# PROOF AND PREJUDICE:
# WHY FORMALISING DOESN'T MAKE YOU A FORMALIST

**MSc Thesis** *(Afstudeerscriptie)*

written by

**Fenner Tanswell**
(born 17th of June, 1988 in Camden, London, United Kingdom)

under the supervision of **Benedikt Löwe**, and submitted to the Board of Examiners in partial fulfillment of the requirements for the degree of

**MSc in Logic**

at the *Universiteit van Amsterdam.*

**Date of the public defense:**
*26th of September, 2012*

**Members of the Thesis Committee:**
Jesse Alama
Alexandru Baltag
Catarina Dutilh Novaes
Benedikt Löwe
Martin Stokhof

INSTITUTE FOR LOGIC, LANGUAGE AND COMPUTATION

**Abstract**

The topic of this thesis is the relationship between formal and informal proofs. Chapter One opens the discussion by examining what a proof is, when two proofs are identical, what the purpose of proving is and how to distinguish the two categories of proof. Chapters Two and Three focus in on informal and formal proof respectively, with the latter also including descriptions of various computational proof checkers and the Formalist family of positions in the Philosophy of Mathematics. In Chapter Four I look at the Formalisability Thesis, that every informal proof corresponds to a formal proof, and argue that this breaks apart into a weak and a strong reading. In Chapter Five, I outline a simple mathematical problem and attempt the practical process of formalisation, following which I consider the decisions that were involved in doing so. Finally, in Chapter Six, I use what was learned from the practicalities of formalisation to argue in favour of the weak reading of the Formalisability Thesis, which I take to be closely related to Carnapian explication, but against the strong reading, which corresponds to the Formalists' approach to formalisation.

## Thanks

Firstly, I would like to thank Benedikt for going well beyond expected supervisory duties: long discussions, detailed feedback, and regular meetings (despite being in different countries!) were invaluable to successful completion of this thesis.

I also thank the Naproche team for a welcoming week in Cambridge learning about Naproche and Jesse Alama for guidance on using Mizar.

Finally, all family, friends and MoL gangers who have offered support, chatted philosophy or made cups of tea: I am incredibly grateful to you all...

... (especially for the tea.)

# Contents

# Chapter 1

# Introduction to Proof and Proving

I begin with an outline of the structure of this thesis and some introductory reflections on the differences between proof in mathematics and proof in logic, what proofs are and what their purpose is.

## 1.1 Outline

The starting point for this thesis is the 1999 paper by Yehuda Rav [55], which brought about a resurgence in interest in the differences between two distinct uses of the term 'proof'. On the one hand, there are *informal proofs*. These are proofs as mathematicians write them; they are presented in textbooks and journals, on blackboards and napkins. They can appeal to meanings, understanding and intuitions and are used to ascertain and transfer mathematical knowledge. On the other hand, we have *formal proofs*.[1] These are proofs as seen in logic: they are syntactic, explicitly defined and strictly formal. The first calling point for this thesis will be to consider each of these concepts in turn, beginning with informal proofs in chapter 2 and then proceeding to formal proofs in chapter 3. In sections 3.4 and 3.5 of the latter I will also introduce several versions of the Formalist position (drawing on Shapiro's outline from [60, pp. 140-171]) and describe how they relate to formal proof.

Having looked at informal and formal proofs individually, I will proceed to the question of what the relationship between them is. A candidate answer that Rav considers is the *Formalisability Thesis*, that all informal proofs can be formalised to be formal proofs (in some appropriate formal system). Different ways of understanding what this means and

[1] It should be noted that Rav uses different terminology to that being used here. In [55] he distinguishes between 'proof' and 'derivation', but I prefer 'informal proof' and 'formal proof' because it doesn't come burdened with the implicature that formal derivations are not proper proofs. This fits better with, for example, the conclusions of the empirical study into researchers' intuitions about proofs in [48], which indicates that knowledge of formal proofs is sufficient but not necessary for mathematical knowledge, suggesting that a formal proof *is* a particular type of proof.

evaluating its philosophical content will be the topic of chapter 4, where I will describe a weak and a strong reading of the Formalisability Thesis. The former is seen as an empirical claim, while the latter is the reading of the Formalist position in the philosophy of mathematics.

In chapter 5, I will work through the laborious practical process of formalising one simple informal proof, that of the mutilated chess board. I will outline the problem, present an interpretation of its mathematical content and lastly produce a formalisation.

Finally, what is learned from working through the mutilated chess board example will motivate the philosophical conclusions of chapter 6. Firstly, I will relate the formalisation process to Carnap's explication and from there argue in favour of the weak reading of the Formalisability Thesis. Secondly, I will reject the stronger, Formalist reading of the Formalisability Thesis. Finally I conclude that one can agree that in principle any proof can be formalised without admitting that rigorous proof inherently needs to be formal.

## 1.2 Why Do We Prove Things?

The question of what a proof is would be the obvious starting point, but it is in my opinion somewhat of an empty question without first thinking about what proofs do and why we prove things. So, let me consider some of the main functions of proof in mathematics.

The first and foremost function of proofs is to *obtain mathematical knowledge of mathematical truths*. Until we have a proof, a mathematical statement will remain conjectural. Of course, there are now other methods of giving evidence for mathematical claims; for example, probabilistic primality testing can give overwhelming inductive evidence for the primality of some given number. Despite their reliability, though, such tests are one of the targets of the PYTHIAGORA oracular computer thought-experiment that Rav uses at the beginning of [55]. In it, the computer can instantly confirm or deny any mathematical statement fed into it, answering with absolute certainty. The result of the thought-experiment is that mathematics is not simply about sorting true statements from false, but also about coming to learn *why* statements are true or false and how they interrelate. Although primality tests are one way of convincing us to believe that a particular mathematical statement is almost certainly true, with many useful applications, it is with a mathematical proof that we learn with certainty how and why it is true. Mathematical proof reveals deductively how a mathematical conclusion follows from previously known mathematical statements.

The moral of PYTHIAGORA tells us that proof does not only need to establish mathematical truth, but also show us how and why something is true. In [34], Hersh argues for the slogan that proof is *convincing* and *explaining*: a proof produced by a working mathematician is primarily aimed at convincing the qualified judges of the mathematical community, but this is not all. Proofs are also presented by teachers to their students yet, Hersh observes, students are too easily convinced. So in this context the proof is not intended to merely convince, but also to explain to the students why it is true. Maybe a better way of putting this is to say that proofs, both for the student and the mathematical community, are used to convince through mathematical explanation, rather than trying to convince by authority (since being convinced by authority is essentially analogous to deferring to a mathematical oracle such as PYTHIAGORA).[2]

The role thusfar ascribed to proof leaves it as a rather tame creature: used and useful, but still essentially subservient to truth. Another proposal by Rav, though, is to reverse this prejudice:

> *...proofs rather than the statement-form of theorems are the bearers of mathematical knowledge.* Theorems are in a sense just tags, labels for proofs, summaries of information, headlines of news, editorial devices. The whole arsenal of mathematical methodologies, concepts, strategies and techniques for solving problems, the establishment of interconnections between theories, the systematisation of results—the entire mathematical know-how is embedded in proofs. [55, p. 20]

The purpose of proof, from this point of view, is no longer aimed at obtaining mathematical knowledge of mathematical truths, but instead shifts the focus to obtaining mathematical knowledge itself, since mathematical knowledge proper consists of knowing and understanding proofs. It could then be said that the purpose of proving is to actively take part in and engage with mathematics.

Another reason for proving and proofs is the end product: that which they reveal. Above, it sounds as if each proof is a self-standing unit that gets us to a single mathematical truth, but in general (and merely having considered the alternative proof-first picture) we know that the reality is much better. For what a proof really reveals is not merely the truth of a mathematical statement, but the *dependencies* that exist between it and other pieces and techniques of mathematics. Doing proofs reveals the rich tapestry of structure and interrelations in mathematics, which is in a large part what the discipline is all about. By producing a proof we can learn which other mathematical facts are needed to support a theorem and how they depend on one another. As is observed by Dawson in [20], we don't just prove theorems but also *re-prove* them, which adds to the web of mathematical knowledge by showing that a proof can go through in different ways using different methods, possibly with different dependencies.

Not only do proofs reveal the relationships that obtain between different mathematical objects or theorems, but also the process of proving is useful for *clarifying and disambiguating* our mathematical concepts. Usually proofs will invoke properties of their mathematical subjects, which means that we become clearer on what the terms of mathematics refer to through their mathematical properties. Of course, this doesn't mean that proving leads to

[2] There is the worry, though, that the idea of proof as convincing and explaining is still idealised when compared to actual mathematical practice wherein some work is too advanced to convince or explain to anyone but the top handful of academics in the field.

a definite answer as to what exactly is meant by a mathematical term, but it is in proofs that we use and manipulate these terms and so it is through this usage that properties are revealed and any problems or difficulties may eventually be found.

Proof and provability are also central features in the study of metamathematics. Proofs are integral to investigating formal systems, in matters of consistency, independence, conservativity and much more.

This may not be an exhaustive list of reasons for proof in mathematics, but it covers the main motivations. Furthermore, it covers a reasonably broad selection, by which I mean that the listed reasons are sufficiently diverse to reveal that different presentations or styles of proofs will be better suited to achieving different ends. For example, a diagrammatic proof may be best for explaining why something is true, but disambiguating concepts is more likely to be achieved when we try to invoke fewer intuitions in a more formal proof while metamathematics is focused on formal proofs and provability.

I will return to the specific purpose of informal proofs in section 2.1, but for now I turn to how exactly such different styles of proof may be distinguished into the two distinct categories of formal and informal.

## 1.3 Two Notions of Proof

Before getting into discussing formal and informal proofs separately in the subsequent two chapters, it is prudent to have a tentative division of the two concepts.

A *formal proof* is a derivation in a formal, axiomatic system. A formal proof therefore is a syntactic, mechanical, gapless and explicit step-by-step piece of reasoning to get from the axioms of the system to the conclusion; a procedure which is therefore, at least in theory, checkable by an automaton or computer. Formal systems will be defined properly in section 3.1, but for now they are calculi which link premises to conclusions, through the application of explicit rules of inference. Such a system should be over some fixed formal language within which the premises are explicitly formulated and which the rules operate on. For example, Hilbert-style derivations, as seen in [27, p. 87]; Gentzen's natural deduction and sequent calculus, as found in [64]; and the various tableaux favoured by Priest in [54] are all examples of formal systems found standardly in proof theory. However, for mathematics such systems are cumbersome and impractical, so in the field of Formal Mathematics popular formal systems are computational and tend to be designed to better simulate the structure of informal mathematical proofs. Examples of such formal systems are found in Mizar, Coq, Naproche, HOL Light and Isabelle.[3] Axioms serve as the ultimate

[3]The difference between the formal systems of proof theory and those of formal mathematics may merely be a reflection of the balancing act between a system's practical use in proving theorems and the ease with which one can prove things about the system. In formal mathematics, the use of computers allows the system to be quite applicable but with less interest in proving theorems about it, whereas in proof theory the emphasis is on showing things about the systems while actual derivations can get long, tedious and overly complicated.

premises in the formal systems and example hereof include the Zermelo-Fraenkel axioms with choice (ZFC) for set theory or the axioms of Peano Arithmetic (PA), to choose but the most obvious two. Each formal system itself has a formal definition[4], hence formal proofs are mathematical objects in their own right, studied in proof theory and manipulated in meta-mathematics.

