goldbachian even number.[23] Suppose further that a number theorist proves *informally* that all even numbers are Goldbachian. If his proof can be formalized within set theory, it will count as a logical falsifier, for the theory will have been shown to be inconsistent. But, Lakatos notes that if the informal

[20] [Currie and Worrall 1978], p.36.

[21] [Currie and Worrall 1978], p.36.

[22] [Currie and Worrall 1978], pp.36-37.

[23] A non-Goldbachian even number is one which is not the sum of two primes. Goldbach's Conjecture asserts that all even numbers are sums of two primes. It has not yet been proven, although it has been confirmed for a large number of cases.

proof cannot be thus formalized, the set theory will not be shown inconsistent, "but only to be a *false* theory of arithmetic. The theory is false in respect of the informal *explanandum* that it had set out to explain." Lakatos does not explain further, but he could mean that since the informal proof has the force of a convincing argument without the rigor of a formal proof, it shows that there is a problem with the formal theory—namely, that it does not explain some fact demonstrated informally. To remedy the problem, Lakatos suggests we check the definitions (in this case the definition of 'natural number' may be suspect) and adjust the definitions to accommodate the heuristic falsifiers.

Lakatos concludes that, as a result of these adjustments, we find that the formal theory is no longer useful as an explanation of arithmetic; the only way to restore its usefulness is to eliminate *all* heuristic falsifiers. It seems as if informal proof has provided an observation that is at odds with the formal theory. In *Proofs and Refutations*, Lakatos suggests two techniques for dealing with heuristic falsifiers:[24] 1) "monster-barring"—rejection of the informal proof on the grounds that it is not really a falsifier of the formal theory; 2) "lemma incorporation"—make adjustments to the formal theory to accommodate the new fact shown by the informal proof. But why does he say that the existence of a heuristic falsifier renders the theory useless? That is an extreme conclusion, especially considering that monster-barring is an option.

Lakatos' description of the Goldbach's Conjecture case leaves a number of questions unanswered. He mentions that the informal proof cannot be formalized—why not? Most proofs are not written in the language of set theory; they are abbreviations of formal proofs. But, given any proof written in this abbreviated style, we can translate it into a formal proof. If we could do that in the GC case, then we would have found a contradiction in set theory. But if we cannot translate it into a

[24]Lakatos has the view that heuristic falsifiers work both against informal and formal theories. He does not distinguish who they work in the different cases. 1) and 2) here are used in *Proofs and Refutations* to work against informal theories, but there is no reason to think that they do not work against formal theories as well.

formal proof, we have to ask *why*. Central to the notion of proof is the fact that it can be tested for correctness by formalizing it and checking it using some mechanical procedure.

Lakatos has presented a curious case; we are left unsure how to react to his scenario. We *know* what it means for something to be a proof in a formal theory—roughly, $A$ is a proof in a formal theory $T$ iff it is a sequence of steps, each of which is either an axiom of $T$ or follows from an earlier step via some accepted rule of inference in $T$. There are standards of rigor for formal proofs— given a sequence of steps we can tell whether it is a proof in $T$. If the informal proof of Goldbach's Conjecture could be formalized, then it would count as a logical falsifier of the formal theory, showing the theory to be inconsistent.

What we do *not* know is what an informal proof is. Since we have no idea *why* it cannot be formalized, only *that* it cannot be formalized, we can exercise one of the following options:

1. We can accept the informal theorist's proof as a rival hypothesis, which may give rise to a new formal theory. We attribute its status as informal to the fact that it is part of a theory which has not been formalized *yet*, but will be. This option classifies the new proof as a logical falsifier in progress.

2. We can reject the informal proof and conclude that the informal theorist has simply made a mistake somewhere. Since it does not meet our standards of rigor for proofs, we need not accept it. Therefore it is not a falsifier at all.

3. We can accept the proof as a heuristic falsifier, but without some criteria to judge correctness of informal proofs and some guidance as to what to do with the formal theory in response to it, this option does not provide a direction for the formal theorist to follow.

The problem with the Lakatos' Goldbach's Conjecture example is that it is radically underdescribed. We do not know what the differences are between informal theories and proofs and formal theories and proofs. We do not know how to judge validity of informal proofs, or even if validity is an appropriate term to apply to them. In order to make such decisions, the situation needs more detailed description— Lakatos needs to *give* us an example of an informal proof and explain how mathematics progresses in the face of heuristic falsifiers.

In *Proofs and Refutations*, Lakatos shows how Euler's Theorem[25] changed as a result of heuristic falsification.

One important thing to note here is that unlike the imaginary Goldbach's Conjecture case, the Euler's Theorem case pre-dated formalization. Until the axiomatization and formalization of mathematics took place (around the late nineteenth and early twentieth centuries) there were no standards of rigor for proofs. It was not until a formal language for mathematics was systematized that standards of absolute rigor in mathematics were possible.

Since then, formalization has played an important role in mathematics. It is not one in which we are *required* to translate all proofs into formal proofs. Formalization is used to make clear what have been obscure or ambiguous notions. For example, formalization of the calculus helped clarify the theory of infinitesimals. It helped make explicit the order of quantifiers. We need not formalize all of mathematics, but we use it to help uncover, clarify and guarantee correctness in troublesome areas of mathematics.

Lakatos discusses the differences between informal and formal proofs in an article entitled "What Does a Mathematical Proof Prove?"[26] For him, all proofs are in one of the following three categories:

[25]Euler's Theorem says that for any regular polyhedron, $V - E + F = 2$. A full treatment of its development is found in [Lakatos 1976].

[26] [Currie and Worrall 1978], pp.61-69. The editors of this volume note that Lakatos changed his mind about some of the points in the paper and did not plan to publish it.

1. pre-formal proofs

2. formal proofs

3. post-formal proofs

Both 1. and 3. are informal proofs. 2. includes standard formal proofs in classical mathematics—finite sequences of statements, each of which is an axiom or follows from a previous statement via some accepted rule of inference. Formal proofs have decision procedures for determining whether some given sequence is a proof. Lakatos says that informal proofs, on the other hand, do not admit of such procedures. He adds that we should not think that an informal proof is merely a formal proof with gaps or suppressed premises; it is not just an incomplete formal proof.

Lakatos cites Euler's Theorem as an example of an informal theorem. It is accepted as a proof, but it contains no postulates and no obvious way to formalize this reasoning.[27] According to him, the proof is a convincing argument that intuitively shows the theorem to be true.

There are no specific criteria for correctness of informal proofs. One can show that something is *not* an informal proof by "pointing out hitherto unthought of possibilities."[28] In the Euler's Theorem case, the possibility that a polyhedron could have a hole in it[29] constitutes a possible falsifier. One can incorporate the counterexamples into the theorem by expanding the concept of a polyhedron. Or, one can limit the concept of polyhedron so to restrict the counterexamples.

None of the falsification are on formal grounds—they are all on the level of what he calls the pre-formal theory. He does not elaborate on what pre-formal theories are. Lakatos does say that pre-formal theories are subject to formalization, which

[27]This view is Lakatos'—Mark Steiner (find ref.) discusses the formalisation of Euler's Theorem through the development of algebraic topology, resulting in a rigorous demonstration of it.

[28] [Currie and Worrall 1978], p.65.

[29]The picture frame counterexample is one of these cases.

may yield "unfortunate results." If we formalize too early, he warns, we limit the subject matter and possibly exclude from consideration new objects of study:

> While in an informal theory there really are unlimited possibilities for introducing more and more terms, more and more hitherto hidden axioms, more and more hitherto hidden rules in the form of new so-called 'obvious' insights, in a formalized theory imagination is tied down to a poor recursive set of axioms and some scanty rules.

In a general pronouncement on formal versus informal theories, Lakatos says:

> [Informal proofs] prove something about that sometimes clear and empirical, sometimes vague and 'quasi-empirical' stuff, which is the real though rather evasive subject of mathematics. This sort of proof is always liable to some uncertainty on account of hitherto unthought of possibilities. [Formal proofs are] absolutely reliable; it is a pity that it is not quite certain—although it is approximately certain—what it is reliable about.[30]

Lakatos seems worried that formalization will restrict the class of mathematical statements we can come to know. He says little about how informal theories are structured, and less about what informal proofs are. If informal proofs have no axiomatic structure, then what structure does he have in mind? Lakatos is trading on the fact that we know exactly what constitutes a formal proof in a formal theory. Instead of providing a clear picture of what informal proofs in informal theories look like, he says that they are *not* like formal theories in certain ways, and that the proofs are *not* restricted in the way formal proofs are. What he fails to give us is any clear *positive* picture of the structure or process of informal mathematics.

Lakatos writes that informal proofs are falsified by hitherto unthought of possibilities. But what counts as such a possibility? Surely we are not required to take every

[30] [Currie and Worrall 1978], p.69.

possible objection seriously, for that would cripple mathematics– no work would ever get done, no proofs would ever be finished. Lakatos realizes that there are no standards of correctness for informal proofs, but he considers that a virtue, for he says it allows us to expand the fields of examination to include bold speculative ideas.

An important question, however, remains: how can we tell the difference between a legitimate flaw in an informal proof and an irrelevant objection? This is a problem for all theories, but in the case of mathematics, one can reply to many objections by pointing to the relevant section of a proof and showing how it is justified.

What Lakatos seems to ignore is that rejecting standards of correctness for proofs undermines one of the major reasons for proofs in mathematics: they serve to convince us of the truth of the theorem proved. Of course, intuition plays a role in accepting a proof—one must agree to the truth of the axioms and the validity of the rules of inference—but what proofs serve an important justificatory role for us. To rob proofs of this most important function of bestowing certainty is to take away most of the power of proofs, This move incurs a great cost for those who are interested in the growth of mathematical knowledge.

Lakatos is willing to bear the costs involved, for he seems to think that having standards of rigor for formal proofs will exclude proofs that he considers to be enlightening. He does offer some restrictions on informal proofs; in informal mathematics, for a proof to be rigorous or valid, there must be no heuristic falsifiers for it. Recall that according to the Principle of Retransmission of Falsity, the falsity of some statement will affect the status of statements further up in the system. He does not provide an account of how this retransmission is supposed to work; there are no specific criteria for recognizing or applying falsifiers.

Lakatos does admit that it is possible for the flow of refutations to stop, at which point we will have reached truth. "But of course we shall not know when. Only refutations are conclusive—proofs are a matter of psychology."[31]

### 2.2.3 Fallibilism as a Philosophy of Mathematics

Lakatos' diatribe against formalism ends in an extreme conclusion—that we reject the possibility of certainty in mathematics, and embrace a fallibilist stance. His conclusion comes as a result of finding that all the major deductive anti-skeptical programs in mathematics—logicism, formalism, and even intuitionism—share the same flaw: they reject criticism too early in order to pursue justifications, which he thinks is antithetical to fostering growth in mathematics. "Different levels of rigor differ only about where they draw the line between the rigor of proof-analysis[32] and the rigor of proof, i. e. about where criticism should stop and justification should start."[33]

In a famous and controversial passage from *Proofs and Refutations*, Lakatos claims that certainty in mathematical knowledge is impossible. " 'Certainty is never achieved'; 'foundations are never found'—but the 'cunning of reason' turns each increase in *rigor* into an increase in *content*, in the scope of mathematics."[34] What makes his comments controversial is how much he discounts the role absolute rigor has played in the development of mathematics. The editors carefully note that, whereas fallibilism with respect to the axioms of mathematics is a reasonable position for well-known reasons, there is no sense in which a rigorously correct proof is fallible; we have methods for checking to see if something is a proof in a formal system. This is a crucial point to which we shall return later.

To make his position more explicit, Lakatos tries to compare what a formal proof adds with what an informal proof adds to our knowledge. Formalism definitely adds something to the certainty of the theorem proved– it guarantees that there will not be

[31] [Lakatos 1976], p.53.
[32]In [Lakatos 1976], the term 'proof-analysis' is used to mean the methods of introducing and incorporating heuristic falsifiers into the informal theory.
[33] [Lakatos 1976], p.56.
[34] [Lakatos 1976], p.56.

a counterexample formalizable within that system. But that is no guarantee against heuristic falsifiers, which must be taken into account. Furthermore, he maintains that a formal proof gives us no guarantee that the formal system has "the empirical or quasi-empirical stuff in which we are really interested and with which we dealt in the informal theory."[35]

Lakatos notes that almost no one has studied the possibility of refutations in mathematics; he mentions Laszlo Kalmar as an exception. We will examine Kalmar's attempt at a refutation in the next section. Lakatos maintains that we cannot take fallibilism seriously without taking the possibility of refutations seriously.[36]

We will take up Lakatos' challenge and consider a possible refutation in mathematics. Church's Thesis is a key assumption of computation theory which identifies the intuitive notion of effective calculability with the mathematically rigorous notion of partial recursive function. It is not a theorem, since it cannot be proven. But, there is considerable evidence that it is true. Using Church's Thesis as a sample case, we can entertain the possibility of heuristic falsifiers for computation theory in the form of arguments against Church's Thesis. Although Lakatos' views need not apply to all mathematics (he points out that not all branches are equally subject to heuristic falsification), computation theory presents a promising opportunity to fill out the details of his alternative philosophy of mathematics. If his fallibilistic philosophy fails to offer a plausible alternative account of mathematical knowledge, that gives the classical philosopher of mathematics less reason to give up the view that formal proofs are primary conveyors of certainty in mathematical knowledge.

---

35 [Currie and Worrall 1978], p.67.

36 [Currie and Worrall 1978], p.42.

## 2.3 Church's Thesis– A Case Study for Fallibilism

### 2.3.1 Introduction

We saw in the previous section that Lakatos criticized what he saw as the major attempts to provide a foundation for mathematical knowledge. He correctly points out that the problem with inductivism is that it does not provide us with a direction after heuristic falsification; it is not as if the categories were already set up, ready for us. In the case of logicism and formalism, we know that there is more to mathematics than proving theorems.