By *informal proofs*, on the other hand, I will refer to proofs as they are usually written in practice, such as in journals, textbooks, correspondence or classrooms. Broadly speaking, they are presented in natural language augmented with mathematical symbols, terminology and diagrams. Their aim is to convince some intended rational audience of the truth of a mathematical statement through a reasoned argument or series of arguments which appeal to the intellect of the audience. They can draw on concepts which have not been explicitly defined and may (in some cases) even engage our intuitions about space, inductions and generalisations. These proofs will be tailored to the audience they are intended for, so that a proof in an academic journal will be different to a proof intended for first-year undergraduates. So long as the method of the proof is clear and convincing to the intended audience it may be that easier or routine steps are left unstated.

That I have presented these two types of proof as separate does not necessarily mean that they are. For one thing, the standard view seems to be that formal proofs are a subclass of informal proofs. In the other direction, a Formalist may argue that all informal proofs correspond to, or are only true in virtue of, some formal proof. So how should this classification go?

## 1.4 What is a Proof and When are Two Proofs the Same?

Having given a general answer to the question of the purpose of proofs and proving in section 1.2, it should now be possible to approach the second question of what a proof is exactly. The question, though, is not quite clear enough, for there is a distinction to be had between proofs as instantiated and proofs *per se*. Consider the following statement by Leitgeb:

> Whereas published mathematical proofs—as of 2009—might be identified with certain traces of chalk on blackboards, printed texts in mathematics journals, or electronic patterns in some technical devices, we regard mathematical proofs *per se* as abstract entities which are independent of any material instantiation. [43, p. 266]

The dichotomy here is between the practical and the theoretical: in practice, proofs are presented in many different ways and have certain practical limitations while in theory they can "transcend all boundaries of a merely pragmatic kind" [43, p. 267]. What this

[4]However, the issues of this thesis are very much present here, for the definition of formal proofs in the computational systems might be far beyond the practical limitations for giving such a definition.

amounts to is that proofs as we create and use them are concrete instantiations of certain abstract 'proof' entities. Formal proofs are then a subset of these which, in addition to being abstract entities, are also mathematical entities.

Of course, this distinction between abstract proofs and instantiations thereof raises further questions: what are abstract 'proof' entities? what sorts of properties do they have? how does a concrete proof instantiate an abstract proof? what can we know about the abstract proof (independent of its instantiations)? The intention of this thesis is not to delve into the murky depths of these metaphysical questions. However, one question that is entangled with these other is worth a moment's thought, and that is the question of proof identity.

On the picture of proof that divides proofs into abstract proofs and concrete instantiations, the natural approach to proof identity is to say that two concrete proofs are identical if they instantiate the same abstract proof. But then, one may ask, how similar do two proofs have to be to instantiate the same abstract proof? The question, I believe, is one of granularity; but let me separate the answers for informal and formal proofs.

Firstly, let's consider informal proofs. A very fine-grained approach could require that the exact wording of two informal proofs be the same for the two informal proofs to be identical, but this seems too restrictive for there are many words that appear in an informal proof that aren't of vital importance to the intended understanding of the proof. Furthermore, it is often acceptable to change the order in which different parts of the proof are done without turning it into a different proof. At the other end of the spectrum, a coarse-grained approach may only require the general proof strategy to coincide for two informal proofs to be the same, yet this may be too liberal because in many respects mathematics must pay attention to details so despite having the same strategy the execution of that strategy might be non-trivially different. My leaning for informal proofs is more towards the coarse-grained approach, as this strikes me as how we really do see proofs: that there is a natural notion of "being the same proof" on which two proofs are the same unless there is some substantial difference between them.

Identity of formal proofs will play an important role in chapter 4, so consider whether the same story that held for informal proofs also holds for these. Let me set aside for now identity between formal proofs in different systems (which will certainly complicate things) by fixing a formal system $\Lambda$. As it was for informal proofs, we expect there to be a natural notion of "being the same proof", but this time relativised to $\Lambda$. Such a notion will induce an equivalence relation of $\Lambda$-*sameness* on formal proofs in $\Lambda$. How to spell out this relation, however, is not entirely clear.

Let us at least give upper and lower bounds for the granularity of the equivalence relation. The strictest, finest-grained approach would take $\Lambda$-sameness to be identity of sequences of $\Lambda$-formulas, but as in the case of informal proofs this is too strict: on this definition two formal proofs may be different despite only having systematically replaced variables to get from one to the other. A bound for liberal, coarse-granularity is to identify all $\Lambda$-proofs of the same statement, but this won't be correct either as we naturally recognise that the same statement can be given different proofs.

To try to refine this further, the immediate thought is that formal proofs, by virtue of their status as mathematical objects, can be identified based on certain syntactic features. An obvious first step is to deal with the example of systematic variable-renaming by having $\Lambda$-sameness identify $\Lambda$-proofs at least up to such renaming (explicitly defined, based on $\Lambda$). Proof theorists have a number of other syntactic features that allow us to identify equivalence classes of formal proofs. For example, in [21] Došen discusses equivalences like normalisation and generality, but these equivalence classes do not properly match the notion of "being the same proof". On the one hand, many proofs we consider the same are still distinguished as not the same, but on the other hand, identifying a formal proof with its normal form is in many cases beyond what we would want to call the same proof.

I do not here have a positive proposal to make for how to further refine or specify what the classes of $\Lambda$-same proofs will look like. However, it can be observed that this will be no easy task, based on work on a similar question on identifying the sameness of stories in narratology, as is investigated in [45] and [46]. The latter paper raises a number of methodological issues for fixing equivalence classes of *structurally the same* stories, issues that can be just as troubling in trying to identify sameness of structure in proofs. Fortunately, making these equivalence classes precise is not all too important for the coming discussion, so long as $\Lambda$-sameness is coarse enough to identify proofs despite the minor differences discussed, but fine enough to separate proofs that we recognise as different. For the remainder of this thesis, I will proceed assuming that each formal system $\Lambda$ has a corresponding equivalence relation of $\Lambda$-sameness.

In this thesis I will be mainly concerned with practical aspects of formalisation and, as such, much of the focus will be on concrete instantiations of proofs rather than abstract proof and provability. It is, however, important to keep in mind the distinction between abstract and concrete instantiations of proofs, the granularity of proof identity and whether this concrete-abstract picture of proofs is the correct one,[5] especially when it comes to thinking about Formalism in sections 3.4 and 3.5, and the philosophical discussion of chapter 6.

[5]The distinction as it has been presented lends itself naturally to a Platonistic approach to the philosophy of mathematics, which is in strong disagreement with several main theses of the general Formalist position.

# Chapter 2

# Informal Proofs

In this chapter I take a closer look at what an informal proof is usually understood to be, considering various ways in which the particular conception I start with could be objectionable. Then I consider a paradox raised in [9] which intends to show that informal proof and mathematics is inconsistent.

## 2.1 Informal Proof, Elaborated

The preliminary description in section 1.3 has left the notion of an informal proof rather threadbare in comparison to the easier definition of a formal proof. So, now, let me proceed with a more detailed examination of what counts as an informal proof.

The standard conception of the purpose of a proof for a theorem in mathematics, I would say, is roughly captured by the following:

**(P)** To give a rigorous, indefeasible sequence of mathematical arguments that convince your intended audience of the certainty of the theorem.

As a starting point this captures many of the purposes, mentioned previously in section 1.2, that apply more naturally to proofs as they are seen in mathematics generally (rather than, say, proof theory and mathematical logic). Nonetheless, I would like to run through and dissect the terms used in (P), discussing the various problems that might arise.

Firstly, *rigour* in mathematics (or, more accurately, in the philosophy thereof) has become closely linked to formality. The emergence of this link historically coincides with the extensive programme of rigorisation that took hold in the $19^{\text{th}}$ and early $20^{\text{th}}$ Centuries, as described by Tarski:

> Until the last years of the 19th century the notion of proof was primarily of a psychological character... No restrictions were put on arguments used in proofs, except that they had to be intuitively convincing. At a certain period, however, a need began to be felt for submitting the notion of proof to a deeper

> analysis that would result in restricting the recourse to intuitive evidence in this context as well. This was probably related to some specific developments in mathematics, in particular to the discovery of non-Euclidean geometries. The analysis was carried out by logicians, beginning with... Frege; it led to the introduction of a new notion, that of a *formal proof*, which turned out to be an adequate substitute and an essential improvement over the old psychological notion. [62, p. 70]

As well as the non-Euclidean geometries pointed to in the above quote, problems in analysis of using implicit or intuitive concepts of continuity, differentiability, convergence etc. were being replaced with explicit $\epsilon$-$\delta$ definitions to try and face up to growing discontent with previous methodology. As Feferman puts it:

> ...the nineteenth century program for the rigorous foundation of analysis and its arithmetization, ... at the hands most notably of Bolzano, Cauchy, Weierstrass, Dedekind and Cantor. That program grew in response to the increasing uncertainty as to what it was legitimate to do and say in mathematics, and especially in analysis. One could no longer rely on calculations that *looked* right, or depend on physical applications to justify the mathematics. [22, p. 321]

The idea, then, is to combat the inadequacies of informality (with its intuitions) and the problems of rigour that accompany this by formalising our proofs. In section 6.4, I shall reject that formality is necessary for rigorous mathematical proof, so this concern can be put aside for the moment.

The next item in (P), that proofs are made up of a sequence of *indefeasible* arguments also has its detractors. Lakatos, in [40], presents a dialogue in which a group of fictional students (named after Greek letters) discuss Euler's formula for polyhedra.[1] The view of proof offered through this dialogue is heuristic, being a component of Lakatos's "method of proofs and refutations"[40, pp. 49-50]. Proofs are given for a primitive conjecture, but the proofs on this picture amount to little more than convincing thought-experiments, which are defeasible in various ways: the various 'lemmas' that make up the thought-experiment may turn out to have counter-examples or the conjecture itself may have a global counter-example. The response to these can be to update and refine the concepts we are using, incorporate conditions into the statement of the theorem or to dismiss the proposed proof altogether. In his Introduction to Lakatos, Larvor describes how Lakatos takes the contexts of mathematical discovery and justification to be far more closely connected than standard accounts of mathematics acknowledge:

> Proofs, for Lakatos, do not just establish the truth of theorems. They can also be engines of discovery... he presents a picture of mathematics in which proofs, tests and thought-experiments are so intimately related that it is impossible to say where one context finishes and another starts. [42, p. 20]

[1] Euler's formula is $V - E + F = 2$ where $V$, $E$ and $F$ stand for vertices, edges and faces respectively.

This methodology for mathematics, which Lakatos is proposing as at least one way that mathematics is practiced, is opposed to the Ravian proof-first perspective (described earlier in section 1.2) in that it sees mathematics as going backwards and forwards between proof attempts and theorem statements. The proofs do not just aim to establish the theorems but also give feedback as to how to improve and refine the theorem. The doctrine advocated by Lakatos is also fallibilistic: that any accepted theory might turn out to be false. A defence of this position needs to explain how the many obvious mathematical truths can turn out to be wrong, which Lakatos does through discussion of conceptual change: that any concept can change in the development of mathematics. However, I don't think that either of the Lakatosian reasons warrant dropping the indefeasibility requirement for informal proofs. A longer piece of work might have the space to engage with these ideas in greater depth, but I will simply say that I think the methodology, while interesting, gives a different usage of the term 'proof' than we really do use it: if a theorem has counter-examples then it cannot be proved.