Lakatos depicts heuristic falsification as an activity that is not accounted for within the process of systemization of mathematics. If we apply Lakatos' suggestions to Church's Thesis, then we should expect to see exemplified there the kind of informal mathematical structure, replete with heuristic falsifiers, that Lakatos points to in mathematics. He cites Kalmar as an example of someone who tried to take on Church's Thesis. Kalmar attacked the half saying that all effectively calculable functions are recursive; he thought the idealization was too restrictive. We will see that in Kalmar's case, his argument is defective, although the reasons motivating his arguments bear addressing.

Rosza Peter also challenged Church's Thesis, attacking the other half—the half that states that all recursive functions are effectively calculable. Peter does not accept that what we can do *in principle* is at the heart of Church's Thesis. Of course Turing and Church were *not* interested in what we can do in practice; they wanted to characterize a notion of effective calculability that abstracts from our physical limitations. However, Peter's objection brings up an important question: is there some notion of effective calculability which is an abstraction of our human capabilities, but is *not* our current view? It is with this question in mind that we will examine

Kalmar's objections to the plausibility of Church's Thesis.

In the Euler's Theorem case Lakatos successfully shows the complex interplay between what can be proved and what is speculated about. Explaining or accommodating the cylinder and picture frame counterexamples is a matter of adjusting our terminology and expanding or contracting the theory depending on the costs (e.g. clarity, explanatory value, naturalness, applicability to a large number of cases).

In the Kalmar case we will not turn up anything promising in the way of a heuristic falsifier or conjecture. Of course, that Lakatos' view does not conform to logicism is no drawback. Also, the fact that no promising heuristic falsifiers have turned up in the Church's Thesis case is not necessarily a point against Lakatos—maybe we have found the correct characterization of effective calculability, which we can now clarify both formally and philosophically; we give a formal explication of the notion using the model of a Turing machine. We can then use that model as an example of what we can do in principle, abstracting away from certain specific physical limitations. A close examination of the model may make clear what we mean by "what we can calculate in principle".

However, we should ask at this point whether there is a perspective from which we can see heuristic falsification not as the main stage, but just as a phenomenon that occurs from time to time. That is, we can see heuristic falsifiers as the occasional by-products of the general program of systematization, an activity fundamental to mathematics.

Lakatos has taken the business of conjectures and refutations as absolutely central to mathematics; recursion theory would be a special case of having gotten it right; the same would be true for the Euler's theorem case, as it has been successfully formalized in algebraic topology. The appearance of heuristic falsifiers is considered a symptom of this phenomenon. However, we will see that heuristic falsifiers can indeed be accounted for within the realm of formal mathematics– they occur as a part of the process of formalization. They are an expected if occasional result of mathematical clarification, a process whose common results are proofs.

### 2.3.2 Church's Thesis

The case I intend to examine is the well-known and well-established Church-Turing Thesis, (hereafter referred to as Church's Thesis). Church's Thesis identifies the intuitive notion of *effectively calculable functions* with the notions of *partial recursive* or *Turing-computable*, or *$\lambda$-calculable functions*, all of which have rigorous, precise mathematical definitions.

Why should we want to try to identify an intuitive notion with a mathematically rigorous one? Robin Gandy[37] points out 3 motives for trying to give a precise definition to a vague and intuitive notion:[38]

- The intuitive notion may be clearly defined in some contexts; one may wish to extend the definition to a wider range of contexts.
- One may be able to get greater precision and/or a wider range of application with a precise definition—it gives more power, more positive results; e.g. the extension of 'integer' and 'prime' from rationals to other algebraic number fields.
- If one wishes to obtain negative results to show that something is *not* true of the notion, then one must give a definition of that notion so as to delimit its extent.

Gandy says that there are problems "which in some special cases can be settled by calculation, but for which a uniform general computational method of solution seems unlikely"[39]

[37] [Gandy 1989], pp.55-111.
[38] [Gandy 1989],p.56.
[39] [Gandy 1989],p.60. He cites as an example diophantine equations, i.e. equations for which the only valid solutions are integers.

New methods for solving various problems, developed in the late 19th and early 20th centuries, involved "only known processes of calculation such as could be performed in principle, by [Babbage's] analytic engine". [But] one might speculate that some as yet undiscovered conceptual framework for a decision might require some as yet undiscovered process of calculation."[40]

Such speculation raises an important question: how do we know that we will not discover new methods that solve non-recursive or non-Turing-computable functions? The arguments in favor of Church's Thesis will have to be partially empirical in nature, since the thesis asserts facts about human computational capacities. We will examine those arguments shortly.

In 1934, Alonzo Church identified "effectively calculable" with "$\lambda$-definable". His choice seemed like a good one, as the $\lambda$-calculus was a powerful system and all the $\lambda$-definable functions were effectively calculable. Gödel had the view that Church's Thesis was not subject to proof. Gandy reports that Gödel wrote in 1934 that the half of Church's Thesis asserting that all effectively calculable functions are $\lambda$-definable seems to be true. However, "...This cannot be proved, since the notion of finite computation is not defined; but it serves as a heuristic principle."[41]

There are a number of standard arguments in favor of Church's Thesis:

1) Many calculable functions have been shown to be recursive. Many natural classes of functions, e.g ones in elementary number theory, turn out to be recursive.

2) No one has produced a calculable function which cannot be shown to be recursive, or even suggested a plausible method for constructing such a function. Gandy offers a slight variant of this argument, called the "criterion of the failure of diagonal arguments": he notes that Kleene was unable to diagonalize out of the class of $\lambda$-definable functions.[42]

---

[40] [Gandy 1989],p.62.
[41] [Gandy 1989],p.72.
[42] [Gandy 1989],p.78.

3) Many methods of obtaining calculable functions from calculable functions have been shown to lead from recursive functions to recursive functions. Furthermore, no such method has led from a recursive function to non-recursive function.

4) Various definitions have been proposed for the calculable functions; in each case, all the functions have been shown to be calculable, and the definitions have turned out to be equivalent.

These argument are not conclusive; they do not constitute a *proof* of Church's Thesis. As we saw above, Gödel thought that recursiveness and effective calculability could not be satisfactorily identified "except heuristically". Certainly argument 2) serves a heuristic purpose; but, it could happen that some genius will discover an entirely new sort of calculation that outstrips the class of recursive functions. We encounter the same problem with 3): some new algorithm might proceed by steps which were not recursive. The equivalence argument also does not provide a proof—it is possible we have formalized the wrong notion in trying to capture effective calculability.

Shoenfield[43] points out that we can *almost* prove Church's Thesis; the problem with proving it is that we have no precise definition of effective calculability. However, we *were* able to prove by induction on recursive functions that every recursive function is calculable.[44]

Can we prove the converse—that every calculable function is recursive—in the same way? According to Shoenfield, the difference between the two cases is the following: in the former, we isolated properties of calculable functions and predicates which were obvious, even from the vague descriptions of calculable functions. However, no one has isolated the properties of calculable functions needed to prove Church's Thesis.[45]

[43] [Shoenfield 1967]
[44] [Shoenfield 1967],p.109.
[45] [Shoenfield 1967],p. 120.

If we try to prove Church's Thesis, we would try to define the notion "calculable" directly. We assume a single calculable function F. From F you apply a simple operation to get a function G, which is H(F), where H is recursive. If G is recursive, then so is F. So, if F is calculable, so is G.

What we want to show here is that applying recursive operations on calculable functions does not lead outside the realm of calculable functions. Furthermore, we want to show that if G is recursive, then the functions it is derived from are also recursive. Of course, in this argument G is assumed to be calculable, which makes the argument circular. However, G results from a single step in the calculation, so it must be a very simple calculable function; therefore, it is likely to be recursive. If we assume this, then we can *prove* that F is recursive. However, since we cannot *prove* that G is recursive (however obvious it may seem), we have no proof.

Since the evidence for its truth falls short of rigorous proof, Church's Thesis is commonly considered an *explication* of the notion "effectively calculable function"; on this interpretation it has actual empirical content. Church's Thesis does bring up an interesting question: what *is* the appropriate idealization for our human calculation abilities? In some cases, we know that a proposed idealization has outstripped our capabilities– e.g. if we allow for infinitely many steps in a calculation, then everything is computable. However, there do seem to be intermediate cases; we can recognize sentences of English, but the set of English sentences is *not* known to be effectively calculable.[46]

Turing wondered what are the possible processes which can be carried out in computing a real number. He was interested in considering the potential abilities of a *computor*, a calculating agent whose capabilities are not subject to certain physical limitations. On Turing's model of calculation, a computation proceeds by discrete steps and produces a record consisting of a finite (but unbounded) number of cells,

[46]I am indebted to Jim Higginbotham for this observation.

each of which is blank or contains a symbol from a finite alphabet, as each step the action is local and is locally determined, according to a finite table of instructions.[47]

According to Gandy, Turing's analysis *proves the following theorem*: Any function which is effectively calculable by an abstract human being following a fixed routine is effectively calculable by a Turing machine– or equivalently, effectively calculable in the sense defined by Church, and conversely.[48]

Gandy thinks that Turing's work has settled the matter completely—"it shows that what appears to be a vague intuitive notion has in fact a unique meaning which can be stated with complete precision."[49]

Gödel acknowledged the importance of Turing's characterization: " It seems to me that [the] importance [of the concept of Turing-computability] is largely due to the fact that one has for the first time succeeded in giving an absolute definition of an interesting epistemological notion, i.e. one not depending on the formalism chosen".[50]

### 2.3.3 Philosophical Arguments in Favor of Church's Thesis

Before we consider an informal argument against Church's Thesis , let us look at one philosophical argument in its favor, and a somewhat radical characterization of Church's Thesis.

#### Mendelson: Church's Thesis or Theorem?

Elliott Mendelson[51] offers an analysis of Church's Thesis which he thinks gives us reasons to believe that it is a theorem. He considers it "completely unwarranted to say that Church's Thesis is unprovable just because it states an equivalence between

[47] [Gandy 1989],p.81.
[48] [Gandy 1989],p.81.
[49] [Gandy 1989],p.86.
[50] Gandy does note that Gödel argued that in our ability to handle abstract concepts we are not subject to the restrictions described by Turing.
[51] [Mendelson 1990]

a vague, imprecise notion (effectively computable function) and a precise, mathematical notion (partial recursive function)."[52] If we had enough facts about effective calculability to prove Church's Thesis the way Shoenfield suggests, then Mendelson would be absolutely right. We shall see if he can make such a case.

On standard views, Church's Thesis is an explication in Carnap's sense. That is, it replaces some intuitive notion with a precisely defined one which may or may not extend beyond the original notion. Furthermore, "confirmation of the correctness of the [explication]... [is not a matter of proof but] apparently must involve, at least in part, some empirical investigation."[53] Mendelson disagrees with this account of Church's Thesis. He could offer what would be a natural opposing view—that the two notions simply turn out to be coextensive, but differ in meaning. On this view, verifying Church's Thesis would not involve complicated analysis of what the terms mean and how they are used, how they are interrelated. Mendelson suggests that Church's Thesis is really a theorem. offering as evidence for his view an analysis of several mathematical theses which are clearly well-accepted parts of mathematics; on his view, Church's Thesis deserves the same status. He considers the following four "theses" as well-accepted as Church's Thesis:

1. the identification of the intuitive notion of function with the definition of function in terms of a set of ordered pairs satisfying the condition $If (y,x) \in f\, and\, (z,x) \in f,\, then\, y = z$

2. the identification of the intuitive notion of truth in a language with Tarski's set-theoretic definition of '**B** is true in **M**', for any sentence **B** and a structure **M** for a language **L**

3. the identification of the intuitive notion of logical validity with the model-theoretic definition of logical validity—that a first-order sentence is logically

[52] [Mendelson 1990],p.232.
[53] [Mendelson 1990], p.229.

valid if it is true in all structures

4. the identification of the intuitive notion of limit with the $\epsilon$–$\delta$ definition of a limit of a function and the corresponding definition for a limit of a sequence

According to Mendelson, the notion of effectively calculable function is, in certain ways, just as precise as that of partial recursive function; the latter notion just happens to be more familiar and formally defined.[54] In the case of the definition of function in terms of set, he says that the notion of set is not clearer than the notion of function. Likewise for the other examples of equivalence above: we have shown that we can replace one notion with another, but we have gotten nothing in the way of improvements in clarity or intuitive appeal.

Mendelson offers another reason in favor of giving Church's Thesis status as a theorem. He says, "the assumption that a proof connecting intuitive and precise mathematical notions is impossible is patently false."[55] It is obvious that one-half of Church's Thesis is true– the half that says all partial-recursive functions are effectively computable. He provides a short argument for its truth[56]. Of course the starting functions (e.g. addition) are computable; there are ways to describe easy procedures to compute them. And the operations of substitution and recursion and the least-number operator also result in computable functions; again we can describe procedures to compute such functions. Clearly, if a function is partial recursive, then there is an *algorithm* to compute it. So one-half of Church's Thesis has been established to everyone's satisfaction. Mendelson says that the fact that this proof is not in ZF "just shows that there is more to mathematics than appears in ZF."[57]

Mendelson is here restating Shoenfield's point we noted earlier: that the use of the predicate "is calculable" prevents us from being about to *prove* Church's Thesis for

---

[54] [Mendelson 1990], p.232.
[55] [Mendelson 1990] p. 232.
[56] [Mendelson 1990], p.232.
[57] [Mendelson 1990], p.233.

the following reason: there are not enough obvious facts about effective calculability to isolate the properties of calculable functions from which it would follow that all effectively calculable functions were recursive. However, he makes the further point that it does not matter that we cannot specify their properties well enough to formalize them in ZF, for we can *know* without the help of ZF that the properties of the effectively calculable functions are exactly those of the recursive functions.