It is also key to ask in what sense indefeasibility is taken. Does it mean just that this particular audience can find no counter-examples? that the journal referees couldn't? that the wider mathematical community hasn't? Or do we have some kind of ideal in mind which means the proof is absolutely free of counter-examples? In section 6.4 I will give an answer which supports the ideal of absolute indefeasibility.

That a proof is even in *sequential* form is also too strict a requirement. As well as the Lakatosian methodological concerns, it is clear that diagrams play an important role in convincing and explaining in proofs. But, where in the proof sequence would an explanatory diagram be positioned? Often diagrams are parallel to the main proof, without being obviously redundant or 'not part of the proof', thus breaking the idea that a proof should be sequential. In fact, this can be taken further by looking at the many example of proofs without words, as collected in [50] and [51]. These books are filled with examples of theorems proved diagrammatically, with the occassional formula, wherein the sequential order of the diagrams only rarely matters.

Another problem with this sequential claim for proofs is that it may depend on the aim of the proof how one should go about presenting it. For although proofs in academic mathematics may at least appear sequential, in our vague characterisation of informal proof above it was also possible to include proofs in classroom. Such proofs may be presented in other ways to maximise student learning. For instance, in [1] the authors compare the efficacy of different proof techniques for different purposes. Their results indicate that explaining a proof strategy first and then filling in the technical details generally worked better for explaining proofs to students, again showing that sequential proofs are not necessarily preferred.[2] I believe the points of this paragraph and the last make it clear that arguments in an informal proof certainly do not need to be sequential, thus we can drop it

[2] It is important to note that the results were from a controlled study, so do not necessarily indicate that this is a universal truth of teaching mathematics. I merely raise it as a potential way that we could gain from not doing proofs sequentially.

from (P).

In the statement of (P) I included the specification that one uses *mathematical arguments* to rule out other types of convincing such as, for example, appeals to authority or probabilistic evidence. However, what exactly is to count as mathematical is once again up for debate. In fact, there is the looming danger of circularity here if the definition of mathematical were to refer to the methodology of proving. Furthermore, mathematics has traditionally been conceived of as a priori, but recent uses of computation to perform long calculations (such as in the case of the famous Four Colour Theorem) may cast doubt on this, as is argued in [35, pp. 158-159]. The general field of mathematics including computation or inductive evidence (such as the primality tests mentioned in section 1.2) is 'Experimental Mathematics': in [8], Baker argues that this field is characterised by the calculation of instances of some general hypothesis, such that it subsumes computer checking, using inductive evidence and statistical tests. Baker then argues that the problems raised by these methods might concern philosophers, but that in practice the concerns are different:

> Mathematicians have certainly expressed concerns over computer-based proofs, but for the most part these concerns do not have to do with the kind of knowledge they produce. Instead they are connected to worries over reliability on the one hand and explanatoriness on the other. [8, p. 341]

Nonetheless, worries about computer accuracy or the lack of explanation that comes with such proofs, Baker claims, do not detract from their practical uses. Mathematical practice is likewise the focus of van Kerkhove & van Bendegem's discussion of Experimental Mathematics in [38], saying that the methods it employs are a fact of mathematical practice and not necessarily incompatible with the use of formal proofs. They make use of an analogy by Hersh from [33] about mathematics having a frontstage of traditional, indefeasible proofs and a backstage of discovery through number crunching, computation and experimentation. In (P) the emphasis is implicitly on the front part of mathematics, so in order to maintain (P) we may want to restrict the requirement on arguments being mathematical to those being of this front stage: explanatory and deductive.

Depending on the answer to the question of how to guarantee informal rigour, there is also the question of how *certainty* is to be ascertained and verified. Since (P) works on the basis of intended audiences and informal mathematics is, at least on the surface, how mathematics is actually practiced, it becomes vital to know how mathematicians judge a theorem to be certainly true. Establishing what informal proof is then requires an investigation into the social and human process of being convinced by proofs. For example, in [44] where Löwe & Müller defend contextualism about mathematical knowledge, it is shown using examples from mathematical practice that what is required from a proof varies with context:

> Nor is proof a fixed notion: There are various forms of proof, and context determines which type of proof, if proof at all, is required. [44, p. 17]

There are also numerous examples of accepted proofs which have later been revealed as flawed— what does this say about certainty of proofs? Could it be that the certainty with which we hold theorems and proofs comes in degrees, depending on how many people understand and have checked the proof? Does certainty depend on the field of mathematics, or the mathematical culture within which the proof is presented?

Finally, in (P) there is a particular emphasis on the theorem that the proof is a proof of. But is this necessarily the correct perspective? In [55] Rav argues that, to the contrary, the main focus of mathematics is the discovery of novel proof-techniques and that proofs are where the informative content of mathematics resides, with theorems just as titles and headings. Such a perspective values even incorrect proofs if they still contain useful techniques and insights. The switch to the Ravian stance on proofs would entail the reformulation of the standard view into a more method-focused view of proofs.

## 2.2 Is Informal Proof Inconsistent?

In [9], while arguing in favour of paraconsistent mathematics, Beall mentions 'Gödel's paradox' which he claims demonstrates that informal mathematics is inconsistent. He asks us to consider the following sentence denoted by $\gamma$:[3]

($\gamma$) This sentence is informally unprovable.

This can be given a proof by *reductio*, for supposing that $\gamma$ is false means that it must be informally provable and thus true. By *reductio* we therefore have that the sentence $\gamma$ is true. This, though, was an informal proof of $\gamma$, which means that it is informally provable, so the sentence itself is false.

Beall argues from this:

> There seems to be little hope of denying that $\gamma$ is indeed a sentence of our informal mathematics. Accordingly, the only way to avoid the above result is to revert to formalising away the inconsistency—a response familiar from the history of naïve set theory, naïve semantic theory, and so on. [9, p. 324]

I disagree. I shall argue that there is good reason to reject $\gamma$ as a sentence of our informal mathematics.

Firstly, I wonder what Beall is drawing on when he says that there is little hope of denying that $\gamma$ is a proper sentence of informal mathematics. For looking at it doesn't reveal very much that is incontestably mathematical. Sure, Beall's statement of the problem (found in fn. 3) has a Greek letter as a label and even some self-reference, but this doesn't get very far, since we can similarly give another sentence and denote it by $\sigma$:

[3] I here stray slightly from Beall's description of the paradox, which stated it as "$\gamma$: $\gamma$ is informally unprovable." [9, p. 324].

($\sigma$) This sentence has a Greek letter as a label and contains self-reference!

This sentence is clearly not a mathematical statement. So what feature does qualify $\gamma$ as mathematical which $\sigma$ fails to include? It appears the only strictly mathematical part of $\gamma$ is the mention of provability. I contend, though, that this does not make $\gamma$ a mathematical statement. Of course I hold that informal proof and provability are very important notions in talking about mathematics, but the crucial thing to observe is that these are notions *about* mathematics rather than *of* mathematics. What I mean by this is that informal proof is certainly used to talk about mathematics and what is done in practicing mathematics, but it is not a concept that is within mathematics itself. For one thing, I am not aware of *any* other proofs that feature informal provability as a mathematical concept. As logicians we are more than familiar with plenty examples of provability featured in other proofs, but all such examples that spring to mind feature formal proof and provability, for which the paradox does not work. Beyond merely citing the circumstantial evidence of not knowing any proofs about informal provability, there is the fact that informal proof does not have well-established mathematical properties nor does it interrelate with other mathematical concepts in the way that standard mathematical concepts (such as group, integer, derivative, line etc.) do. The only clear interaction it has is with mathematical truth, as exploited in the paradox, but if anything the notion of truth in mathematics (before being formalised into some formal theory of truth) belongs to the same category of notions about mathematics rather than within. I say this for precisely the same reasons: informal truth is not something that is an essential ingredient of any informal proof, nor does it have well-established mathematical interactions or properties.

By denying that informal provability is a concept within informal mathematics, we consequently also deny that $\gamma$ is a sentence of our informal mathematics, thus deny that Beall has showed that mathematics is inconsistent.[4] However, dealing with the paradox in this way does serve to bring to the foreground the fact that formal proofs do have an important role to play, for these formal proofs are actually mathematical objects themselves so do fall into the domain of mathematics. Proof formalisation can therefore serve as an analytical tool, converting an informal proof that is not a mathematical object into a formal proof that is, which in turn allows mathematical investigation. This description is, as yet, rather simplistic; a fuller discussion of formalisability will begin in chapter 4. First, though, I shall use chapter 3 to conduct a proper investigation into what is meant by *formal proof*.

[4]Note, though, that the paradox is not hereby resolved: it still works fine as a paradox for natural language.

# Chapter 3

# Formal Proof and Formalism

In this chapter I will first state explicitly and mathematically what formal proofs are, then move on to discuss the Mizar and Naproche computational formal systems for mathematics. The second half of this chapter will give a brief history of Formalism, followed by a more contemporary account which has been offered as a response to Rav. Finally, an argument against Formalism based on the practical limitations on formal proofs will be rejected based on the success of developments in Formal Mathematics.

## 3.1 Formal Proofs

As already observed in section 1.3, there is quite a selection of formal systems within which it is possible to present formal proofs. Depending on which formal system is being used and even which variant of each system, in proof theory there are standardly two types of presentation for formal proofs, although this is complicated somewhat by formal systems for mathematics such as Mizar and Naproche which are more computational. I shall define each of the former in turn then in the following two sections look at the two automated proof-checkers and their associated formal systems.

**Definition.** An *alphabet* is a collection of symbols which may be strung together to form formulae.

Standardly, the alphabet for a formal language will include logical connectives, quantifiers, variables, constants, relations, brackets, reserved words[1] etc.

**Definition.** A *(well-formed) formula* is a string of symbols from the language, constructed by adhering to strict *formation rules* which allow symbols of the language to be combined in specific ways (rejecting all other strings of symbols as ill-formed).

[1]Reserved words are generally found in the computational systems but not in the well-known proof-theoretic systems.

**Definition.** A *formal language* is the set of well-formed formulae over the alphabet, as determined by the formation rules.

**Definition.** A *formal system*, over some formal language, is made up of axioms, axiom schemata and inference rules. The axioms and schemata are well-formed formulae (or sets, sequences, multisets etc. of formulae), while the inference rules explicitly determine how to proceed from certain formulae to others.

Depending on the inference rules of a formal system, proofs given in the standard proof-theoretic systems will either be sequential or tree-like.

**Definition.** A *sequential formal proof* in some formal system $S$ is a sequence of well-formed formulae in the language of $S$, where each formula is either an axiom or follows from previous formulae by the inference rules and the last formula in the sequence is the conclusion.

**Definition.** A *tree-like formal proof* in some formal system $S$ is a tree whose nodes are well-formed formulae in the language of $S$, where each formula is either an axiom or follows from its immediate predecessors by the inference rules and where the last formula is the conclusion.

But this does not capture the shape of formal proofs as they are done in the computational formal systems, such as Mizar and Naproche. These try to more closely emulate informal proofs, in which different proof elements are introduced as separate theorems, lemmas, axioms and definitions. Later items will be able to refer back to earlier ones without re-proving them. In the subsequent sections 3.2 and 3.3 I will describe these two systems and how they work. First, though, let me mention three of the other systems that will not get such a full investigation, but are referred to elsewhere in this thesis.

Firstly, *Coq* is an interactive proof assistant which allows proofs to be written in the Gallina specification language. Coq is named for and based on what is called the Calculus of Inductive Types, a higher-order type theory which includes higher-order functions and predicates as well as a hierarchy of sets. Interestingly, Coq has recently been used to check a fully formalised version of the Four Colour Theorem, as described in [28]. The full reference manual can be found at [19], though for a brief introduction a better start is [11].

The other two systems *Isabelle* and *HOL Light* can be considered together, stemming from the same source: the HOL family of interactive theorem provers, with HOL standing for higher-order logic, as they are both assisstants for classical higher-order logic. HOL Light was originally designed to, and still does, have a simpler logical core than the original HOL system, which has been maintained despite growing into a well-established variant of HOL. According to [67] where Wiedijk list 100 theorems along with which systems have checked them, as a gage for the success of Formal Mathematics, HOL Light is currently the most successful having formalised 85 of the theorems (Mizar is listed as second with 57, followed by Coq and Isabelle at 49 and 47 respectively). A good starting point for looking into HOL Light is the in-depth tutorial by Harrison [29]. Isabelle, rather than being a variant of HOL is better seen as its successor. Isabelle is less simple than HOL Light, but its uses do extend beyond Formal Mathematics into projects in logic and program development and specification languages.[2] An authoritative introduction and guide to Isabelle can be found in [52].

## 3.2 A Look at Mizar

In this section I will discuss one prominent system in the field of Formal Mathematics: Mizar, which is making advanced proof formalisation ever more achievable and has already succeeded in formalising a wide array of famous proofs. In giving the description of the system, I mainly draw on information in [26] and [49].

The Mizar project consists of three main components: the Mizar language, the Mizar computer program (for checking proofs in the Mizar language) and the Mizar Mathematical Library (hereafter MML, containing definitions and proved theorems in Mizar). The main aim of the project is to provide a practical and convenient language and associated proof-checking software with which mathematicians can formalise and verify mathematics. The language is designed such that it is sufficiently flexible to formulate statements in almost any mathematical field.[3] Formal mathematicians can write articles in the language containing definitions and theorems which have previously been proved informally, and now can be checked by Mizar. Finally, the definitions and results are added to the MML, which other articles can then specify are being used in their own articles, allowing the library to develop a catalogue of interconnected proofs, theorems and definitions. Now I shall look at each component in slightly more detail.

The Mizar language[4] was designed to have the following key properties: firstly, the wide applicability mentioned above; secondly, legibility to the point that the proofs are still possible to read and understand for members of the mathematical community; and lastly, that proofs written in it are formal and mechanically checkable.[5] Mizar contains all the usual first-order logical connectives and quantifiers, plus the ability to use free second-order variables to form schemata. The language then reserves a large number of mathematical terms to be used in a specified sense relating to their original meaning, then combining these with a number of grammatical features of Mizar means proofs can be written for the theorems in each article. These proofs can proceed in standard mathematical ways

[2] As can be seen on the Projects page for Isabelle at ⟨https://isabelle.in.tum.de/community/Projects⟩.

[3] Only avoiding the universal claim here because mathematics is so broad that it is seems likely that there may be some counter-examples.

[4] Technically, the Mizar language has had several different versions, but for our purposes we can focus on the modern incarnation which is used in the system primarily referred to as Mizar.

[5] The applicability and checkability are both achieved in Mizar, but reading Mizar proofs is certainly an acquired skill. The Naproche system described in the next section aims to do better at this.

(by induction, contradiction, reasoning by cases etc.) under the standard macrostructural elements (dividing proofs into axioms, definitions, theorems, lemmas etc.). Proving in Mizar is aided by the ability to type the variables and objects being considered (e.g. as reals, integers, sets etc.) and furthermore it is possible to define new types to work with. Structurally, Mizar proofs are required to follow the standard style of forward reasoning. The proofs in this language differ from informal proof, however, because they need to be fully formalised into the language of Mizar and explicit, in that every proof-step needs to be written, which means that Mizar proofs are generally longer than informal proofs.

Once one has written an article in the Mizar language, the text is sent to the Mizar checker, which mechanically checks the given proofs. The technical side of the computational process involves putting the proof through six different modules sequentially. First, the *scanner* makes an initial pass of the text, primarily to divide the text into its component parts or tokens, picking out the various symbols, numerals, reserved words etc. Secondly, the *parser* module takes the stream of tokens from the scanner and checks for syntactic correctness with respect to the grammatical rules of the Mizar language. Furthermore the parser generates the representation of the article that can be computationally handled by later modules. Next, the *analyser* works on disambiguating notation and constructions by using type information from the text.[6] The *reasoner* then verifies that the strategy of the proof will actually prove the formula it intends to prove. Then the *checker* goes through the proof verifying each individual inference. This is done by checking that the conjunction of the premises and the negation of the conclusion together classically imply a contradiction. This requires more mathematical content than first-order predicate logic provides, so Mizar proofs assume the axioms of Tarski-Grothendieck Set Theory. However, the checker skips the formula schemata which use free second-order variables; these are checked separately by the *schematizer* module at the end. If the proof fails in any of the modules, an error code is generated and the user can rework their proof, otherwise the proof is accepted.

The third part of the Mizar system is the MML. The MML by now contains a huge amount of formalised mathematics; at the time of writing it contains 51762 theorems, 10158 definitions and 787 schemata in 1150 articles.[7] More and more theorems and definitions are constantly being added to this collection. Allowing articles to reference and utilise results from other articles in the MML means that formalised mathematics in the Mizar system can emulate the cumulative aggregation of mathematics that is found standardly in regular mathematical practice. This, in turn, makes it possible to approach and formalise mathematics well beyond the basic foundations that are usually tested by such systems. The MML contains a large number of reasonably advanced theorems including the non-denumerability of the continuum, Ramsey's theorem, Sylow's theorem and key results from

[6]For example, the analyser may be required to choose between different operations with the same symbol, which is done using the types of the objects that the operation is applied to.

[7]These figures were taken from the MML Query homepage ⟨http://mmlquery.mizar.org/⟩, maintained by Grzegorz Bancerek of Bialystok Technical University in Poland, on 28/04/2012.

diverse areas of mathematics. Proving such theorems clearly indicates the level of progress being made by formalised mathematics.

## 3.3 A Look at Naproche

In the previous section I mentioned that reading formal proofs in Mizar, while very much possible, does require familiarity with the system and it's phrasing. There is an automatic translation system which presents Mizar proofs as readable mathematics ready to be published in the Journal of Formalized Mathematics, however this only makes the Journal reader-friendly, not the Mizar language itself. In contrast, the Naproche system is designed to be much easier for the everyday mathematician to get to grips with. Naproche is a front-end system for Coq, already described in section 3.1, and takes (LaTeX) textual inputs of mathematical texts, converts them into first-order logic and then feeds them into Coq for verification. "Naproche" itself stands for *Natural Language Proof Checking* and puts greater emphasis on incorporating the linguistic study of informal mathematics into Formal Mathematics. The main introduction to Naproche is [39], which I will draw on in this section's explanation of the system.

The central idea of Naproche is to develop a controlled natural language which can be straightforwardly used to read and write mathematical proofs but which simultaneously has a canonical representation in first-order logic. A controlled natural language is a language which maintains many features of natural language (in the case of Naproche: standard mathematical vernacular) while sufficiently limiting its notoriously unruly character by restricting the number of words and sentence structures that are allowed to be used, thereby allowing the text to be read and interpreted by a computer. To do so the Naproche system takes advantage of the highly structured nature of mathematical proofs, even in their informal form. As we shall be seeing in section 3.4 on Formalism, this structure and the heavy use of pieces of formal mathematical terms and interrelations is one of the main motivations for the Formalist belief that maths is actually a formal enterprise. The Naproche project is about maintaining the ease of using natural language combined with mathematical terms, while controlling this language such that it is sufficiently simple that an automated system can pinpoint the canonical and intended meaning.

The method for achieving this is that a mathematical text given in Naproche has a unique corresponding Proof Representation Structure (PRS) which are modified versions of Discourse Representation Structures for natural language semantics as in [36]. These are used to parse and disambiguate the natural language components of the text and their relation to the mathematical terms. The PRSs keep track of three different types of referents in the text and a list of conditions on those referents. The *discourse referents* are used to pick out the objects that are in the domain of discourse, i.e. those that can be referred to. In the case of informal mathematics, this can be either mathematical objects like numbers, sets, groups etc. or the symbols and formulae of the language.

Then *mathematical referents* are specifically the terms and formulae of the text themselves (rather than the objects referred to, as is the case for discourse referents), but these can be bound to discourse referents by a listed condition. The textual referents are used to pick out other proofs or lemmas that are being used to support the inference step. A full proof is then built out of a large number of these PRSs, which can be related in various ways and even nested within each other as conditions. More on the technical details of PRSs can be found in general overview in [39] or in depth in [37], but the important point is that each Naproche text has a unique PRS and this, in turn, has a unique translation into first-order logic. The computational system for Naproche thus takes the outputted canonical first-order form of the PRS for the proof and verifies it with Coq, reporting back the results to the formaliser.

As in Mizar, Naproche has standard mathematical macrostructure; mathematical texts are divided into axioms, definitions, theorems and lemmas. Many standard mathematical proof-techniques are implemented, such as proof by contradiction, reasoning by cases and all of the valid proofs of natural deduction. On top of this it allows various types of definitions, including recursive definitions of functions. Naproche implements a simple to use system of assumption introduction and discharging via its natural language phrasings, as well as useful and intuitive variable-binding methods.

A comment on the current status of Naproche is needed to warn that it is still a work in progress. Its development has been guided by a motivating example of wanting to formalise a foundational construction of elementary mathematics, that of Landau [41], meaning that features necessary for that task have taken priority. Two missing pieces that I would've liked for the example I will give in section 5.3 are set-theoretic notation and functions. Although Naproche does have the membership relation, to properly formalise the set-theoretic version of the problem that I will attempt, more set theory is needed (especially because the obvious way to bypass the lack of curly-brackets notation by reducing everything to first-order logic with $\in$ is blocked by limitations on ways in which definitions can currently be constructed in Naproche.) Secondly, function, though present in naproche, are restricted to elementary arithmetical functions which is understandable when motivated by Landau's text, but is prohibitive when it comes to wider applications.

Furthermore, Naproche as it stands avoids worries of types and typing by not mixing different types in different proofs. A long-term solution will be necessary if Naproche is to proceed. On top of this, in order to do anything beyond the very basics of elementary mathematics, it needs some way of invoking previously established results is needed, as in Mizar's MML.

I will return to Naproche in section 5.3, where I will work through an example of formalisation in the Naproche system. For now, though, I will move on to the philosophical position that draws most heavily on formal proof and formalisation: Formalism.

## 3.4 A Brief History of Formalism

The basic observation behind Formalism is that a great deal of mathematics is about the use of formal calculations, manipulations and rules, which might lead one to wonder whether this is all there is to mathematics and mathematical proof. A lot of what is taught at an elementary level is about how to flip the puzzle pieces of arithmetic around the pivot of the equality sign, which is essentially a formal exercise involving little or no reference to the terms' content. It is easy to see how this line of thought might lead someone to the idea that mathematics is nothing more than performing moves in some elaborate game or the shuffling of empty symbols. In this section I will look at the history of Formalism as a position in the philosophy of mathematics, following the historical accounts given in [60] and [65].