Mendelson's third point is perhaps his most controversial. It is an expansion of the above point about proof. He thinks that underlying the standard views regarding Church's Thesis is the opinion that the only way to ascertain the truth of a statement in mathematics is to *prove* it. He points out that proofs assume axioms and rules of inference; also, many equivalences like the ones he cites as theses 1–4 above are "often simply *seen*(my emphasis) to be true without proof, or are a mixture of such intuitive perceptions and standard logical and mathematical reasoning."[58] What he seems to be saying here is that we should expand the language of mathematics to include as theorems statements that are true but not proven in ZF; we can increase the number of mathematical truths by means other than proving them in a formal system.

The four theses Mendelson cites are, in certain superficial ways, like Church's Thesis. Each thesis states an equivalence between two notions, and none of them are verified by proofs. The evidence given in favor of them seems to be informal or intuitive rather than rigorous, or they are treated more as definitions than theorems. Unlike Church's Thesis, though, they are not explications. Showing them to be explications in Carnap's sense would require that they have some empirical content; the investigation to settle the matters would involve asking questions outside of the realm of mathematics. In the case of Church's Thesis it is clear that such questions are appropriate. We want to know what is the correct idealization of our calculating

[58] [Mendelson 1990], p. 233.

abilities. Determining the answer will at the very least require information about our *actual* calculating abilities, which is clearly outside the realm of mathematics.

In the four cases Mendelson describes, we will see that verifying them does not require such extra-mathematical investigation. If that is the case, then there is good reason to think that Mendelson's four theses are *not* explications.

Consider first the Weierstrassian definition of limit. It is *not* the case that we had an clear intuitive notion of limit which was replaced by a more rigorous notion. Before we had a formal notion of limit of a function or a sequence, there was no uniformly acceptable way to explain concepts like continuity. Mathematicians in the eighteenth century used the vague term 'infinitesimal' to explain such phenomena.

The $\epsilon$–$\delta$ definition made possible the solution to many problems which had, up to that time, no satisfactory explanations. Weierstrass's well-known results, along with others' work, provided the formalism to show that some infinite sequences converge to finite limits. The definition was accepted and is now included as a standard part of every elementary calculus textbook.

There is a difference between learning what a partial recursive function is and learning what a limit is. In the first case, we are replacing a prior intuitive notion with a rigorous one. In the case of limits, there is a prior intuitive notion (the geometric one) and a rigorous notion ( the analytic $\epsilon$–$\delta$ notion). For some questions of the form "Does $\lim_{x\to a} f(x) = y$?", we will be able to give a conclusive answer if we use the analytic notion, but *not* if we use the geometric one.

However, the asymmetry does not hold in the case of Church's Thesis. We could consider ourselves as having two notions of computability— the intuitive one (effective calculability) and the rigorous one (partial recursivity). But, there are no cases that come to mind in which a function can clearly be found to be partial recursive, but there is no clear verdict about whether it is effectively calculable. How can we explain this apparent asymmetry?

If we want to analyze Church's Thesis as an explication, we could say one of two things: 1) it is analogous to the case of limits but that the arithmetization of analysis has no counterpart in recursion theory; or 2) the two cases are not analogous. It seems to me that they are *not* analogous; the other cases Mendelson discusses are about about mathematical notions relative to a given formalism. Effective calculability is not just a mathematical notion; effectively calculable is a property independent of the formalism used. Effectively calculable *means* computable by humans.

Mendelson stated that one of the properties of an explication is that it cannot be proven; its correctness is a matter of fit, which involves linguistic as well as empirical study.[59] However, mathematicians do not speculate about whether the formal notion of limit is an apt replacement for the informal one. Formally defining the notion of limit helped to tie the calculus to the arithmetic of real numbers, which led to further foundational work on number theory.[60] For these reasons, the status of thesis 4 is quite different than that of Church's Thesis.

We have seen that thesis 4 cannot be interpreted as an explication; neither can thesis 1 be interpreted as identifying the notion of function with that of set of ordered pairs. What thesis 1 assures us of is that the notion of function is definable for all purposes within set theory, that the notion of function is not an additional notion. But, it certainly does not follow that people believe that a function *is* a set a ordered pairs in the same way they believe that the calculable functions *are* the partial recursive functions.

Mendelson asserts that in theses 1–3 the notions being defined are in some ways no

---

[59]It has been suggested that some explications might be provable—a promising candidate is Frege's definition of the ancestral in terms of parent. From facts about the parent relation, along with mathematical facts about ancestral induction, perhaps we could actually *prove* the equivalence of Frege's definition with the intuitive notion. As Shoenfield pointed out, if we could reduce Church's Thesis to a minimal set of assumptions about human computation and apply simple mathematical operations on the starting functions, it might be possible to prove it. However, in the case of Church's Thesis, it is not clear how to specify the set of intuitive assumptions about effective computability.

[60] [Davis and Hersh 1981] p. 246.

more vague than the definitions. He mentions that the notion of set cannot be thought of as clearer than that of function. But thesis 1 just asserts that what we can do with functions we can also do with sets; the notion of function is not primitive within set theory. Even Mendelson does not think that set theory completely characterizes the general notion of function.

Also, it is far from obvious that the foundations of set theory are just as vague as our intuitive notions having to do with functions. While there may be some obscurity associated with treating a collection of many objects as one object, a function is an object that takes *arguments* and returns *values* in some way or other; those notions are in many ways less well-understood than the notions of set and member.

Thesis 3—the identification of the intuitive notion of logical validity with the model-theoretic notion—shows that our intuitive notions of validity coincide with the formal ones. According to our informal notions, whatever is provable is valid; also, whatever can be made false is non-valid. Our intuitive notions, along with the completeness theorem, allow us to prove thesis 3. However, in the case of Church's Thesis, we do not have the sufficient facts about intuitively calculable functions that, combined with formal apparatus, allow us to prove Church's Thesis.

Mendelson asserts that one-half of Church's Thesis is obviously true—all partial-recursive functions are calculable. The starting functions are all calculable; there are simple procedures to compute them. It follows from the inductive definitions of the starting functions that the operations of minimization and composition, when applied to calculable functions, yield calculable functions.

Why does Mendelson think that no one doubts the half of Church's Thesis that says that all recursive functions are calculable? Answer: the argument is considered trivial. But why? Rosza Pèter objected to Church's Thesis on the grounds that the characterization of effective calculability was too sweeping, that there were recursive functions that were not effectively calculable.

The reason this view is not widely shared is not easy to pinpoint, but we know that if a function is recursive, it comes from either one of the starting functions, which we recognize as effectively calculable, or it comes from applying one of the extremely simple operations of minimization or composition to an effectively calculable function. We cannot give a sense to the idea that applying one of those steps to something calculable results in something *not* calculable.

We standardly consider the problematic half of Church's Thesis to be the part stating that all functions calculable in the intuitive sense are partial recursive. Finding a function that is calculable in the intuitive sense but is not partial recursive would falsify Church's Thesis.

Mendelson offers a final point in his argument that proof is not the only way in which we come to accept the truth of a statement. It is true that axioms are not proven. Giving a satisfactory story of why they *are* accepted is a difficult task, one which I will not attempt here. But he is not maintaining that Church's Thesis should be considered an axiom, merely a theorem. Some equivalences between intuitive notions and rigorous mathematical ones need not be constrained by the requirements of proof, but "often are simply 'seen' to be true without proof, or are based on arguments that are a mixture of such intuitive perception and standard logical and mathematical reasoning."[61]

Perhaps Mendelson is suggesting that we should expand the language of mathematics to include statements containing predicates that have not been defined in ZF, like effectively calculable. The statement "if f is partial-recursive, then f is effectively calculable" will be a mathematical statement, but not one provable in ZF, since the predicate "effectively calculable" is not defined in ZF. The problem with admitting that statement as a theorem is that the evidence for it (e.g. that we are subject to finiteness and other restrictions in our calculating abilities) is circular. The question

[61] [Mendelson 1990], p.233.

keeps presenting itself to us because we are not sure what creatures like us are capable of computing. If it turns out that we do indeed have the computational abilities that Church and Turing proposed, then perhaps Church's Thesis is more like a theorem. However, to say this may commit Mendelson to expanding the realm of mathematics more than is reasonable; admitting psychology or biology into mathematics (since the evidence for Church's Thesis may involve those areas) is imprudent at best.

## Post's View: A Psychological Interpretation

Emil Post maintained that Church's Thesis should be treated not as a definition, but rather as an empirical claim about the limits of the formalizing powers of humans. He says in his 1936 paper[62], that Church's developments form a "working hypothesis", although he thinks that Church's Thesis has progressed far beyond the hypothesis stage. However, he adds the following caveat in a footnote:

> To mask this identification under a definition hides the fact that a fundamental discovery in the limitations of the mathematicizing power of Homo Sapiens has been made and blinds us to the need of its continual verification.

Enderton[63] says that Post considered Church's Thesis to be more of a natural law than an axiom or definition. Post was interested in distinguishing what can be done in mathematics by purely formal means from the work which depends on understanding and meaning. He believed that a true account of human mathematical intelligence must be non-mechanical:[64] "Mathematical thinking must be essentially creative: postulational thinking will then remain as but one phase of mathematical thinking." Gandy notes that both Post and Gödel believed that a satisfactory theory of mathematical intelligence must take account of creative and non-finitary reasoning.

---

[62] [Post 1936], p.105.
[63] [Enderton 1977]
[64] [Gandy 1989],p.93.

Church responded to Post's characterization, saying, "Since effectiveness in the ordinary sense hasn't been defined, the working hypothesis does not have an exact meaning. Defining effectiveness to be computability by an arbitrary machine subject to finiteness restrictions is a good definition; if we do this, we have no need for a working hypothesis."[65]

Church's response ignores Posts's main point, which is that since the thesis is an idealization of actual computing powers of humans, we should focus at least part of our attention on determining what we can do in practice, not just in principle. The idealization is already somewhat restrictive, as the model prohibits use of infinitely many steps in calculations. It is still an open question whether the idealization is correct.

We will now turn to an attempt to present a potential refutation using philosophical arguments.

### 2.3.4 An Argument Against Church's Thesis

Laszlo Kalmar[66] also considers Church's Thesis to be more of an explication, not a theorem in formal mathematics subject to proof or disproof, since it identifies two notions, only one of which is has a rigorous mathematical definition. He considers most arguments for or against it to be *pre-mathematical.* Kalmar argues against its plausibility not by giving a counterexample, but by presenting what he considers to be strange consequences of one half of Church's Thesis.

Kalmar attacks the half of Church's Thesis asserting that every effectively calculable function is general recursive[67]. He focuses on the ramifications of assuming the contrapositive—that all non-general recursive functions are not effectively calculable.

Consider the following function $\psi(x)$:

[65] [Gandy 1989],pp.85-86.
[66] [Kalmar 1956]
[67]Kalmar uses this terminology instead of partial-recursive.

$$\psi(x) = \mu_y(\varphi(x,y) = 0) \begin{cases} \text{the least natural number y for which} \\ \varphi(x,y) = 0 \text{ if there is such a y} \\ 0 \text{ if there is no natural number y such} \\ \text{that} \varphi(x,y) = 0 \end{cases}$$

Kalmar says that $\psi(x)$ is an example of a non-general recursive function, and $\varphi$ is some appropriate general recursive function of 2 arguments. He maintains that the supposition that $\psi$ is not effectively calculable "has strange consequences." He provides an explanation of what he means below.

Kalmar analyzes $\psi(x)$ as follows:[68]

1) For any natural number p for which $\exists y\, \varphi(p,y) = 0$, then there is a method to compute the least such y, i. e. $\psi(p)$: compute $\varphi(p,0), \varphi(p,1), \varphi(p,2)$, etc. (possible since $\varphi$ is recursive) until you get a $q$ such that $\varphi(p,q) = 0$. In this case $\psi(p) = q$.

2) But, for any p for which we can prove, "not in the frame of some fixed postulate system but by means of arbitrary—of course, correct—arguments that no natural number y such that $\varphi(p,y) = 0$ exists, we also have a method to calculate $\psi(p)$: prove that no natural number $y$ with $\varphi(p,y) = 0$ exists, which requires ... a finite number of steps," whose result is that $\psi(p) = 0$.

Kalmar claims that his analysis is based on the definition of the function and use of the law of excluded middle- no other assumptions are made.

Kalmar concludes from the fact that $\psi$ is not effectively calculable and applying the law of excluded middle, "we infer the existence of a natural number $p$ for which, on the one hand, there is no natural number $y$ such that $\varphi(p,y) = 0$, on the other hand, this fact cannot be proved by any correct means, a consequence of Church's Thesis which seems very unplausible."[69]

When Kalmar speaks of a "method" for calculation, he says he is not assuming

[68] [Kalmar 1956],p. 74.
[69] [Kalmar 1956],p.74.

it to be "uniform".[70] In order for a method to count as a decision procedure or algorithm, it must *not* depend on the inputs. It must output a value for any input, not just for some inputs. Whether the "method" Kalmar suggests constitutes a decision procedure is the crucial to his case against Church's Thesis. We will discuss this question in detail later.

Kalmar states that the proposition that for some natural number $p$, there is a natural number $y$ such that $\varphi(p, y) = 0$, would be undecidable, "not in Gödel's sense of a proposition neither provable nor disprovable in the frame of a fixed postulate system... but *not even admitting any correct means.*" The fact that Church's Thesis identifies recursivity with effective calculability by any correct means shows for Kalmar that the above proposition is undecidable in a "really absolute sense".[71] However, he argues that this absolutely undecidable proposition is really decidable after all—it is false:

> ...this "absolutely undecidable proposition" has a defect of beauty: we can decide it, for we know it is false. Hence, *Church's Thesis implies the existence of an absolutely undecidable proposition which can be decided, viz. it is false,* or in another formulation, *the existence of an absolutely unsolvable problem with a known definite solution,* a very strange consequence indeed.

Kalmar does qualify his result by saying that this consequence cannot be proved by any correct means since it would have to contain a proof of the undecidability of the proposition plus a proof of its negation, which is impossible.[72]

How could it be possible that we can know of the falsity of an undecidable proposition? If the proposition that for some natural number $p$, there is a natural number

[70] [Kalmar 1956],p. 73.
[71] [Kalmar 1956],p. 75.
[72] [Kalmar 1956],p. 75.

$y$ such that $\varphi(p,y)=0$ is undecidable, that means that there is no single algorithmic procedure that, given the value of $p$ as input, will output a $y$ such that $\varphi(p,y)=0$. In order to be in a position to assert the falsity of the above proposition, we must use a method of computation that outputs its negation. Knowing that we will never have a effective procedure to decide the proposition is not sufficient. Kalmar says that we can "see" its falsity.

Kalmar further claims that even the undecidability of the proposition in question cannot be proved by any correct means. He argues for this conclusion by considering a general proposition of the form $\exists y\ \mathbf{P}(y)$ with a general recursive predicate, hence effectively decidable property $\mathbf{P}$.

Suppose $\exists y\ \mathbf{P}(y)$ is true. Then, since $\mathbf{P}$ is decidable, there is some $q$ such that $\mathbf{P}(q)$, and this $q$ can be found in a finite number of steps. So it would follow that $\mathbf{P}(q)$ holds and $\exists y\ \mathbf{P}(y)$ *can* be decided.[73]

However, if a proposition of the form $\exists y\ \mathbf{P}(y)$, with $\mathbf{P}$ recursive, is undecidable, then it does *not* hold. "Hence, if the undecidability of that proposition could be proved, then the negation of that proposition could be proved too. Thus, the proposition could be decided, so it would not be undecidable; but that is impossible if only correct means are allowed.[74]

Kalmar concludes, "the fact that some of the consequences of Church's Thesis cannot be proven by any correct means is an argument against its plausibility."[75]

Kalmar, citing another paper of Church's on a related topic[76], suggests that Church's Thesis is a challenge "to find, instead of the class of general recursive functions, either a less inclusive class which cannot be shown to exclude some function which ought reasonably to be allowed as effectively calculable, or a more inclusive class which cannot be shown to include some arithmetical function that cannot be

[73] [Kalmar 1956],p.75.
[74] [Kalmar 1956],p.76.
[75] [Kalmar 1956],p.76.
[76][2]in [Kalmar 1956], p.80.

seen to be effectively calculable."[77] He proposes to answer Church' challenge in the following way: add to the class of general recursive functions all the arithmetical functions $\psi(x)$ defined by an equation of the form $\psi(x) = \mu_y(\varphi(x,y) = 0)$ with a general recursive function $\varphi$ of two arguments.

To calculate $\psi(p)$ in a finite number of steps, he advises using the following method: Calculate in succession $\varphi(p,0)$, $\varphi(p,1)$, $\varphi(p,2)$ ...and simultaneously try to prove by all correct means that none of them equals 0, until we find either a $q$ for which $\varphi(p,q) = 0)$ or a proof of the proposition stating that no natural number $y$ with $\varphi(p,y) = 0)$ exists. the result of the calculation with be either some $q$ in the first case, or it will be 0 in the second case.

Kalmar acknowledges that he is not presenting a disproof of Church's Thesis, for there is not an actual *proof* that $\neg\exists y\ \varphi(p,y) = 0$, for any given $p$. He does, however, consider the above argument to be a challenge to the defender of Church's Thesis, since he offers what he thinks is a *method* for calculating $\psi(p)$. However, his method relies on the notion of arbitrary correct means, which is not a mathematical notion; hence, his arguments against Church's Thesis are pre-mathematical rather than mathematical.

Kalmar presents all of these arguments in service of his view about the status of the concepts related to effective calculability. He summarizes his view below:

> There are premathematical concepts which must remain premathematical ones, for they cannot permit any restriction imposed by an exact mathematical definition. Among these belong, I am convinced, such concepts as effective calculability, or of solvability, or of provability by arbitrary correct means, the extension of which cannot cease to change during the development of Mathematics[78].

[77] [Kalmar 1956],p.76.
[78] [Kalmar 1956],p.79.

What is crucial to examine here is what Kalmar's view comes to. He is not offering a disproof of Church's Thesis, but he does seem to be saying that Church's Thesis restricts what we can count as correct methods of proof. No doubt there are proof techniques that have not been discovered yet. But for Kalmar's case to be convincing, he has to offer a method of calculating $\psi(x)$ that is an actual algorithm; this means that the method must fall within certain restrictions. Any method of computation for p must not depend on p; no matter what p is[79], the procedure must output a value.

In order to make his notion of proof by arbitrary correct means explicit, he would have to do the following: pick some appropriate postulate system, adopt the proposition $\exists p \neg\exists y\ \varphi(p,y) = 0$ to the new system. But how do we prove the consistency of the new system? Kalmar claims that "we [would] have to prove (by some correct means) the verifiability of the new postulate..." So there is no improvement here.

Kalmar responds to this complaint, saying "The consistency of most of our formal systems is an empirical fact...why do we not confess that mathematics, like other sciences, is ultimately based upon, and has to be tested in, practice?"[80]

I would like to make a few formal observations about undecidability that will show how Kalmar's view is misguided. Let us return to the function $\psi$. It is the characteristic function for the following set, call it $A$.

set $A = \{p : \exists y\ \varphi(p,y) = 0\}$

$A$ is not recursive, but it is recursively enumerable, i. e. for any $p$, if $p \in A$, we have an single algorithmic procedure that outputs precisely the members of $A$. But, for all $p \notin A$, there is no single algorithm whose output is exactly those things that are not members of $A$, i. e. the members of $\bar{A}$, the complement of $A$. So for any $p \notin A$, there is no guarantee that we can ever find an effective procedure to show that $p \notin A$.

[79]Of course, p must be an input of the appropriate sort, say a number, if the algorithm computes some arithmetic function.

[80] [Currie and Worrall 1978],p.27.

As noted above, it follows from the fact that $A$ is r. e. for any $p$, if $p \in A$, then we have a effective procedure to show that $p \in A$. But, there are also some $p \notin A$ such that we can show that $p \notin A$– we have a number of methods at our disposal.[81] But which method we use to show that $p \notin A$ will depend on $p$. Different methods will allow us to show, for different values of p. that $p \notin A$. These methods do *not* count as algorithms, for algorithms do not depend on the inputs. One method of finding for *all* $p \notin A$ that $p \notin A$ is simply to list the members of $\bar{A}$. However, that method requires an infinite number of steps, which violates conditions on algorithms.

Kalmar's general argument says that from the undecidability of $\exists y\ \mathbf{P}(y) = 0$ we should be entitled to infer that there is no $q$ such that $\mathbf{P}(q)$. But that does not follow from the undecidability of $\exists y\ \mathbf{P}(y)$ . We *may*, through a combination of insight and luck, find a $q$ such that $\mathbf{P}(q)$. All the undecidability of $\exists y\ \mathbf{P}(y)$ shows is that we cannot prove for *any* given $q$ that $\mathbf{P}\ (q)$.

Kalmar thinks that the formalism of computation theory excessively limits our notion of effective calculability. Assuming that a non-general recursive function is not effectively calculable does *not* have strange consequences—we have seen that shown formally.

A perhaps more interesting question to consider is why Kalmar views the formalism as so restrictive, if not for technical reasons. He mentions that arguments against the plausibility of Church's Thesis are philosophical in nature. From what he has said it appears that he views the boundaries between classes of calculable functions, e. g. recursive, recursively enumerable, as arbitrary, not useful, or unnatural. But if that is the case, the Kalmar should give an argument stating *why* they are unsatisfactory. Unsurprisingly, he gives no such argument, for it would require giving reasons for his objections that are presumably motivated by some alternative view of computability.

[81] We know that there are $p \notin A$ for the following reason: all the $p \notin A$ are in the complement set of $A$, called $\bar{A}$. $\bar{A}$ is called co-r. e. (co-recursively enumerable), which means that its complement is *not* recursive and its complement is r. e. . Since $\bar{A}$ is not recursive, we know it is non-empty (since the empty set is recursive). All non-empty sets have finite subsets, so $A$ and $\bar{A}$ have finite subsets.

This would constitute a major breakthrough in mathematics.

Of course it is possible that we will discover different classes for calculable functions. Work in complexity theory has already created many structures for classifying problems. But this new field does not threaten the pre-existing structure of computation theory; it is a supplement to it, increasing our knowledge and understanding of notions related to calculability. We now have a new vocabulary of terms, with many new directions to explore.

Key notions in mathematics do change over time and get refined or discarded through new discoveries. And along with the notions, the formal systems which support them also get revised, through addition or deletion of axioms and especially introduction of new proof techniques, definitions. However, just to say that the extension of concepts like effective calculability "cannot cease to change during the development of mathematics" is to take an untenable position. Without alternative directions, we are left with no coherent way of explaining the mathematical phenomena we set out to understand in the beginning. The classical mathematician has no reason to abandon what is at this point a fruitful, powerful, truth-conveying set of structures. It is unreasonable to expect Kalmar to some up with new mathematics to justify his claim, but likewise it is unreasonable to expect anyone to jettison useful structures, without which there is no obvious mechanism for any growth in mathematical knowledge.

But what *is* central here is the process of mathematical systematization; In the case of computation theory, formalization identifies an intuitive notion—effective calculability—with a number of equivalent rigorous notions—partial recursive function, Turing-computable function. Whether we have accurately captured the former notion by defining it as one of the latter ones is a legitimate question. Formalization cannot help us answer it, but it can give us tools to increase our knowledge of the latter notions. If they outstrip or diverge from the original intuitive ones, then depending on what directions we wish to pursue, we may adjust the formalization or accept that the intuitive notion has replaced or improved upon by the formal one.

## 2.4 Concluding Remarks: The Plausibility of Fallibilism as a Working Philosophy of Mathematics

We have seen that in order to take Lakatos' views seriously, we have to consider the possibility of refutations in mathematics. This, in turn, requires that he give a sensible account of falsifiers for theories. Lakatos' account requires the existence of heuristic falsifiers in informal mathematics coming into conflict with systematic exposition in formal mathematics. We have, however, failed to discover any sense in which there are such falsifiers.

The history of mathematics contains many accounts of logical falsifiers; for example, theories have been shown to be inconsistent due to faulty axioms. It is the slippery concept of a heuristic falsifier that makes fallibilism less coherent, for it relies on a distinction that has never been made clear. Furthermore, Lakatos states explicitly that although not just anything can count as a heuristic falsifier, there are no standards of correctness for the informal proofs which are the candidate heuristic falsifiers. If the criteria for what count as heuristic falsifiers are not only partially dependent on fuzzy notions but deliberately kept vague, then it is difficult to make a plausible case for fallibilism.

Lakatos bases his position on a reaction to what he sees as the failures of inductivism, logicism, and particularly formalism. Lakatos is right to point out that there is more to mathematics than creating a formal system for the proving of theorems. Certainly judgments are important in picking new axioms to introduce, new definitions to incorporate; sometimes new results trigger such moves, forcing us to change our formalism.

In the case of computation theory, formalization identifies an intuitive notion—effective calculability—with a number of equivalent rigorous notions—partial recursive function, Turing-computable function. Whether we have accurately captured the former notion by defining it as one of the latter ones is a legitimate question. Formalization cannot help us answer it, but it can give us tools to increase our knowledge of the latter notions. If they outstrip or diverge from the original intuitive ones, then depending on what directions we wish to pursue, we may adjust the formalization or accept that the intuitive notion has replaced or improved upon by the formal one. According to this picture of the development of mathematics, heuristic falsification is not a process that is orthogonal to formal mathematics, but rather one which is *crucial* to it.

Lakatos does not suggest that informal mathematics is indiscriminate, sloppy mathematics. But the lack of any real guidelines as to what are criteria for these crucial notions force the classical philosopher of mathematics to reject his view out of hand. Fallibilism is an extreme view, with serious ramifications throughout philosophy. Unless he makes a compelling case, we need not give up proof as a way of conferring certainty. Although he has reminded philosophers of the lessons we learned about the limits of formalization, within those limits there are myriad possibilities for expanding the base of knowledge and moving in new directions.

# Chapter 3

# Surveyability and the Four Color Theorem

## 3.1 Introduction

Mathematics has come a long way since the Pythagorean theorem. Not only have we opened up new fields of study, but mathematics has undergone drastic changes in what count as appropriate methods of proof.

While intuitions often drive the directions we take, we are still constrained by restrictions on correct methods for doing proofs. Traditionally, one of the prime characteristics of proofs is that one can follow a correct proof so as to arrive at a priori knowledge of the theorem proved. According to some accounts of mathematical knowledge, the process of following a proof meets the constraints imposed on processes that purport to confer a priori knowledge. I would not presume to try give an account of exactly what those constraints are or ought to be; however, we ought to be able to look at some new mathematical practices and see if they fall outside the bounds what we used to consider constraints on appropriate methods for doing proofs.

For several decades, computers have been used to help ease the computational burdens on those doing difficult and lengthy calculations. They have been used, for example, in number theory to generate data on the distribution of primes. In applied mathematics, we use computers to generate approximate answers to problems, and the degree of precision will depend on the task at hand.

The Four Color Theorem is an example of a theorem that was proved using the computational help of a computer. The Four Color theorem states that every map can be colored using at most four different colors so that no two neighbors are colored alike[1]. What makes this theorem worthy of philosophical discourse is that the proof relies crucially on the results of a computer program written to test about 1500 cases of map configurations. It is not the first computer proof[2], but it is the first computer proof of a mainstream mathematical problem of general interest to mathematicians.