Let us begin with two early versions of Formalism, Term Formalism and Game Formalism, which are given their academic pedigree more by the criticisms received from Frege in [24] than by their own positive positions. The proposals for these two types of Formalism criticised by Frege were forwarded by Thomae in [63] and Heine in [32], but as Weir points out in [65] they fail to distinguish properly between the two positions.

First up, Term Formalism is the view that the subject of mathematics is not the abstract referents of mathematical terms but, rather, the terms themselves. By seeing things this way, the hope is that they will avoid some of the trickiest questions in the philosophy of mathematics, those of ontology and knowledge. Their answer provides a straightforward ontology of mathematical terms, which we can know about by writing down tokens of them, manipulating them and stating facts about them. But really, this idea fails to get off the ground; questions of truth mostly fail to have any sort of proper answer and even the ontological simplicity leaves us in a limited position when it is observed that most mathematical objects don't have individual terms to refer to them (think of the real numbers, most of which do not have specific terms that name them or to write them down with). It is not entirely clear what the Term Formalist is even to make of proof: do they show us new properties and relations between the tangible signs of mathematics? If so, then it seems they step beyond the bounds of Term Formalism, but, if not, it is necessary to give an account of what proofs do do.

A slightly stronger proposal is Game Formalism which draws on the idea that mathematics is about nothing more than knowing the rules for mathematical 'games'. For each part of mathematics there are rules as to how we may manipulate and move between propositions, calculations and formulae; rules which make up elaborate games which are, on this view, the essence of mathematics. Any further meaning of mathematical terms and objects can be set aside when actually practicing or playing the game of mathematics, or even more radically be said not to exist at all. As with Term Formalism, there is a strong nominalism running through this stance: mathematics isn't about objects but about the rules[8], epistemology is simply the knowledge of rules and truth is established by showing

[8] The over-used analogy to chess comes in handy here, for it is easy to argue that the pieces on the

that something follows about the game pieces of mathematics by these rules. Even proof makes more sense now because it can be seen as a description of some series of turns of the game being played out. Although this position may be more coherent than Term Formalism, it is still not particularly convincing in its given form. For one thing, examining mathematical practice reveals that the game analogy isn't perfect because mathematicians are guided by content and meaning. Furthermore, the point of rules in a game is usually to get somewhere and achieve some goal (to win the game!) but if maths is just a collection of rules, what is the point? what guides us to find some results interesting rather than others? Frege's criticism [24] of Thomae's version of Formalism [63] presses the point that mathematics is highly applicable in ways that other games are not, which at least warrants explanation.

While both of these two simplistic versions of Formalism were found wanting by Frege, this did not mark the end of Formalism. The development of logic, proof theory and axiomatics at the end of the 19$^{\text{th}}$ led to the possibility of giving a new version of Formalism. A new position, called *Deductivism*, saw mathematics as investigating logical consequences of stipulated axioms. Mathematical practice is about taking a few foundational statements which introduce your primitive terms and discovering the logical consequences of them by explicit logical inferential rules, leading to the moniker "If-Then-ism". Mathematics on this view is about formal proofs and proving as defined in 3.1. A great benefit to this is that it removes all reliance on intuitions in mathematics, making all properties and rules clear and syntactic. Although historically mathematics is heavily rooted in its applications (as described by Maddy in the first chapter of [47]), Deductivism treats the axioms for well-known mathematics as if they are arbitrarily stipulated and reads the primitives to have no content beyond that ascribed by the axioms. Hence the famous story of Hilbert's declaration that

> ... one must always be able to say, instead of "points, straight lines, and planes", "tables, chairs and beer mugs". [31, p. 403] (translation from [60, p. 151])

The point is that the interpretation of these terms is of no relevance to the consequences in the formal system, for the theorems established in mathematics are logical consequences of the axioms of one's system. So if the axioms are interpreted in such a way that they come out true, then it follows as a matter of logical necessity that all of the formal consequences in the system are true on that interpretation too.

Hilbert is commonly seen as the figurehead of Deductivism because of his work in [30] in which he applies the Deductivist methodology to Euclidean geometry, giving axioms for geometry and justifying all inferences based on logical relations. All spatial and geometric intuitions are reduced to mere aids to understanding for the reader. We can apply the

chessboard are not crucial for playing a game of chess: so long as the rules of chess are observed, it would be equally acceptable to draw one's board in the dirt or even picture it in your mind. The ontology of chess is therefore not important, so by analogy that of mathematics need not be either. Such stark minimalism is not to my liking: I will raise an objection to this in section 6.4.

system of geometry by interpreting the primitives in the axioms with actual lines, points etc and thereby also apply the geometric theorems.

The study of deductivism opens the way for new meta-mathematical investigation. As above, models can be given for the axioms to demonstrate satisfiability; models can be given to satisfy some but not all of the axioms, to show the logical independence of the axioms from one another; and questions of consistency can be asked and investigated. Eventually such investigations did lead to problems for Hilbert's grander ambitions[9], but this should not detract from the wide influence Deductivism has had. As evidence for this we can cite the fact that a view along these lines is the target for Rav's criticisms.

I will now brush over several other Formalist positions such as those proposed by Curry, Carnap, Quine and, more recently, Weir, details of which can be found in the Stanford Encyclopedia of Philosophy article [65]. Instead, let me turn to a new version of Formalism that has arisen to meet Rav's challenge.

## 3.5 The Derivation-Indicator View

One of the main Formalist responses to Rav's [55] is Azzouni's paper [3] which outlines what he calls the *Derivation-Indicator View* of mathematical practice.

Azzouni does not wish to argue with those who claim that informal proofs are indispensible in mathematical practice, for just looking at mathematical practice reveals this to be the case. However, he contends that something more than social agreement among mathematicians is required to explain why there is such an overwhelming consensus over what counts as a formal proof. The reason that we are so good at reaching such an agreement, he claims, is that informal proofs *indicate formal derivations*: whenever a mathematician presents an informal proof, that proof actually reveals a derivation of that theorem. Unlike other Formalist positions though, he says:

> ... a standard mathematical proof *indicates* any of a family of derivations without those derivations (1) being what standard proofs abbreviate, (2) being, in some more extended sense, the 'logical forms' of such proofs, or (3) being items that such proofs are 'reducible to'. [4, p. 142]

In an analogous way to proof sketches indicating how a full proof must proceed, Azzouni sees proofs as derivation sketches which indicate how a formal proof would proceed. An informal proof can then be filled out into a full derivation by formalising missing syntactic steps that mathematicians gloss over as obvious or trivial. But informal proofs do not specify which system or language they are working in or what the acceptable manipulation rules are for their indicated formal proofs. This is why he says they indicate families of derivations rather than specific individual formal proofs. For

[9] I am, of course, referring to plan to use Finitary Mathematics to prove the consistency of Ideal Mathematics, shown to be impossible by Gödel's Incompleteness Theorems.

> ...since algorithmic systems embedded in one another are so embedded to conserve derivational results, we can take the derivation indicated to be one located in any algorithmic system within which the result occurs and is surveyable. [3, p. 94]

Furthermore, these systems make use of *implicit* rules and axioms, which mathematicians recognise as being used in the mechanically checkable informal proof[10] without needing them explicitly stated to understand the theorem, because mathematicians know and understand how such rules work. This view of mathematical practice also holds that in practice we do not worry if we add new rules, i.e. augment the system we were implicitly working in with new implicit rules, the only restriction being that the newly augmented system is conservative over the old system. Conservativity is important because this should allow conservative translations between systems in the way that Azzouni describes:

> ...we (implicitly) stipulate a co-referential identification among *terms* in different systems; this in turn induces an identification among statements in different systems, and of course, an identification among concepts. [3, p. 100]

With this added observation, an informal proof will indicate a derivation which can be found in a whole family of formal systems with the conservativity result holding over them. Therefore, Azzouni paints a picture of mathematical practice where

> ... the mathematician is seen... as gracefully sprinting up and down algorithmic systems, many or which he or she invents for the first time. [3, p. 103]

Having outlined the Derivation-Indicator view, in the next section I will examine an argument levelled against it based on the practical limitations of formal proofs.

## 3.6 Grandmother, What Large Derivations You Have

In this section I will consider the argument in [53] where Pelc argues against the Derivation-Indicator version of Formalism by claiming that it is theoretically impossible to mechanically check advanced mathematical theorems once they are given as formal derivations and that therefore we get no epistemic benefit from them. In my view, Pelc's argument demonstrates an underestimation of the field of Formal Mathematics that leads to dismissing the Formalist position for the wrong reasons.

Pelc begins the argument as follows:

> We want to argue that not only do mathematicians not in practice use derivations to get or increase confidence in their results, but that in the present state

[10] That informal proofs are all mechanically checkable seems to follow, for Azzouni, from all mathematics being somehow algorithmic, but how a mechanical check of most informal mathematics should proceed is, according to this thesis, rarely straightforward.

> of knowledge it is theoretically impossible to achieve such a gain of confidence in the case of most interesting mathematical theorems. [53, p. 88]

Pelc's claim is thus that in order for formal derivations to provide an increase in confidence over informal proofs alone there must be at least the theoretical possibility of mechanical verification of that derivation. He then tries to demonstrate that the hypothetical length of derivations corresponding to deep theorems makes it theoretically impossible to carry out such a mechanical check.

As an example (and explicitly avoiding the commitment to this as anything more than an example) Pelc chooses to consider derivations in ZFC as a commonly accepted axiomatisation to act as a foundation for mathematics within which one could theoretically formulate all or most of mathematical reasoning.[11] As an example of a deep theorem with a long and extended proof he selects Fermat's Last Theorem (FLT).[12] Using $\mathcal{L}(T)$ to denote the length of the formal derivation of $T$, he argues:

> We conjecture that in the case of most 'complex' 'deep' theorems $T$, and in particular in the case of FLT, it is impossible to provide (and justify) any upper bound on $\mathcal{L}(T)$. We agree that this is a bold statement but we would like to challenge a skeptical reader to provide (and justify!) any upper bound on $\mathcal{L}$(FLT). [53, p. 91]

Next, Pelc calculates the (extremely large) number of mechanical checks that could potentially be performed in our universe.[13] The result of this calculation, the integer $M$, is a very generous upper bound on the number of individual mechanical checks we could conceivably carry out. From this it is argued that

> Consider a derivation of length larger than $M$. The hypothetical existence of such a derivation of a theorem T could not possibly contribute to our confidence in T because we could never have any kind of access (even theoretically) to all

[11] An interesting issue that arises is that Pelc may not be right to suppose that ZFC is powerful enough to formalise Fermat's Last Theorem. A post of Dec 2007 from the Foundations of Mathematics mailing list in which Timothy Y. Chow quotes an anonymous correspondent discusses this particular pairing. The worry is that the proof of Fermat's Last Theorem uses cohomology theory, which may require the use of Grothendieck universes. These are generally of a consistency strength of ZFC plus countably infinitely many strongly inaccessible cardinals, so strictly stronger than ZFC. Fortunately for us, when we switch to making an estimate for the length of a formal proof of FLT in one of our proof checkers, we can suppose we are using Mizar since the axiomatic basis for this system is Tarski-Grothendieck Set Theory which guarantees the existence of these universes, sidestepping the problem entirely.

[12] Presuming, of course, that Fermat himself did not have a significantly shorter proof as hinted at in that famous margin.

[13] This is done in the following way: he takes the unit of Planck time, the lower bound on the duration of any observable physical event and therefore the shortest amount of time that any one single check could take, and sees how many could fit into the entire duration of the universe (before the heat death caused by reaching maximum entropy), then multiplies this by the number of particles in the universe, which acts as an upper bound on the number of potential parallel processing units.

> the terms of such an extremely large sequence of formulae, and hence we could never verify that it is indeed a correct derivation. [53, p. 93]

Calling a theorem *reachable* if the length of the shortest proof in the chosen axiomatic system (in this case ZFC) is less than $M$, Pelc concludes that since we currently do not have the fact of the matter settled on whether FLT is reachable, it is unknown whether it is theoretically possible to mechanically check the formal derivation of it, so the Formalist cannot claim any gain in confidence for the theorem based on this.

Unfortunately, FLT is not yet among the list of theorems proved in Mizar, nor of any theorem checker. However, the continually growing list of successful formalisations in systems such as Mizar can be used to make a convincing argument in favour of the reachability of FLT or, for that matter, any other given theorem that has been proved informally.

Pelc spends a lot of time emphasising just how huge $M$ is and allows generous over-estimations at every point of its definition. In contrast, the number of computations needed to mechanically check the theorems already proved in Mizar is really rather small. Of course, there is a lot of mathematics to formalise before one could even start to formalise FLT in Mizar: Wiles's proof proceeds by proving a special case of the Taniyama-Shimura-Weil Conjecture (or the modularity theorem now it has been fully proved) which would need Mizar to have considerable background in the modular forms and elliptic curves that the theorem deals with, as well as category theory, Iwasawa theory and many other areas of mathematics that are made use of in the proof. On top of this, Wiles's original paper [68] is over 100 pages long. Now supposing, as Pelc does, that all of the required mathematics could in theory be formalised, the challenge he presents is to give a reason to expect that the final formalisation is reachable.

That reason, I claim, is that for a formalisation of a theorem already proved informally to be unreachable, that is to be longer than the gigantic $M$, would require a massive blow-up of length in the translation from the informal to the formal proof. The choice that Pelc makes to consider ZFC as the formal system is less harmless than it seems: experience tells us that ZFC derivations are particularly long and cumbersome; thereby making us suspect that deeper theorems may get exponentially longer, and that it might well follow for it to be possible for a full derivation to become unreachable. Mizar articles, however, certainly are longer than their informal counterparts but generally this is only estimated to be around four times longer.[14] Thanks to the use of the MML, it is also reasonably clear that this will stay constant rather than growing exponentially as the proofs get deeper. Even if this four were instead one million it would not be sufficient to make FLT even close to unreachable. Coupling this with the efficiency provided by the accumulation of

[14] Based on a survey of articles by Wiedijk found at [66]. Wiedijk is looking at the 'de Bruijn factor', named after N. G. de Bruijn because he describes the phenomenon of 'blow-up' from informal proof to formal proof in [13]. He calculates this by comparing the length of the TEX encoding of various informal proofs to that of their formalisations.

results provided by the MML, it seems obvious to me that any proof that could be humanly written without the aid of a computer, and theoretically has a formalisation, will without a doubt also be reachable.

Recall that the argument provided by Pelc relied on the following:

> Since no upper bound on $\mathcal{L}$(FLT) has been justifiably provided, the answer to the question of whether FLT is reachable is unknown. Given the fact that the number $M$ is so enormous, one would be tempted to give the answer 'yes' to this question, i.e., to establish $M$ as an upper bound on $\mathcal{L}$(FLT). However, as observed before, this has never been done (and seems impossible to do, although we do not need this stronger conjecture in our argument). [53, p. 93]

Well, let me do some calculations. The journal article proof of Fermat's Last Theorem by Wiles [68] has 247,666 characters, but we recall that Wiedijk looked at TEX files rather than the journal article. Based on experience, TEX files are slightly larger but not incredibly so, especially since Wiles's article contains large amounts of text which has essentially no blow up at all. To give a generous bound, let us suppose that for Wiles the TEX file was an order of magnitude larger at around 2.5 million characters.[15] Then I should take the largest scaling (rather than the average) of Wiedijk's survey since we are looking at upper bound: this was Topological Analysis in HOL Light which scaled up 8.8 times from informal to formal proof. Multiplying these up leads us to a generous estimate of 22 million characters for the length of the formal proof of Fermat's Last Theorem.

Is this enough to satisfy Pelc's demand? Certainly I think so: it has been justifiably provided estimate based on empirical data, the proof available and the current forefront of theorem provers. What has been provided is not a proof and is undeniably defeasible, but Pelc's requirement cannot be to *prove* that this upper bound is correct, that is, to provide an indefeasible estimate, as this simply demands too much. Even if a convincing argument could be made that these estimates were not sufficiently generous (which I do not expect to be the case), the given estimate is tiny compared to $M$. I think it fully justified to therefore state that, in all likelihood and by a reasonable estimate, FLT is reachable.

In this section we have seen how advances in Formal Mathematics can be used to defend Formalist positions such as the Derivation-Indicator view from arguments against the inadequacy of formal proofs. Such a link is made more sturdy by looking at [15] by Carl & Koepke of the Naproche project, specifically presenting the Naproche system as a potential way to elaborate some of the vaguer aspects of the Derivation-Indicator view of mathematical practice, such as what "indication" amounts to exactly and how informal proofs go about indicating formal proofs.

In the next chapter I will explore this link more thoroughly.

[15] As a comparison, this thesis only grows by a factor of 1.15 from pdf to TEX.

# Chapter 4

# The Formalisability Thesis

In his discussion of informal and formal proofs (in his terminology: proofs and derivations), Rav considers the following relationship that could obtain between them:

> ...it has been suggested to name *Hilbert's Thesis* the hypothesis that every conceptual proof can be converted into a formal derivation in a suitable formal system: proofs on one side, derivations on the other, with Hilbert's Thesis as a *bridge* between the two. [55, pp. 11-12]

This, though, is entirely opposed to Rav's own views:

> ...there is no proof, in the technical sense, of Hilbert's Thesis... This is an article of faith, and the number of believers in it is constantly dwindling. And even if such a complete formalisation were in principle possible, who will do it and who will guarantee that the formalisation has been carried out correctly, *before* being fed into the computer for computer verification? [55, pp. 35]

What we see in these quotes are the outline of a particular version of the Formalisability Thesis and a particular route for arguing against it. In this chapter I will make the case that the Formalisability Thesis, as it is standardly considered, may be divided into a weak and a strong reading. I will examine what the significance of these differences is and how this should affect the debate surrounding the formalisation of proofs.[1]

A note on terminology: I shall from here on out use 'formalisation' to refer to the process linking informal and formal proofs, and 'formalised proof' as the result of this process.

[1] Much of the work done in this chapter is influenced by a discussion on the Foundations of Mathematics mailing list between December 2007 and January 2008, initiated by Timothy Y. Chow. In particular, the division I will present between the weak and strong readings of the Formalisability Thesis is an attempt to clear up what was a central difficulty in that discussion.

## 4.1 The Formalisability Thesis

It strikes me, in looking through the literature on the relationship between formal and informal proofs, that statements and interpretations of the Formalisability Thesis which are along the same lines as Rav's version of Hilbert's Thesis fail to provide a proper grounding for a fruitful philosophical debate of the issue. This is because these statements conflate two different ways of thinking about proof formalisability. In this section I will extract these two different positions, showing the problem with trying to argue for or against both simultaneously.

As a point of departure, let's first begin with Rav's statement of Hilbert's Thesis:

**HT** Every informal proof can be converted into a formal proof in a suitable formal system.

The main new addition here that I have not yet considered in previous sections is the suitability requirement. That the formal system should be suitable might be justified by not wanting to trivialise HT. For example, any informal proof could in theory be converted into a formal proof in the system which has the theorem to be proved as an axiom, but this will lose all meaning of the informal proof and, furthermore, has no assurance (intuitive or mathematical) that the formal system will be consistent. Although such a procedure would convert informal to formal proofs, it certainly isn't what is intended. Another possibility is that 'suitable' is specifying that each discipline should be formalised in a particular associated system or family of systems. For example, that arithmetic should be formalised to an appropriate formal system of arithmetic while geometry should be formalised in a formal system of geometry etc. This doesn't really match up with practice, though, since formalisation does often cross the borders between different areas of mathematics. For instance, set theory is regularly used to formalise informal proofs from all over mathematics, not just those pertaining to sets. Also, consider as an example Barwise and Etchemendy's [6] where they present their system *Hyperproof* which has a diagrammatic form of first-order logic, which strongly suggests that the boundaries between subfields of mathematics and logic should not be so strictly segregated in the formalisation process. Alternatively, a weaker and more acceptable justification for adding a suitability requirement may simply be that some formal systems are inappropriate for certain informal proofs to be formalised into. It may be that the chosen system is logically too weak, that the formal language cannot express key concepts of the proof or any other such formal limitation which might prevent the formalisation being successful.

For the sake of continuing, and in line with the natural reading of the requirement, from here on I will take a suitable formal system to be one which can adequately represent the concepts, structure and reasoning of the informal proof (so is not deficient in its language or logical strength) and for which there is no suspicion of inconsistency.

I am now claiming that HT separates into two distinct lines of thought concerning formalisability.

The first family of readings are those associated with the Formalist tradition in the philosophy of mathematics, described above in sections 3.4 and 3.5. The Formalist approach to HT is roughly along the following lines:

**SFT** For any correct informal proof and suitable formal system $\Lambda$, there is a corresponding correct formal $\Lambda$-proof unique up to $\Lambda$-sameness.

SFT, here, stands for *Strong Formalisability Thesis* and we recall the notion of $\Lambda$-sameness was introduced in section 1.4. This is a philosophical claim about the nature of informal mathematical proofs and what it means for a proof to be correct. The SFT does contain certain elements that need to be filled out more, and how this proceeds will depend on the philosophical inclinations of the particular Formalist you are dealing with. Let me consider two readings that could be given.

The first reading is the idea that informal proofs abbreviate formal proofs. Such a view could be associated with Game Formalism or Deductivism, where the argument would be that because all mathematics is about formal rules in our game of mathematics (for Game Formalism) or derivations in formal axiomatic systems (for Deductivism), all informally-phrased mathematics is actually just short-hand for a particular formal proof in a suitable formal system $\Lambda$ (unique up to $\Lambda$-sameness). Therefore, this version of the SFT sees formalisation as the process of expanding all such short-hand and hidden syntactic manipulation to reveal the formal proof underlying one's informal proof. However, I believe we should call this abbreviative reading the *Formalist Caricature* for it is not much more than the simplest way of thinking about Formalism and is a rather inadequate position we cannot expect anyone to seriously defend. The second reading, that correct informal proofs indicate formal proofs, is the Derivation-Indicator view we saw Azzouni advocating in section 3.5. On this view, we can see the correspondence of the SFT as that linking a correct informal proof to one of the formal proofs it indicates.

For both the Formalist Caricature and the Derivation-Indicator view, a correct informal proof is only correct in virtue of there being a correct formal proof that it corresponds to in some suitable formal system (up to the sameness equivalence for that system). It also follows on the account of the SFT that for any given informal proof and suitable formal system $\Lambda$, that a formal $\Lambda$-proof will either be right or wrong, in the sense of belonging to the unique $\Lambda$-sameness equivalence class corresponding to the informal proof. The uniqueness condition of the SFT should be emphasised because this is the Formalists' reason for believing anything in informal mathematics can be correct at all: that each informal proof depends on a specific formal correspondent in a formal system, which underlies the informal proof, is abbreviated by the informal proof, is indicated by the informal proof etc. Of course, this proof is only unique in each suitable system $\Lambda$ up to $\Lambda$-sameness, but this equivalence class is specifically taken to track when the mathematician would call formal proofs "the same" so is still in line with the natural understanding of uniqueness.