The proof of the Four Color Theorem is too long for any human being to survey or check in a lifetime. This fact makes some philosophers of mathematics uneasy. The Four Color Theorem fails to meet a major desideratum for proof: that it be checkable by a person. Does the existence of unsurveyable proofs force us to change what we mean by "proof"?

Thomas Tymoczko says yes—since the proof of the Four Color Theorem is not surveyable, it is not a proof in the traditional sense. Furthermore, accepting computer-aided proofs into mathematics introduces experiment in mathematics, showing that it is at least a partly empirical discipline.

By "surveyable", Tymoczko must mean "surveyable in practice" rather than "surveyable in principle"; otherwise, the Four Color Theorem would be surveyable. Distinguishing these two notions will be crucial to uncovering what Tymoczko finds objectionable in computer-aided proofs.

I agree with Tymoczko that introducing computers into mathematics may in-

[1]A more technical explanation of the theorem will be presented in section 2.
[2]We will look at a computer proof predating the Four Color Theorem in a later section.

troduce experimental methods into proofs; however, whether a particular use of a computer in a proof counts as an experiment will depend on *which* computer method is used. Computers have been used to execute procedures in probabilistic algorithms. Some of the results have the status of proofs.[3] However, in the case of the Four Color Theorem, a computer serves merely as a computational workhorse, performing millions of operations on a determined number of cases. A person could in principle perform the operations were it not for physical limitations on lifespan, etc.

I maintain that although it is not humanly surveyable, the proof of the Four Color Theorem is still a mathematical proof in the traditional sense. The proof of the Four Color Theorem has more in common with other traditional mathematical proofs than with results in experimental science.

I will proceed as follows: I will give an overview of the Four Color Theorem, including some facts about how it was produced and what strategy was used. I will introduce some philosophical issues that it raises. Then I will look at a series of arguments by Tymoczko designed to show that mathematics is quasi-empirical. I discuss some of the popular objections and offer responses to his arguments; however, I acknowledge that although the Four Color Theorem is a proof in the traditional sense, it is possible that our conception of mathematical proof will be expanded by introducing computers into mathematical practice.

## 3.2 History of the Four Color Theorem

The Four Color Theorem has been a subject of interest since 1852, when Francis Guthrie first wrote to his brother Frederic that it seemed that countries of every map could be colored with only four colors such that neighboring countries were

[3]Later in this paper I will discuss Michael Rabin's probabilistic algorithm for determining the primality of large numbers; I maintain that counting it as a "proof" certainly introduces experiment into mathematics.

colored differently.[4] By neighbors, we mean countries that share a border rather than countries that meet at a single point (like wedges of a pie); otherwise, the map would require as many colors as countries. Further constraining what counts as a country is the requirement that no country completely surround another.

In 1878, Arthur Cayley proposed the Four Color Theorem as a problem to the London Mathematical Society; Arthur Kempe soon afterwards published a paper claiming to have solved the problem. To understand his purported solution, we will need some terminology.

### 3.2.1 Kempe's Attempted Proof

A map is called *normal* if none of its regions encloses any other region, and no more than 3 regions meet at any point. Kempe tried to prove the Four Color Theorem by reductio ad absurdum. He assumed that there is at least one 5-colorable map (a map that requires 5 colors), and tried to derive a contradiction. Kempe assumed that if there is a five-colorable map, then there is a normal five-colorable map, and furthermore, a minimal one (one such that any map with fewer regions would be four-colorable). To prove the Four Color Theorem, it suffices to show that a minimal five-colorable map is impossible.[5]

Kempe correctly showed that in any normal map there is at least one region with five or fewer neighbors, which means that one of four configurations (as seen in Figure 3-1) must appear on any normal map:

To say that one of these configurations must occur means that the set of configurations is *unavoidable*. Kempe argued that if a minimal normal five-colorable map had a region with 5 or fewer neighbors, then there would also have to be a normal map with fewer regions that was also five-colorable. But this contradicts the original

[4]Steen, [Steen 1978]

[5] [Appel and Haken 1980]

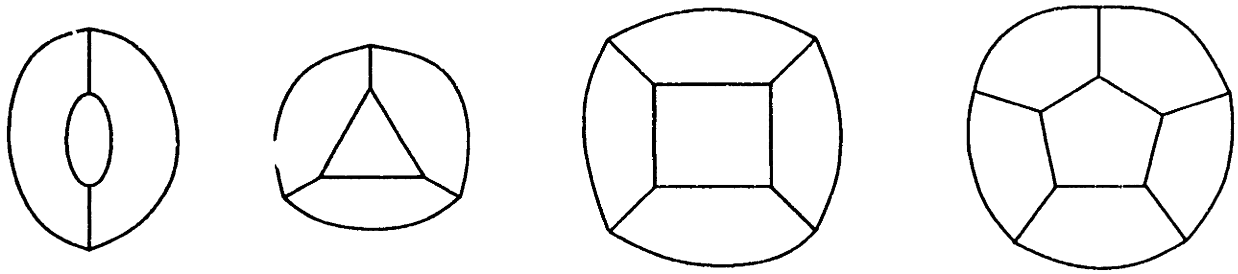

Figure 3-1: Kempe's unavoidable set of configurations

assumption that a minimal five-colorable map exists, thus completing the reductio ad absurdum argument.[6]

Kempe was able to derive the contradiction in the case of regions with 2, 3, or 4 neighbors; however, he was not able solve the problem for the 5-neighbor case.

P. J. Heawood pointed out the problem with Kempe's proof in 1890. He also studied the more general problem of how to color maps on surfaces other than a plane, e. g. a torus. Heawood was able to prove many theorems about the number of colors needed to color such surfaces, but he was never able to use his arguments in the case of planar surfaces. He was never able to prove the Four Color Theorem.

Kempe's argument did point out two important concepts needed to prove the theorem:

1. the idea of an unavoidable set of configurations

2. reducibility of the configurations in that set

What reducibility amounts to is the following: if there is a way of showing, by examining the configuration and the way chains of regions can be aligned, that the configuration cannot appear on a minimal 5-colorable map, then the configuration is reducible.

Since Kempe introduced this method, mathematicians have been working on ways

[6] A more detailed version of Kempe's argument can be found in Steen [Steen 1978].

to show that large numbers of configurations are reducible. But, a computer is required since the number of configurations is so large.

### 3.2.2 20th Century Developments on the Four Color Theorem

Work on the theorem continued throughout the 20th century. Heinrich Heesch developed a method called *discharging*, which was like moving charge in an electrical network, to find an unavoidable set of configurations. All vertices in a graph are assigned a "charge", determined by the number of neighbors at that vertex. All vertices of degree five are assigned positive charge, and vertices of degree greater than five are assigned a negative charge. The purpose of the discharging procedure was to develop a way to insure that all vertices of positive charge (those of degree five) belong to a reducible configuration. Then, since all triangulations (that is, all graphs under consideration) must have vertices of positive charge, the configurations in this set are unavoidable.[7]

In 1970, Wolfgang Haken started working on discharging with the goal of being able to show all configurations reducible. He and Kenneth Appel worked with many others to try to overcome two major problems:

1. reducing the number of configurations in the set, since computer power and memory requirements were enormous for a problem of this type.

2. reducing the ring size of the configurations.

A ring is a region bounded by circuit of vertices. The number of vertices determines the size of the ring. We see in Figure 3-2, a ring with 6 vertices, called a 6-ring. Over the next 6 years, many people contributed to the effort of reducing the complexity

[7] [Steen 1978], pp.169-170.

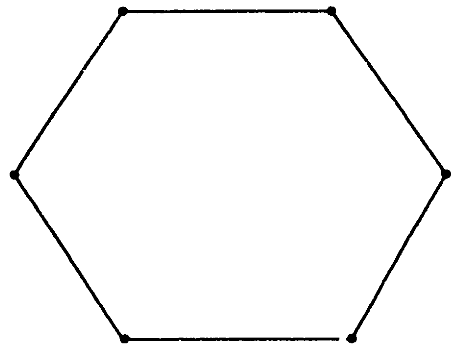

Figure 3-2: a sample 6-ring

of the problem. John Koch, then a graduate student, wrote programs to check for reducibility for configurations up to ring size 11.

In 1976, Appel and Haken arranged to have their program run on about 1500 cases. They emphasize that the computer's role in the proof was a crucial one:[8]

> A person could carefully check the discharging procedure that did not involve reducibility computations in a month or two, but it does not seem possible to check the reducibility computations themselves by hand. Indeed the referees of the paper resulting from our work used our complete notes to check the discharging procedure, but they resorted to an independent computer program to check the correctness of the reducibility computations.

It should be clear by now that the procedures for proving the Four Color Theorem involved extensive computational work, far more than had previously been attempted by any group of mathematicians. In the following section, we will see exactly how computationally demanding Appel and Haken's task was.

[8] [Steen 1978], p.178.

## 3.3 Computer Facts about the Four Color Theorem

Appel and Haken's program took more than 1200 hours of CPU time on an IBM 370-168 in 1976. The program tested the unavoidable set of 1478 configurations for reducibility. The analysis of each case involved generating all possible colorings of that configuration and checking, using the rules of the discharging procedure, for reducibility. Demands on run-time and memory increase by a factor of approximately 4 with the addition of one to the length of the ring. For example, one particular 13 ring had 66,430 colorings, whereas one 14 ring had 199,291 colorings. Testing the 14 ring for reducibility took about 10 minutes of CPU time.

As is clear by now, this was a task requiring enormous computational capacity. Computing speed and available memory have increased at quite a rapid rate since 1976. However, even running these cases on a Cray supercomputer now would, according to a computer scientist I asked, require order-of-magnitude 40 CPU hours. He based his estimate on the hypothesis that computing speed has roughly doubled every three years. Even taking into account expanding technology in the computer field, proving the Four Color Theorem is a laborious, not to mention expensive, enterprise, requiring the most that state-of-the art equipment can offer.

## 3.4 How the Four Color Theorem Challenges the Classical Conception of Proof

Some mathematicians do not like the Four Color Theorem because the proof is inelegant and non-algebraic (unlike the proof of the Five Color Theorem for planar graphs), but they accept it as a proof. What problems does its acceptance into mathematics present for philosophers who hold a traditional view of the role of proof in mathematical knowledge? According to this view (at various times put forth by Frege, Russell and many others), a proof in a system is a sequence of sentences in the language of the system such that each member of the sequence is either an axiom of the system or a sentence which results from previous members of the sequence in accordance with some rule of the system.

It is true that what we mean by 'proof' is 'proof in a standard formal system with a certain form...', But, that does not completely explain why we consider those particular sequences to be proofs. What makes them proofs is that they do a certain job—they convince us of the truth of the theorem proved, using clear, explicit, accepted reasoning.

Proofs serve a prescriptive, normative function. If I have followed a proof of the Pythagorean theorem, then I can conclude with impunity that whenever I do computations involving right triangles, if I add the squares of the lengths of the two shorter legs, the sum will equal the square of the hypotenuse. Following a proof of a theorem gives me good reasons to believe that it is true, and these reasons justify my belief in the theorem. In fact, following a proof *compels* my belief in the theorem.

Frege was disturbed that some mathematicians "confuse the grounds of proof with the mental or physical conditions to be satisfied if the proof is to be given".[9] He cites one of his favorite examples from the literature of his time: Schroeder's "Axiom of Symbolic Stability. It guarantees us that throughout all our arguments and deductions the symbols remain constant in our memory—or preferably on paper".[10] That psychology could affect the foundations of mathematics to the extent that we needed safeguards against mysteriously changing variable letters seemed absurd to Frege. What he thought affected the foundations of mathematics was the degree of rigor with which many results were formulated.

But it is not the degree of rigor that is troubling in this case; the Four Color

---

[9]Grundlagen, p.VIII
[10]Grundlagen, pp.VIII-IX.

Theorem was presented by Appel and Haken as a definitive, rigorous, complete proof. Their discharging procedure was proven in a mathematically rigorous fashion to produce an unavoidable set U of configurations (the actual number of configurations was around 1475; it has more recently been reduced to around 1000). Computers were used to develop the discharging procedure and the set U, but once it was produced it could be surveyed; one can give a surveyable proof that this set U is unavoidable.

However, the last step of the proof—showing that every configuration in U is reducible— *cannot* be surveyed in detail. Verifying that last step requires running a computer program on the configurations to test them for reducibility. An actual printout of this step would be practically impossible to obtain and certainly impossible to attend to in a reasonable length of time ( I will treat this issue in some detail later).

Does the existence of a computer-assisted proof force us to change our view of what it means for something to be a proof? In particular, does the introduction of computer-verified steps in a proof introduce empirical methods into mathematical practice? Given that the mathematical community accepts the proof, are we then forced to accept that mathematics is quasi-empirical after all? Before drawing any conclusions, we must see what the criterion of surveyability comes to, why it might be considered important, and whether it conflicts with our intuitions about what mathematical proofs actually *do*.

## 3.5 Thomas Tymoczko on the Four Color Theorem

Thomas Tymoczko [Tymoczko 1979], in his well-known 1979 article, asserts that acceptance of the Four Color Theorem does indeed force us to adopt what he calls a "quasi-empirical" account of mathematics. On his view, the existence of computer-assisted proofs introduces experimental methods into pure mathematics and the philosophical ramifications of such an introduction are quite serious:[11]

> If we accept the Four Color Theorem as a theorem, then we are committed to changing the sense of "theorem", or more to the point, to changing the underlying concept of "proof".

In service of his case, Tymoczko presents what he considers to be three major characteristics of mathematical proofs, and then questions the extent to which the Four Color Theorem fits his characterization. He concludes that while the computer proof of it *is* a real proof, it represents a departure from the traditional conception. In particular, it is a proof that is known *a posteriori.*

Tymoczko lists three characteristics that are true of proofs:

1. Proofs are *surveyable.*

2. Proofs are *convincing.*

3. Proofs are *formalizable.*

   A proof is *surveyable* if it is checkable, comprehensible in its entirety. It must be possible to be checked definitively by members of the mathematical community, although such a procedure could take months. According to Tymoczko, surveyability makes proofs accessible to any competent mathematician. It is the lack of surveyability that gives Tymoczko pause when deciding whether to accept the proof of the Four Color Theorem.

   Proofs are also *convincing*; this is a fact about the anthropology of mathematics. Surveyability and formalizability help explain *why* they are convincing.

   Proofs are *formalizable.* In practice, we do *not* formalize proofs, for it would make them too long (in many cases) to survey or even comprehend. All correct

---

[11] [Tymoczko 1979], p.58.

proofs do, however, have the property that they can be converted into proofs in some formal system or other. Having this structure helps explain why proofs are so convincing.

A natural question that comes to mind is whether these two requirements of surveyability and formalizability are at odds. Tymoczko briefly notes that most mathematicians consider surveyable proofs to be formalizable (Heyting and Lakatos are notable exceptions), but he focuses on whether formal proofs are surveyable. The answer is an easy no. Formalizing a proof drastically increases its length, so there must be formal proofs that are not surveyable. By surveyable, Tymoczko means "can be read over by a mathematician in a human lifetime".

In general, we come to know formal proofs either directly or their existence is established by means of informal surveyable arguments. Of course, few (if any) proofs are written formally; what usually happens is that a mathematician gives an informal surveyable argument that the formal proof exists. He notes that "there are general surveyable arguments that any proof in, say, elementary arithmetic can be formalized in Zermelo-Frankel set theory."[12] So in practice, the only way we come to know formal proofs is through the existence of surveyable proofs.

Tymoczko maintains that the proof of the Four Color Theorem drives a wedge between the criteria of surveyability and formalizability—the proof of the Four Color Theorem is formalizable but *not* surveyable.

However, Appel and Haken's work *does* convince us of the truth of the theorem; we are convinced by surveying a proof with a key lemma which is justified by citing the results of running a computer program; Tymoczko therefore concludes

[12] [Tymoczko 1979], p. 62.

that appeal to the lemma is justified on empirical grounds.

According to Tymoczko, this fact—that appeal to the lemma is justified on empirical grounds—is both surprising and important, for it has serious ramifications for the philosophy of mathematics. We accept that the proof of the Four Color Theorem is convincing, but being a convincing argument is not sufficient to establish it as a proof. Tymoczko says that the Four Color Theorem does not have a surveyable proof—no mathematician can survey the proof of the reducibility of the unavoidable set U. Another way to put it is that Appel and Haken's proof *is* surveyable, except that the key lemma is justified by appeal to computer, a process which is *not* surveyable. Either way, he would have us believe that this evidence forces us to change our conception of mathematical proof.

To illustrate what he thinks is going on when we use the phrase "appeal to computer", he offers what he considers an analogous case:

Imagine that on Mars there are mathematicians like there are here on Earth, except that on Mars there is a genius mathematician called Simon. He can prove lots of theorems that other people have proved, but he can also prove theorems that no one else has been able to prove. He justifies steps in his proofs with "proof too long, but I have verified it". Sometimes people are able to reconstruct his results, giving traditional proofs, but not always. However, since Simon is such a mathematical genius, people accept his results, incorporating his results into their own proofs, justifying them with the line "Simon Says".[13]

Tymoczko says that appeal to computers and appeal to Simon are similar. If we consider computers to be a legitimate authority but not Simon, then it is because we have some evidence for the reliability of computers. What

[13] [Tymoczko 1979], p.71.

kind of evidence we provide will be crucial to deciding the status of computer-assisted proofs. Tymoczko claims that whatever evidence that is, it *cannot* be in the form of a surveyable proof, for the proof of the reducibility lemma is not surveyable. Therefore, appealing to computers "introduces a new method into mathematics".[14]

Does the Four Color Theorem have a formal proof? Most mathematicians think so. But, the reasons for believing that a formal proof exists are because of the *current* proof, which involves appeal to computers. One might object that this appeal is only a harmless extension of human powers, that the computer just traces out the steps of the formal proof. Tymoczko says that our reason for believing that a formal proof exists is the surveyable proof containing the reducibility lemma that is justified by appeal to the results of a computer-run experiment. Our evidence presupposes the reliability of computers.

What factors do we consider when assessing the reliability of computers? Tymoczko mentions two:

(a) reliability of the machine

(b) reliability of the program

We have to rely on engineers and physicists to design machines that work; we rely on programmers to write good assemblers, compilers, languages and programs. In the case of the Four Color Theorem, many mathematicians believe that the appeal to computers is justified, that computers are a reliable means of generating correct information. However, Tymoczko claims that since that guarantees we get from the reliability of computers are not the same guarantees we get from traditional methods of proof, the Four Color Theorem is *not* known

[14] [Tymoczko 1979], p.72.

with the same degree of certainty. In fact, when put to the test, our faith in the reliability of computers can be shaken. He gives an example:[15]

> Suppose some supercomputer were to work on the consistency of Peano Arithmetic and it reported a proof of inconsistency, a proof which was so long and complex that no mathematician could understand beyond the most general terms. Could we have sufficient faith in computers to accept this result, or would we say that the empirical evidence for their reliability is not enough? Would such a result justify a mathematician's claim to know that PA was inconsistent, and would such a mathematician have to abandon PA? These are bizarre questions, but they suggest that the reliability of computer-assisted proofs in mathematics, while easy to accept in the case of the Four Color Theorem, might some day be harder to swallow.

Common philosophical wisdom distinguishes a priori truths from a posteriori truths in the following time-honored (if imprecise) way: a priori truths are known independent of experience; a posteriori truths are known only through experience. Tymoczko concedes that we indeed know many theorems a priori, but the Four Color Theorem is not one of them. We may know a priori that the proof with the reducibility lemma implies the Four Color Theorem, but our knowledge of the reducibility lemma does not take the form of a proof that we know a priori.

Our knowledge of the reducibility lemma is *a posteriori* knowledge, for it rests on empirical assumptions about a computer-assisted procedure.[16] He adds that it is unlikely that anyone will ever come to know the Four Color Theorem a

[15] [Tymoczko 1979], p.73.
[16] [Tymoczko 1979], p.72.

priori, since it is unlikely that anyone will ever come up with, say, an algebraic proof of it. Therefore, it is an a posteriori truth, proved via the first a posteriori mathematical proof.[17]

Crucial to Tymoczko's case is what he takes "surveyability" to mean. He seems to be saying that it means "surveyable in practice". We shall see that this explication may present problems for Tymoczko; it will at least obligate him to give a further account which may unduly restrict what we can count as proofs in the traditional sense.

## 3.6 Objections to Tymoczko's View

### 3.6.1 Teller's Comments on Surveyability

Following the publication of Tymoczko's article were a number of replies. Paul Teller [Teller 1980] considers surveyability to be important, not because proofs that are not surveyable are proofs in some different sense, "but because without surveyability we seem not to be able to verify that a proof is correct...it is a characteristic which some proofs have, and which we want our proofs to have so that we may reasonably assure ourselves that what we take to be a correct proof is so."[18]

As the field of mathematics progresses, we acquire new methods of surveying, which allow us to expand our ways of checking proofs. But, we needn't change our conception of proof to accommodate the shift in the methods of surveying. Some of the methods we have used include the use of pencil, paper, slide rules, calculators, and log tables. Using a computer to check the key reducibility

[17] [Tymoczko 1979], p.73.
[18] [Teller 1980], p.798.

lemma in the Four Color Theorem represents merely "an extension in our means of surveying, not a change in our concept of proof."[19]

The point is well-taken; relying on computers is not so different from relying on, say, log tables, presses that print log tables, or calculators. We certainly have to rely on at least the use of pencil and paper to help us record and remember lines in a proof when it gets too long for us to apprehend in its entirety. Tymoczko seems to allow that we can survey proofs using paper and pencil, but not proofs using computers; he owes us an explanation of the difference between the two cases, but none appears to be forthcoming.

Tymoczko's Simon example shows us that we might be skeptical about a computer-generated unsurveyable proof whose structure was too complex for us to understand. Teller admits that we would worry about such a proof, but the worry consists in whether the proof is correct, not whether the proof (if correct) is a proof in some new sense. If the proof is correct, then it is just as much of a proof as those that are humanly surveyable.

The Simon analogy is also used to show that although we consider the appeal to computers as legitimate (as opposed to the appeal to Simon), our evidence for their reliability is somewhat shaky. I maintain that we consider computers to be a legitimate authority because we *know* how computers work. We have no idea how Simon works, what laws under which he operates, how his computational processes work. At this point, the former can be formalized, whereas the latter cannot.

Tymoczko concedes this last point, and he rightly points out that the kind of evidence we provide for the reliability of computer-assisted proofs is important. To distinguish computer-assisted methods from non-computer-assisted ones, he

---

[19] [Teller 1980], p.799.

must make exactly the distinction mentioned above: he must give a more detailed account of surveyability in practice that distinguishes traditional cases like following a proof of, say, the Pythagorean theorem from both the Simon case and the computer cases.

Tymoczko brings up an important challenge for advocates of a classical view of mathematics: Does the use of computers introduce experiment into mathematical proof? Teller responds to the challenge in two ways. First, he denies that there is any principled difference between the performance of computers and the performance of mathematicians. Although Tymoczko says that whether we describe computer-assisted methods of proof as experiments or new methods of proof is "largely a matter of notational convention"[20], Teller sees no reason for us to describe them as the latter. The use of computers represents an expansion in our means of checking proofs, not a shift in the foundations of mathematics.

Teller's second response trades on a standard intuition about the difference between "mathematical" facts and "scientific" ones. Experiments establish spatio-temporal facts like "the meter reading was 4.5 on June 1, 1992, at 2PM." Correct mathematical proofs, on the other hand, establish non-spatio-temporal facts. He gives an example:[21]

> If one repeats a proof of a fact about numbers, unlike a measurement of the charges of an electron, one has to get the same result as before, again on the assumption that one does not use a mistaken method of proof and as long as one makes no mistake in applying that method of proof. And all this goes for computer-executed proofs as much as for proofs executed by human organisms.

[20] [Tymoczko 1979], p.76.
[21] [Teller 1980], p.799.

### 3.6.2 Experiment and Mathematical Proof

Other critics[22] disagree with Teller's view, and hold that use of calculation *does* in fact introduce experiment into mathematics. Michael Detlefsen and Mark Luker[23] agree with Tymoczko that the proof of the Four Color Theorem relies on empirical evidence, but this is not a novel event; many proofs involving calculation depend upon empirical evidence. They also argue that surveyability of a proof does not guard it against reliance on empirical factors.

Detlefsen and Luker offer reasons why the Four Color Theorem should not be treated as novel. First, there are many computer-assisted proofs which predate the Four Color Theorem[24]. Second, and more importantly, they argue that there is no real difference between a proof that involves calculation by a computer and one that involves calculation by a human. They use Tymoczko's paradigm case of a proof known a priori, the theorem of Gauss that the sum of the first 100 positive integers is 5050. The proof consists in writing down the numbers in two rows of fifty columns, as follows:

$$\begin{array}{ccccccc} 1 & 2 & 3 & 4 & \ldots & 49 & 50 \\ 100 & 99 & 98 & 97 & \ldots & 52 & 51 \end{array}$$

We notice that the sum of each column is 101 and that there are 50 columns. We can easily determine by quick calculation that the sum of all 100 integers is 5050.

[22] notably Michael Resnik [Resnik 1989].

[23] "The Four Color Theorem and Mathematical Proof", Journal of Philosophy 76, February 1979, pp.803–820.

[24] They cite the Lucas-Lehmer algorithm for finding Mersenne primes, and Cerutti and Davis' computerized proof of the theorem of Pappus, the latter of which we will examine later in this paper.

Tymoczko says that a proof "is a unit of reasoning that contains everything within itself needed for conviction".[25] However, Detlefsen and Luker take issue with his claim. They point out that for the above computations to take place, we must be certain of many things, among them that[26]

- the underlying algorithm to be used is mathematically sound.
- the program used is a correct implementation of this algorithm.
- the computing agent correctly executes the program.
- the reported result was actually obtained.

Tymoczko readily admits that we rely on factors like the above to establish the truth of the Four Color Theorem. Detlefsen and Luker maintain that if Tymoczko's analysis of proof as "needing nothing outside itself to carry conviction" is correct, then we also must rely on factors like the above to derive Gauss's conclusion from our simple observations. Therefore, they conclude that empirical considerations enter into the process of proving most theorems in mathematics.

One might object to their conclusion by trying to give a non-empirical account of computation. They suggest a promising candidate: "an episode of computation is taken as being something that is composed of elementary steps, each of which 'in and of itself produces complete conviction' and is crystal clear to the intellect. Thus, 'computation produces knowledge which is a priori"'.[27]

Detlefsen and Luker respond to the charge of the apriorist, claiming the above characterization of computation is misguided. The steps of a computation, they say, are not supposed to be intuitively obvious, rather simply mechanical steps

[25]Tymoczko. p.59.
[26]Detlefsen and Luker, p.810.
[27]Detlefsen and Luker, p.813.

that require no cleverness or insight. Their crucial point, though is this: these mechanical operations may (and often are) performed on (physical) symbols, *not* the things that the symbols represent. Calculation, then, is a physical activity on physically traceable objects, and therefore relies upon empirical factors for its success.