For the SFT, there is no requirement that the correspondence between an informal proof and the unique formal proof in a given system is actually accessible in a practical way. For example, the Formalist Caricature could hold that an informal does abbreviate a fully formal proof in some system, but that practical limitations prevent reasonable access to it, nonetheless convincing the Formalist that such a fully-expanded version does exist. Likewise, Azzouni says:

> My uses of "indication" (in "derivation-indicator" and the like...) may have been unfortunate in the impression that they can give that the "indicating" of a derivation by an informal proof is supposed to be phenomenologically visible to the working mathematician. This was never my intention... [5, fn. 17]

From this we can conclude that the indication does not necessarily mean that we can track the correspondence that Azzouni's Formalism relies on.

While the Formalist SFT reading places the Formalisability Thesis firmly in the domain of philosophers and conceptual analysis, this does not seem to agree with the fact that HT reads very much like an empirical claim. The question of whether every proof can be formalised is something that we can check, for any individual instance of an informal proof, by going about the business of actually formalising that proof. Whether HT is true or false does seem to be closely related to the success or failure of the field of Formalised Mathematics described in chapter 3. I thus present a second reading of HT that emphasises this empirical dimension of going about the process of formalising informal proofs. I shall call this the *Weak Formalisability Thesis*:

**WFT** For every correct informal proof and a suitable formal system $\Lambda$, a sufficiently skilled agent can transform this proof to a correct formal $\Lambda$-proof (possibly making some non-trivial decisions in the formalisation process that result in formal proofs that are not $\Lambda$-same).

The emphasis on the transformative process of formalisation here is key, for this reading is all about the possibility of actually working through the process. The perspective on formalisation, here, is also one that is agent-centric, in that it is carried out by some sufficiently skilled agent who may be making important decisions which affect to formal proof that is produced. The agent draws upon their understanding of the informal proof to try to give a formal reconstruction of the proof. Formalisation for the WFT is different to the 'correspondence' account emerging from SFTf: the process of formalisation may well be possible in many different ways. It may be that different formalisers will produce different formal proofs by making different decisions as to what the mathematical content of the informal proof is, by making different choices as to how to best represent that content or, most likely, both. The SFT corollary that a formalised proof is either right or wrong, is not warranted for the WFT because many different formalised proofs may be acceptable formalisations for any given informal proof. Take note that these different proofs are not just distinct, but may be sufficiently different to no longer be $\Lambda$-same in the chosen system $\Lambda$. As a result, an informal proof can correspond to a selection of substantially different formal proofs. Implicit in this is also that formalisation must *change* an informal proof.

The WFT lends itself to pragmatic considerations: different choices during formalisation will result in formalised proofs that may be good or bad for different purposes. The choice of which system will be the $\Lambda$ will affect this, but also different decisions may reveal different things about the informal proof. The fact that the informal proof is changed can be beneficial in additional ways, such as casting a known proof in a new light or in revealing new mathematics (as happened in formalising the Four Colour Theorem in [28]).

Since these positions have not been distinguished in this particular way in the past, let me re-emphasise the differences. On one hand we have a proposal about the philosophy of mathematics and the nature of mathematical proofs. A correct informal proof corresponds to a formal proof in some suitable system $\Lambda$, where this formal proof is unique up to $\Lambda$-sameness. The process of formalisation is not itself important except as an epistemic tool to discover the deeper nature of the given proof, but may even be inaccessible to us. On the other hand, there is the agent-centric approach in which the process of formalisation and the decisions that come with it are of main importance. Formalised proofs are no longer necessarily right or wrong but instead are pragmatically useful. Formalisation does change the informal proof, the outcome of which depends on the formaliser and the decisions they make to get there.

## 4.2 From Two Readings to Two Debates

Having separated the two different ways of looking at the Formalisability Thesis, I will outline the two separate debates that take shape around it. To my knowledge, this has not been explicitly discussed in previous works[2] and I believe it is an important observation to make, which I will show by taking some examples of the previous literature where the arguments misfire by conflating the two positions. First, though, let me give very general characterisations of the two debates as they stand.[3]

Formalism, as outlined earlier in section 6.4, has been a prominent position in the philosophy of mathematics so the debate over the Formalist reading of the Formalisability Thesis is fairly well-known. Proponents of this view, including Azzouni, will make a philosophical case for Formalism. They may invoke mathematical practice and examples of pure calculation and symbolic manipulations to support their position, but the use of these is usually to draw upon the readers' intuitions that mathematics is inherently about formal manipulations. As mentioned previously in section 2.1, a common line of argu-

[2] This is not entirely accurate. In [23] Feferman emphasises the difference between the Formalisability Thesis holding that every informal proof can in principle be formalised, and the "Formalist Thesis according to which mathematics has no content but merely consists in following formal rules" ([23, pp. 373-374]). However, his interpretation of the Formalisability Thesis is in fact far closer to the Formalist reading I gave than to what is claimed by the agent-centric approach to formalisation.

[3] Although, I think the lack of distinction between the SFT and the WFT has caused confusion in the past, so the recaps here divide up which of the two debates different lines of thought should be in rather than where they actually are.

ment in favour of Formalism comes from equating formality with rigour, that mathematics is only rigorous if all assumptions and rules of inference are entirely explicit. To argue against Formalism therefore requires an account of mathematical rigour that separates it from formality, or at least reasons to think that such an account will be the correct one. The debate does not concentrate on the practicalities of formalising proofs, nor does it invoke the more advanced areas of formalised mathematics when such actual formalisation is mentioned. This may be because those arguing against Formalism often point out that formalisation changes the proof, an objection which may be effective against the Formalist position but does far less damage against the weak reading of the Formalisability Thesis. The debate can thus be primarily characterised as based on philosophical discussion on both sides, with a focus on giving a correct account of mathematics. The mathematics being treated by the Formalist, however, is often rather idealised so many detractors point out that mathematical practice does not lend great support to Formalism.

The debate over the practical, empirical side of the Formalisability Thesis is different. Whether or not every proof is formalisable, on this reading, is a question which is possible to investigate. An ideal refutation would offer a counter-example, some correct informal proof which could not be formalised.[4] Such a counter-example, however, is not forthcoming, nor should we really expect one. For presenting a proof as 'unformalisable' offers a very clear target for those who believe all proofs can be formalised to attempt to formalise (as I will do in chapter 5!). A slightly more promising avenue of criticism against the agent-centric approach will thus be to give a solid reason why certain parts of mathematics will fail to be formalisable. It seems all that supporters of the WFT need in response to this is to watch as more and more sub-disciplines of mathematics are successfully formalised, while fending off arguments as to why it is conceptually impossible to complete this programme.

Let us see how distinguishing these two readings is advantageous in dealing with the discussion found in previous literature.

The most obvious first candidate for reconsideration is Rav. Above, in the introduction to this chapter, we saw Rav saying:

> And even if such a complete formalisation were in principle possible, who will do it and who will guarantee that the formalisation has been carried out correctly, *before* being fed into the computer for computer verification? [55, pp. 35]

Rav's project is more aimed at undermining the Formalist reading of the Formalisability Thesis. Yet in this quote we find a problem that in essence challenges the WFT, since this is the reading on which we find practical computer verifications (the SFT does not require that we know which formal proof an informal proof corresponds to). A glib reading of Rav's problem could be fended off easily: we can simply deny that there is a 'correct' formalised proof because, as we know, the weaker reading isn't about correct or incorrect, but about the pragmatic benefit of formalised proofs. However, there is a deeper point to

[4] In [43], Leitgeb refers to this as the "Holy Grail"!

Rav's questions which needs a response: how can we say that the formalised proof correctly formalises the informal proof? A pragmatist position, though, allows for practical answers and in this case I think the answer can be as simple as community consensus. As in section 1.4 where I discussed the sameness of proofs for informal proofs and formal proofs separately, suggesting that there is a natural notion of sameness of proofs, there is no reason for the mathematical community not to be able to apply the notion across and between the two categories, especially in systems like Naproche and Mizar which try to follow standard mathematical style. Appealing to a mathematical community has advantages and disadvantages: on the one hand, it explains why there are times when there is serious disagreement about correctness but, on the other hand, it does give an entrance for human fallibility to make an appearance through. Nonetheless, without making the two different debates explicitly distinct such a path to respond to Rav would not have been open to us.

Making the distinction between the two readings also allows us to reject arguments against formalisation of proofs based on the purpose of proof, such as [34] in which Hersh argues that

> All real-life proofs are to some degree informal. A piece of formal argument— a calculation— is meaningful only as part of an informal proof, to complete or verify some informal reasoning. The formal-logic picture of proof is a fascinating topic for study in logic. It is not a truthful picture of real-life mathematical proof. [34, p. 391]

Central to Hersh's argument is that the role of proof in mathematics is to convince and explain, thus informal proofs are the only required type of proof in mathematics. Having observed our distinction, though, it is revealed that such argumentation begs the question against the Formalist and misfires against the proponent of the weaker reading. For the Formalist position is precisely saying that mathematics is only formal manipulations, while the WFT picture does not strictly require the truthfulness to real-life mathematics, since the formalised proof is a reconstruction rather than a picture going for total accuracy. The purpose of giving a formal proof, as Hersh acknowledges, does not necessarily need to be to convince or to explain, but this does not need to lead to his stronger conclusion that the formalised proof is no longer a proof.

Another example of where such a distinction adds clarity is in the work of Robinson who has written extensively on formal and informal proofs and mathematical rigour (and who will figure prominently in the coming chapter 5). In his [59] he does seem to be addressing what amounts to the agent-centric reading of the Formalisability Thesis, yet his problems like

> Formalization of a given informal proof then often turns out to be surprisingly difficult. The translation from informal to formal is by no means merely a matter of routine. In most cases it requires considerable ingenuity, and has the feel of a fresh and separate mathematical problem in itself. [59, p. 54]

and

> There is no formal criterion for judging the correctness of a formalization—indeed, how could there be? The judgement is necessarily intuitive, and it is not at all clear why the intuition should be granted the final word here when it is deemed untrustworthy as an arbiter of the validity of the proof itself. [59, p. 54]

That translation from informal to formal is not a matter of routine is a problem for the AI work Robinson is mainly interested in, but the requirement of having a skilled agent carry out the process is even built into the WFT. Furthermore, the WFT does not see formalisation as revealing that corresponding formal proof which the validity of the informal proof relies on, so the intuitive nature of the judgement is not problematic for this reading. Once again, though, these can be re-phrased as more direct criticisms of the SFT. The defence the SFT would have against Robinson's questions would be the lack of need for the correspondence to be accessible to us. So now it can be asked: how does it make sense for the correspondence to be inaccessible to us (as in the Formalist Caricature) or, more bizarrely, phenomenologically invisible as Azzouni claimed the derivation-indication can be (in the quote used in section 4.1). Certainly, such questions need answers and are now directed at their proper targets.