Michael Resnik echoes the sentiments of Detlefsen and Luker on this point. He notes that computation is used in mathematics to support many non-deductive arguments; for example, in number theory, one can test a general conjecture by computing some of its instances.[28] The case of the Four Color Theorem is also a case in which empirical factors were considered in the execution of the necessary calculations:[29]

> In the esoteric computation for the Four Color Proof, the odds are good that nobody involved knew all the mathematical, computer, electronic, chemical and physical theory required to give a complete account of why the computer's computation counts as reliable evidence for the mathematical facts. At least in this sociological respect, the Four Color Computation was very much like a scientific experiment.

Resnik anticipates the apriorist's objection to viewing calculation as bringing empirical methods into mathematics. We can in principle deduce mathematical statements from purely mathematical premises which contain no reference to physical events. It is these ideal reconstructions that play a normative role in defining standards of proof.[30] Resnik agrees (at least in general) with Tymoczko

[28]Resnik, p.132.
[29]Resnik, p.134.
[30]Resnik, p.140.

that these normative standards are inappropriate, since we never in practice give formal proofs.

The purpose of this small digression is to suggest that if Tymoczko wants to allow that some proofs are known a priori, then he would have to distinguish between cases in which empirical considerations are relevant and ones in which they are not.

Tymoczko concedes that we rely on a host of empirical factors when employing even traditional methods of proof and still maintain that computer-assisted methods change the face of mathematical proof. He seems to be distinguishing the cases by saying that the guarantees we get from the reliability of computers are not the same as the guarantees we get from traditional methods of proof.

Let us now consider a hypothetical case. Suppose it was possible to prove the Four Color Theorem using just a few axioms and modus ponens. What kind of guarantees does modus ponens offer us? Well, using modus ponens guarantees that if you have $A$ and $A \to B$, then you can conclude $B$. But you also have to rely on the use of pencil and paper, memory, mental acuity, etc. in order to be in a position to draw the conclusion.

What kinds of guarantees do computer-aided methods give us? Are they so different? If you write a computer program correctly and compile it using a functioning compiler, then, if the power does not go out and you do not run out of memory or run short on CPU time, then the program will return the correct result.

It is true that some empirical circumstances will interfere with theses processes, and likely *more* circumstances can interfere with computer-aided processes than traditional processes; however, the difference seems to be one of degree, not kind.

Detlefsen and Luker note that surveyability of a proof does not guard against empiricism in mathematics. Tymoczko uses the Simon example to compare the logic of "appeal to authority" and "appeal to computer". Although there are many differences in the two cases, neither is as reliable as appeal to first-hand survey, i.e. looking over the proof oneself. However, there are cases in which first-hand survey fails to guard against error.[31] Purported proofs can contain errors that go undetected for years. Some mathematicians go even further; they quote Philip Davis:[32]

> A derivation of a theorem or a verification of a proof has only probabilistic validity. It makes no difference whether the instrument of derivation or verification is man or a machine. The probabilities may vary, but are roughly of the same order of magnitude when compared with cosmic probabilities.

Detlefsen and Luker do not consider the Four Color Theorem to be an extreme case of a proof in a new sense, for they view it as merely another computer-assisted, deductive proof. What they find more intriguing is proofs that use non-deductive methods, such as Michael Rabin's probabilistic algorithm for determining the primality of large numbers.[33] They believe that probabilistic methods should be allowed in mathematical proof, but that those methods will drastically alter the nature of mathematical proof. We will discuss this point in more detail in section 3.7.1 and section 3.7.2.

---

[31] See section 7.2 for an example.
[32] Detlefsen and Luker, p.816.
[33] We will discuss Rabin's proof in section 3.7.1.

### 3.6.3 More on Surveyability—Has the Proof Really Been Surveyed?

Israel Krakowski[34] contends that the proof of the Four Color Theorem has actually been surveyed; the *computer* "has, in a step-by-step fashion, surveyed and proved this lemma [the reducibility lemma]." [35] Although Tymoczko objects that justification by appeal to computers is similar to the unsatisfying Simon case, Krakowski says that he is not concerned with the process of justification, but rather with the process of proving that the computer has completed.[36]

What is important to that process is that each step in the calculation *has* been taken. To suggest that the computer has not taken the steps assumes some kind of alternative view of what counts as calculation. Of course things go wrong with computer calculations, but these are at least as well-understood and predictable as (if not more so than) the problems that go wrong with human calculations.

Krakowski admits that the proof of the Four Color Theorem "highlights the already existing empirical elements of mathematical knowledge."[37] But he maintains that by bestowing surveying capabilities on a computer, the Four Color Theorem can be known a priori.

In the introduction to a collection of essays on empiricism in mathematics, Tymoczko responds to Krakowski, saying that he does not think admitting computers to the American Mathematical Society will solve the problem. We can avoid the problem by treating computers not as new colleagues, but as new tools for us to use. Then he is yet again faced with the problem of saying which so-called empirical methods provide us with traditional guarantees, and which

[34]'the Four Color Problem Reconsidered", Philosophical Studies 38, 1980, pp.91–96.
[35]Krakowski, p.92.
[36]Krakowski, p.93.
[37]Krakowski, p.95.

methods do not.

### 3.6.4 Another Classical Defense of A Priori Proof

Margarita Levin[38] objects to the notion that a computer proof of a mathematical theorem can be like an experiment. Like the other critics, she points out the obvious analogies between worries about the accuracy of computers and the accuracy of mathematicians. However, she warns that Tymoczko's argument may result in having to classify many theorems proved by traditional method as experiments, too. Since Tymoczko considers the Four Color Theorem a novelty, then he must not think that mathematical empiricism holds true because of the human epistemic condition. Otherwise, there would be no reason to distinguish the Four Color Theorem from any other mathematical theorem.

Levin is right; We do take into account a number of empirical factors in assessing the reliability of computer-aided methods, but we also use some of those same factors in assessing the reliability of traditional methods as well. Of course we have to rely on the laws of physics in order to believe the results of a computer-aided proof, but we have to rely on the laws of physics to believe most things, including that a mathematician's work is correct. In addition, we have to rely on factors less reliable than the laws of physics: the fact that mathematicians are conscientious and attentive, that we are not hallucinating, etc.

Levin examines a case in which the population of China is called upon collectively to do the calculations needed for some proof. Certainly such a proof would not be surveyable in practice. If Tymoczko considers this proof classically acceptable, then it is unclear why he would not accept the computer proof,

[38] "On Tymoczko's Argument for Mathematical Empiricism", Philosophical Studies 39, 1981, pp.79-86.

which would differ only in the number of calculations and the agents carrying out the calculations. If he does not consider the proof classically acceptable, then his point can be made without adverting to computer methods at all, but then he has the problem that many traditional proofs will now be considered experiments. In either case Tymoczko must explain what about the Four Color Theorem makes his case.[39]

If trust in computer hardware is the problem, then Tymoczko still fares no better, for as Levin says, we place our trust in log tables, printing presses as well. Does trust in those things bring empirical factors into our proofs which use them?

Tymoczko's arguments so far fail to convince the philosopher who is a realist about mathematical proof that lack of surveyability introduces a new element into mathematical proof. On the realist's view, there exist theorems whose proofs are too long for us to follow. Nonetheless, there *are* procedures to determine whether such formal proofs are correct. These theorems are also, on traditional accounts, considered to be known a priori. Tymoczko would have to deny the a priori status of such theorems if he maintains this strict view of surveyability. It is not clear that he wants to do so, for if he does, he may find himself between the rock and a hard place that Levin describes above.

### 3.6.5 A Computer Proof Predating the Four Color Theorem

Mathematicians were worrying about the possibility of new problems being ushered in by the advent of computers long before the Four Color Theorem was proved. In a 1969 article on a computer proof of the Theorem of Pappus, Elsie

[39]Levin, p.84.

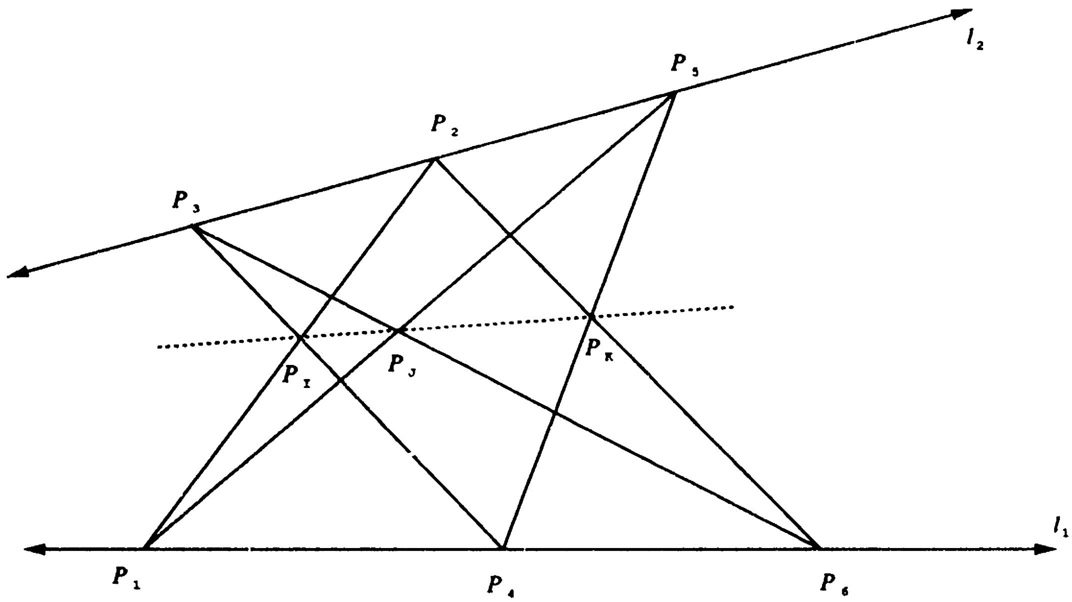


Figure 3-3: geometric illustration of the Theorem of Pappus

Cerutti and Phillip Davis[40] consider the question "What constitutes a proof in mathematics?" They bewail the difficulties one encounters when trying to prove theorems in analytic geometry by doing long algebraic computations; obviously these kinds of tedious computational tasks are well-suited to computers. Their paper describes how they developed a computer-executed computational proof of the theorem of Pappus (for which there exists an analytic proof).[41]

The Theorem of Pappus states the following:

> Let $l_1$, $l_2$, be straight lines in the plane. On $l_1$, take 3 points $P_1$, $P_4$, $P_6$ arbitrarily and on $l_2$ take the points $P_2$, $P_3$, $P_5$ arbitrarily. Now, connect the points in a criss-cross fashion indicated in Figure 3-3.
>
> Call the points of intersection $P_i$, $P_j$, and $P_k$. They will turn out to be collinear.

[40]"Formac Meets Pappus: Some Observations on Elementary Analytic Geometry by Computer", American Mathematical Monthly 76, 1976, pp.895-904.

[41]See citation 4 in Cerutti and Davis, p.905, for one location of the proof.

The information was represented in the program was as follows: The points were represented as coordinates. Solving the problem required basic but tedious algebra, involving solving for the determinant DE of (before reductions) 3,072 monomials. The output (after 4.52 minutes of execution time which included compiling and preprocessor time) was the line $DE = 0$[42].

If the computer printed out the line DE=0, that was sufficient to prove the Theorem of Pappus[43].

Cerutti and Davis also indicated that their computer-assisted methods would allow them to derive some new theorems or generalizations of old ones. They discuss examples of generalizations of the theorem of Pappus. One such theorem they describe as being "derived after an inspection of a machine printout... and this process can be described as *computer assisted theorem derivation*."[44] They acknowledge, however, the limitations of such methods: "Even with a computer at one's disposal, transformations and shorthand notations may therefore be sought to reduce storage requirements and to interpret the output."[45] At that time memory was limited, so problems had to be formulated around this constraint.

Cerutti and Davis suggest some possible objections to the proof:[46]

> What if the programming were erroneous? What if the initial data were false? What if there was a machine malfunction? What if the programmer, in a moment of pique, simply programmed the computer to type out $DE = 0$ and let it go at that?

---

[42]Cerutti and D vis, pp.898-9.

[43]The details of the proof are found in Cerutti and Davis, including how they reduced the number of monomials to be evaluated.

[44]Cerutti and Davis, pp.902.

[45]Cerutti and Davis, pp.903.

[46]Cerutti and Davis, p.903.

> These are certainly valid concerns. Similar objections, however, can be raised in the case of conventional proofs. One aspect of mathematical proof is that it consists of a finite string of symbols which must be recognized one by one and processed either by a person or a machine or by both. Now symbols must have physical traces on paper, in the brain, or elsewhere and cannot be reproduced and recognized with perfect fidelity. Human processing is subject to such things as fatigue, limited knowledge or memory, and to the psychological desire to force a particular result to come out.

Cerutti and Davis point out that we do have ways to overcome the obstacles involved in doing computer proofs: we can run the program over and over to check for errors, we can check the steps in the program ourselves, and we can ask colleagues to inspect the program and try running similar programs. All of these activities increase the credibility of the computer-assisted proof. In the case of traditionally proved theorems, we also go through processes of checking and rechecking, but despite our best efforts many faulty theorems remain. Detlefsen and Luker cite group theorist Daniel Gorenstein on the problems with solutions to the classification problem for finite groups:

> ...it seems beyond human capacity to present a closely-reasoned, several-hundred page argument with absolute accuracy. I am not speaking of the inevitable typographical errors... but of "local" arguments that are not quite right—a misstatement, a gap... there is a prevalent feeling that, with so many individuals working... every significant configuration will loom into view sufficiently often and so cannot remain unnoticed for long. On the other hand, *it clearly indicated the strong need for continual reexamination of the existing "proofs"*(my emphasis).

Cerutti and Davis conclude that a mathematical proof "has much in common with a physical experiment; that its validity is not absolute, but rests upon the same foundation of repeated experimentation."[47]

The fact that there are problems in the practice of mathematics is not sufficient to preclude our being justified in having a priori knowledge of a theorem once we have followed it. Tymoczko uses the example of imagining a computer proof of the inconsistency of PA to try to show how our faith in the reliability of computers could be shaken. Certainly it is unclear exactly how we would take such news, but it is likely that we would regard this putative result with great skepticism. However, we would be extremely skeptical of a purported traditional proof as well. A case like this does *not* show that the reliability of computer proofs is hard to swallow; rather, it shows that there are some statements in mathematics that we are loathe to give up it would take something drastic to convince the mathematical community that PA is inconsistent, so *any* methods used to arrive a such a conclusion would be strictly scrutinized.

## 3.7 What is the Epistemological Status of Computer Proofs in General?

### 3.7.1 Probabilistic Methods in Computer Proofs

How can we know that attempting an computer experiment is the best way to go about solving a mathematical problem? Tymoczko points out "even where questions of the form P(n) are decidable and we have the techniques to program a computer to check the instances, we cannot simply run the computer as long

[47]ibid. , p.904.

as it will go, hoping that it finds, say, that $\exists x P(x)$ before the computer reaches its limits. There must be some reason to expect that the computer will stop with an answer within a reasonable time."[48] In the case of the Four Color Theorem we can ask why anyone thought that an unavoidable set of reducible configurations each of ring size less than or equal to 14 could be found. From the outside, 14 looks no more probable as a bound than 20 or 50 or even 100. But, if the minimum ring size were 20 or more, the experiment would not have been feasible.

Mathematician Edward Moore proved that the unavoidable set must include configurations whose ring size is at least 12. Perhaps Moore would discover a map requiring the minimum ring size to be 20. Why did Appel and Haken think their experiment would work?

Tymoczko answers: "they used a sophisticated probabilistic argument, not a proof, that the ring size could be restricted to 17 or less, and that restriction to 14 was a good bet. They provided an argument that invested statements of the form 'there is an unavoidable set of reducible configurations each of which has a ring size less than or equal to n' with a probability derived from the ratio of the number of vertices in the configuration to the ring size n." [49]

Their strategy is not uncommon in mathematics. One of the most famous cases of a probabilistic "proof" is Michael Rabin's probabilistic algorithm for determining if a given number is prime. He contends that it may be possible to "prove" many statements using computers if we allow the computer to err with a predetermined low probability.[50] The summary below is based on Kolata' explication of Rabin's proof.

[48]Tymoczko, p.79.

[49]Haken, p.202. A more detailed explanation of their argument can be found in Appendix A.

[50]G.B. Kolata, "Mathematical Proofs: The Genesis of Reasonable Doubt", Science 1976, pp.989–990.

**Details of Rabin's Proof**

Rabin's test for primality was based on a discovery that if a number $n$ is prime, then every integer between 1 and n will pass a certain test (called "being a witness" for n). If any integer fails the test, then n is not prime. Rabin discovered that if $n$ is not prime, then at least half the integers between 1 and $n$ will fail Miller's test. If some number between 1 and $n$ is chosen randomly, then there is a 50% chance it will fail the test. In general, the probability that $k$ numbers chosen between 1 and $n$ will fail the test is $1 - \frac{1}{2^k}$. So we can test enough numbers until the probability of n's being composite, i.e. not prime, is acceptably low.

Exact testing of potential primes larger than, say, $10^{60}$ takes a very long time, and may outstrip the computational capacities of our current computers, so the probabilistic method, in addition to being faster and more efficient, is also a practical solution to the problem of testing large numbers for primality.

### 3.7.2 Does the Use of Probabilistic Methods Alter What Counts as a Proof?

Whereas the use of computer-assisted steps in a mathematical proof does not necessarily force classical philosophers of mathematics to change their conception of proof, introduction of probabilistic methods certainly does. Probabilistic proofs demonstrate the truth of a theorem only within certain degrees of error.

Advocates of the legitimacy of probabilistic proofs argue that (very long) classical proofs can be considered only probably correct,[51] as they are subject to errors. Probabilistic proofs may be technically easier to understand, much shorter,

[51]DeMillo, Lipton, and Perlis, "Social Processes and Proofs of Theorems and Programs", Communications of the ACM 22, v.5, May 1979, pp.271–280.

more perspicuous, and may allow us to isolate important mathematical notions useful for further research. In the case at hand, some mathematicians maintain that they have *more* confidence in results that could be obtained by probabilistic methods than in many 400-page mathematical proofs[52]. Classical proofs may be so long that no one will be able to comprehend more than the barest outline of its reasoning, and therefore less able to find errors.

Detlefsen and Luker hold that we should accept Rabin's methods as valid mathematical proof techniques because of two things: 1) the high degree of certainty conferred by his algorithm; and 2) the fact that many proofs using classical deductive methods fall prey to many kinds of uncertainty.

A case of this kind, involving so-called classical proofs[53] actually happened. Two groups of topologists, one American, the other Japanese, independently announced results concerning a topological object called a homotopy group. Their results contradicted each other; since both proofs involved complex symbolic and numeric computations, it was unclear who was wrong. The groups exchanged proofs, looking for errors; however, they found none, even though each group was keen on doing so. A third groups enters the scene with another proof, this one in support of the American result. The Japanese tactically withdrew to reconsider their proof.

While anecdotal, this story does present a challenge for the proponent of the classical view. Of course it is true that many factors influence our ability to assess proofs. But, the fact that verifying the correctness of a traditionally proven theorem may be incredibly difficult just shows that some proofs are of sufficient complexity to be (for the time being) beyond our cognitive reach. It remains that we *do* have methods to determine if a given proof in a formal

[52]Kolata, p. 990.
[53]DeMillo, Lipton, and Perlis, p.272.

system is correct. We have correctness conditions for such proofs; they may be impossible to implement in the case of a particular proof because of constraints on attention and comprehension, but that does not affect their status as proofs, merely our abilities to verify that status. Furthermore, worries about such proofs are worries about their correctness, not necessarily worries about that status of the methods used to prove the theorems.

In the case of computer-assisted proofs, if a program is provably correct, then computer's role in the proof is a trivial one; it is similar to that role played by pencil and paper in a traditional proof.

However, if the program is *not* provably correct, then the fact that it runs on such-and-such a machine *is* part of the evidence that the program works. If we have worked out correctness conditions for computer programs, then, depending on what those conditions are, the role played by the computer will be as trivial as that played by pencil and paper. Correctness conditions have *not*, as of yet, been worked out. However, at first blush, it would appear that the computer-verified procedure of the reducibility lemma would meet any reasonable set of correctness conditions, for it was used for purely computational purposes.

## 3.8 Closing Comments

Tymoczko has presented us with an account of surveyability from which it follows that whether an argument counts as a proof (in the traditional sense) hinges on biological/psychological facts about humans; proofs that can be surveyed in say, less than 60 years count as proofs, but those not surveyable in less than 60 years do not count as proofs. He also says that surveyability in practice applies to many traditionally proved theorems but not to theorems proved by computers. What is required but not supplied by him is an explanation of how to distinguish these cases.

There is a further problem with using the notion surveyability to mean "surveyable in practice". Can we distinguish between what cannot be surveyed in practice for merely adventitious reasons (e.g. we cannot pay people enough to survey some proofs) and what cannot be surveyed for intellectual or cognitive reasons (e.g. humans do not live long enough or have enough cognitive stamina to survey really long proofs)? Presumably Tymoczko would want the notion "surveyable in practice" to be immune from limitations on character, but not biological limitations. However, it is not clear how to make this distinction so that the cases are divided the way he wants.

The classical account of proof allows for the separation of the questions of whether something *is* a proof and whether we can *recognize* that something is a proof. What separates mathematical proof from experiment is that there are standards of rigor for formal proofs; given a sequence of statements in a formal system, we have a procedure for determining whether it is a proof of some theorem in that system.

It is obvious that using computers to solve formerly practically unsolvable problems will change our mathematical practice. Tymoczko rightly points out that certain crucial questions will have to be answered, like the the following: since not *everything* that claims to be a computer proof can be accepted as valid, what are the criteria for acceptable computer proofs? It seems likely that standards will be developed and methods refined as we use computers to do more powerful and complex computational work.

One big issue that has already changed the face of mathematics is the use of computers in probabilistic algorithms. Michael Rabin's probabilistic algorithm for determining the primality of large numbers is an example of a a procedure that mathematicians accept as a proof, but the result is only an answer with a certain (albeit high) degree of probability. Using computers in probabilistic arguments may force a revision of the notion of mathematical proof or force creation of a new notion in addition to the old one. But what will determine whether use of a certain computer method is not the fact that it *is* one, but rather what*kind* of method it is. The use of a computer in the proof of the Four Color Theorem does not seem to represent an introduction of a new method of proof, but rather an improvement of old methods.

As for the status of the Four Color Theorem, it is hard to accept that this seemingly essential fact about the nature of planar graphs is an empirical fact. The Five Color Theorem for planar graphs has an algebraic proof, which is both surveyable and formalizable; it counts as a traditional proof according to everyone, including Tymoczko. Assuming that both the Four Color Theorem and the Five Color Theorem are true, it is odd to attribute a priori status to the latter and a posteriori status to the former.

There is, however, something unsatisfying about the proof of the Four Color Theorem; the discharging procedures do offer some information about the configurations in the unavoidable set U, but they do not provide a perspicuous explanation of what property or properties of planar graphs give rise to four-colorability.

On the other hand, after the results of the Four Color Theorem were published, improvements were made in the discharging procedure to reduce the number of configurations in the unavoidable set to less than 1000.

Maybe what we can learn from the work done on the Four Color Theorem is how to apply our methods of proof to take advantage of the computational resources of digital computers. It is an open question what count as acceptable methods for use in computer-assisted proofs, but one worthy of attention both by mathematicians and philosophers.

# Bibliography

[Appel and Haken 1980] Kenneth Appel and Wolfgang Haken. The solution of the four-color-map problem. *Scientific American*, pages 108–120, September 1980.

[Asch 1956] Solomon Asch. Studies of independence and conformity of a minority against a unanimous majority. *Psychological Monographs: General and Applied*, 70(9), 1956.

[Ayer 1946] A.J. Ayer. *Language, Truth and Logic*. London, 1946.

[Cerutti and Davis 1976] Elsie Cerutti and Phillip Davis. Formac meets Pappus: Some observations on elementary analytic geometry by computer. *American Mathematical Monthly*, 76:895–904, 1976.

[Church 1936] Alonzo Church. A note on the entscheidungsproblem. *Journal of Symbolic Logic*, 1:40–41, 1936.

[Church 1937] Alonzo Church. Reviews of turing 1936 and post 1936. *Journal of Symbolic Logic*, 2:42–43, 1937.

[Church 1938] Alonzo Church. An unsolvable problem of elementary number theory. *American Journal of Mathe-

*matics*, 58:345–363, 1938.

[Clay and Lehrer 1989] Marjorie Clay and Keith Lehrer, editors. *Knowledge and Skepticism.* Westview Press, 1989.

[Currie and Worrall 1978] Gregory Currie and John Worrall, editors. *Mathematics, Science and Epistemology: Philosophical Papers Volume 2- Imre Lakatos.* Cambridge University Press, 1978.

[Davis and Hersh 1981] Philip J. Davis and Reuben Hersh. *The Mathematical Experience.* Houghton Mifflin Company, 1981.

[Detlefsen and Luker 1980] Michael Detlefsen and Mark Luker. The four color theorem and mathematical proof. *Journal of Philosophy*, 77:803–820, 1980.

[Enderton 1977] Herbert Enderton. *Handbook of Mathematical Logic*, pages 547–712. Studies in logic and the foundations of mathematics. North-Holland, 1977.

[Gandy 1989] Robin Gandy. The confluence of ideas in 1936. In Rudolph Herkel, editor, *The Universal Turing Machine*, pages 55–111. Oxford University Press, 1989.

[Kalmar 1956] Laszlo Kalmar. An argument against the plausibility of Church's thesis. In Arendt Heyting, editor, *Constructivity in Mathematics*, pages 72–80, 1956.

[Kant 1965] Immanuel Kant. *The Critique of Pure Reason.* Macmillan, 1965. translated by Norman Kempe Smith.

[Kitcher 1984] Philip Kitcher. *The Nature of Mathematical Knowledge.* Oxford University Press, 1984.

[Kornblith 1985] Hilary Kornblith, editor. *Naturalizing Epistemology.* MIT Press, 1985.

[Krakowski 1980] Israel Krakowski. The four color problem reconsidered. *Philosophical Studies*, 38:91–96, 1980.

[Lakatos 1976] Imre Lakatos. *Proofs and Refutations: The Logic of Mathematical Discovery.* Cambridge University Press, 1976.

[Levin 1981] Margarita Levin. On Tymoczko's argument for mathematical empiricism. *Philosophical Studies*, 39:79–86, 1981.

[Maclane 19xx] Saunders Maclane. The nature of mathematical proof. *American Mathematical Monthly*, 10(1):465–480, 19xx.

[Mendelson 1990] Elliott Mendelson. Second thoughts about church's thesis and mathematical proofs. *Journal of Philosophy*, 1990.

[Post 1936] Emil Post. Finite combinatory processes— formulation 1. *The Journal of Symbolic Logic*, 1936.

[Putnam 1979] Hilary Putnam. Analyticity and aprioricity: Beyond Wittgenstein and Quine. *Midwest Studies*, pages 97–98, 1979.

[Resnik 1989] Michael Resnik. Computation and mathematical empiricism. *Philosophical Topics*, 17(2):129–144, 1989.

[Shoenfield 1967] Joseph Shoenfield. *Mathematical Logic.* Addison-Wesley, 1967.