The main point of separating these debates, however, is that it now allows me to defend two positions which on the conflation of the two debates would have been opposing. In section 6.3 I will argue in favour of the WFT, then in section 6.4 argue against the Formalist reading given by the SFT. Without dividing the Formalisability Thesis properly these stances may have appeared inconsistent, but now I can show that accepting the one while rejecting the other presents the best position regarding proof formalisation.

# Chapter 5

# A Sample Formalisation

This chapter is dedicated to an attempt at going through the practical process of formalisation. I begin by presenting the mutilated chess board example that will be the subject of the formalisation. Next, I extract what I see as the mathematical content of the example, still informally but as a step towards formality, which I then follow up with the formalisation itself. I finish this chapter with some discussion of what the example involved on the part of the formaliser.

## 5.1 Mutilating Chess Boards

In [58], Robinson discusses the relationship between formal and informal proofs. On formal proofs, he says

> ... all systems of formal proof are essentially similar in that they are really no more than notational systems for providing a display or trace of a systematic virtual search computation. [58, p. 270]

Robinson's focus is proof produced by automated theorem *provers* for which this is broadly correct. The essence of Robinson's problem with formal proofs, though, is that while these search algorithms may verify a theorem to have no counter-examples, in doing so they obscure or lose the essence of the proof. He argues that what makes informal proofs preferable is that they still manage to be rigorous, while also presenting the insight that justifies to us and convinces us of the truth of the mathematical claims. The point seems to be that formal proofs as constructed by computers will always fail to be capable of working with such proof ideas, forever doomed to approaching problems by brute force and enumeration:

> What I have tried to do is to suggest why it is that 'real' informal-but-rigorous proofs are so different from the formal ones which our theorem-provers construct, and why I fear they must always remain so. [58, p. 280]

Yet, Robinson's characterisation of formal proofs by automated provers may not necessarily transfer to formal proofs simpliciter thanks to expansive developments in the field of Formal Mathematics that we saw in part 3. For one thing, for automated provers the search for counter-examples makes up the whole of a formal proof whereas in the systems we considered each step of a proof is checked individually. The main difference, then, is that a human formaliser can now retain large parts of an informal proof in the formalisation process. The importance of the agent to the formalisation process will feature heavily in sections 5.4 and 6.3.

Robinson used two main examples to illustrate his points and I will here focus on one of those: the Mutilated Chess Board Problem. This is the classic brain-teaser of whether a chess board with two opposite corners removed can be completely covered by dominoes (where each domino covers two squares of the board). To make the point discussed above, Robinson compares an informal proof with a proof by total enumeration, illustrating the fundamental differences between them. In this part of the thesis I will attempt to formalise the reasoning that was used in Robinson's informal version of the proof. The process of attempting to formalise such an informally presented problem will be helpful to the philosophical quest into the relationship between formal and informal proofs. It should be noted, however, that this formalisation has already been done in Mizar, found in [7]; in section 5.4 I will compare my Naproche formalisation to this Mizar version.

To begin formalising the mutilated chess board problem and the impossibility of domino-covering, we first need its informal proof. The problem and answer are set up with the following two quotes, taken from the earliest recorded sources of the puzzle. So as not to misrepresent the informal proof, let me reproduce these quotes in full.

> An ordinary chess board has had two squares—one at each end of a diagonal—removed. There is on hand a supply of 31 dominos, each of which is large enough to cover exactly two adjacent squares of the board. Is it possible to lay the dominos on the mutilated chess board in such a manner as to cover it completely? [12, p. 157]

> It is impossible ... and the proof is easy. The two diagonally opposite corners are the same color. Therefore their removal leaves a board with two more squares of one color than of the other. Each domino covers two squares of opposite color, since only opposite colors are adjacent. After you have covered 60 squares with 30 dominos, you are left with two uncovered squares of the same color. These two cannot be adjacent, therefore they cannot be covered by the last domino. [25]

Finally, note that in several of the formulations of the problem the colour of the chess board's tiles is not mentioned, but is instead part of what needs to be figured out to find the solution.[1]

## 5.2 Extracting the Mathematics

In the above description of the puzzle, mathematical content is not separated from other narrative embellishments that primarily serve to make the puzzle more fun and engaging. However, to even begin formalising it will be necessary to isolate the mathematics that is at work. In this section I will briefly go about doing this, setting out the mathematical content of the above proof in a stricter mathematical style. The mathematics here is naturally very elementary (as is the nature of the original puzzle), but the point to observe is not the content itself, but the way in which this content relates to the informal version of the proof above.

First, the actual area of the tiles of the chess board and dominoes serves no mathematical purpose. So we can represent each square on the board by natural number co-ordinates. So, using standard notation for cartesian products and natural numbers defined as the set of their predecessors, we get:

We define the *set of squares* $S$ to be equal to the set $(8 \times 8) - \{(0,7),(7,0)\}$. Next, we define the *adjacency relation* $A$, to be a relation on $S$ such that for $(a,b),(c,d) \in S$ we have that $(a,b)A(c,d)$ iff either $a = c \wedge (b = d+1 \vee d = b+1)$ or $b = d \wedge (a = c+1 \vee c = a+1)$.

Finally, we define the mutilated chess board $C$ to be the pair $(S, A)$.

Obviously this satisfies a very crucial property for an adjacency relation:

**Lemma 5.2.1** *The relation $A$ is symmetric.*

**Proof** Suppose $(a,b)A(c,d)$. There are two cases to consider.
*Case 1:* $a = c \wedge (b = d+1 \vee d = b+1)$. But since $a = c$, if either $b = d+1$ or $d = b+1$ then it follows from the definition that $(c,d)A(a,b)$.
*Case 2:* $b = d \wedge (a = c+1 \vee c = a+1)$. Similarly, since $b = d$, if either $a = c+1$ or $c = a+1$ then it follows from the definition that $(c,d)A(a,b)$. □

This, though, only formalises the squares of the chess board, what is still missing is the colouring of black and white tiles. The colourings of the tiles will partition the squares into two sets and there are three immediate ways to go about this. Firstly, one could simply enumerate the squares and assign them each a colour corresponding to the chess board we already know. This is hardly the mathematical way, though, and would lie closer to Robinson's enumerative strategies which we are explicitly trying to avoid! A better way

[1] Recently, at the 2012 *Foundations of Mathematics* conference in Cambridge, Timothy Gowers presented the talk "Could there be a foundation for mathematical discovery?" in which he used the mutilated chess board problem as a central example to motivate how we might examine the process of mathematical discovery. In his talk, the black and white tiling of the board was also presented as part of the puzzle's solution.

would be to give a general rule for partitioning the set. Either, one can give such a general rule for which numbers belong to each set, then check that we do have the property that no two adjacent squares are the same colour, or else, we can define the property using that property and then check it does partition the set. I'm going to use the second of these three approaches, since it should be the easiest to formalise.

We say a square $(a, b) \in S$ is *black* if $a + b$ is even, and we call the set of these squares $B$. We say a square $(a, b) \in S$ is *white* when $a + b$ is odd, calling the set of white squares $W$.

**Lemma 5.2.2** *The set $\{B,\ W\}$ is a partition of $S$.*

**Proof** Every natural number is either odd or even so each square belongs to either $B$ or $W$, so the partition covers $S$, and no square can belong to both, so they are pairwise disjoint. Neither $B$ or $W$ are empty since $(0, 0) \in B$ and $(0, 1) \in W$. $\square$

For the proof, a vital fact is the discrepancy between the numbers of tiles of each colour. Once again this can be checked through counting, but to try to follow the original proof more closely it is preferable to formalise the appeal to the fact that the full chess board has the same number, so removing two leaves one colour with fewer.

**Lemma 5.2.3** *The sets $B$ and $W \cup \{(0, 7), (7, 0)\}$ contain the same number of elements, that is $|B| = |W \cup \{(0, 7), (7, 0)\}|$.*

**Proof** It is straightforward to construct a bijection here. For example $f : B \to W$ where

$$f((a, b)) = \begin{cases} (a, b + 1) & : \text{ if } a \text{ is even.} \\ (a, b - 1) & : \text{ if } a \text{ is odd.} \end{cases}$$

Checking that this is a bijection is straightforward. $\square$

**Corollary 5.2.4** *For the mutilated chess board, we see that $|W| < |B|$.*

**Proof** Follows directly from lemma 5.2.3, since $|W| < |W \cup \{(0, 7), (7, 0)\}| = |B|$. $\square$

Now we can consider the problem of covering the board with dominoes. As with the tiling, we want to cover the whole board, thus a domino covering can be defined as a partition of the whole board into pairs of squares (each pair being one 'domino').

A domino-covering of $S$ is a partition of $S$ into pairs of adjacent squares.

This leads us to the dénouement of the piece:

**Theorem 5.2.5** *There is no domino-covering of $S$.*

**Proof** By contradiction. Suppose there is a domino-covering $D$. Its elements are pairs of adjacent squares such that $\bigcup D = S$.

So let $D'$ be the set of ordered pairs $(a, b)$ such that $\{a, b\} \in D$ and $a \in B$ and $b \in W$. Since each element of $B$ and $W$ belongs to precisely one pair in $D$, it follows that $D'$ is a bijection between $B$ and $W$, contradicting corollary 5.2.4. □

Now we have a solid piece of mathematics that we believe can be formalised, it is time to put this to the test.

## 5.3 Naproche Formalisation of the Mutilated Chess Board

In this section I will give the text of a Naproche formalisation of the Mutilated Chess Board problem, but it comes with some provisos. The first is that Naproche, as a work in progress as described in section 3.3 was not yet up to the task of formalising the problem. Despite the many useful features we saw that Naproche has, the lack of implemented set theory and functions is a prohibitive problem in trying to formalise the mutilated chess board. In order to attempt a formalisation nonetheless, I have simply proceeded as if these features were implemented in the natural way they would be extended from the current version of Naproche. It is important to emphasise, though, that this is an assumption and could end up being done differently or never being done at all! Let me be clear then: this is a "Naproche formalisation" that will not run in the current system of Naproche.

Our other two worries from section 3.3 also reappear. The formal version of the chess board problem that I wish to give involves seeing numbers as sets, which could (if the above problem were resolved) lead to difficulties of types and typing. On top of this, Naproche doesn't yet have a way of referring to other results or a catalogue of results to refer to, but to enumerate all of the basic background information on numbers, sets, pairs, cardinalities etc. needed even for the mutilated chess board example would have left this as a exceedingly long and tedious thesis. Instead, where I think these other facts needed to be referred to and proved, I have supplied a bracketed name/ description of the relevant theorem, lemma or definition that would be invoked. That this was necessary does not seem to present an insurmountable problem for the formalisation, but does reveal a practical limitation of putting the current version of Naproche to use.

The rest of this section is the Naproche formalisation:

`Definition 1: Define` $Squares$ `to be` $(8 \times 8) - \{\langle 0, 7\rangle, \langle 7, 0\rangle\}$`.`

`Definition 2: Define` $Adj(m, n, v, w)$ `iff` $\langle m, n\rangle \in Squares$ `and` $\langle v, w\rangle \in Squares$ `and` $(m = v \wedge (n = w + 1 \vee w = n + 1)$ `or` $n = w \wedge (m = v + 1 \vee v = m + 1))$`.`

`Lemma 1: For all` $m, n, v, w$`, if` $Adj(m, n, v, w)$ `then` $Adj(v, w, m, n)$`.`
`Proof. Let` $m, n, v, w$ `be given.`

Suppose $Adj(m,n,v,w)$. Then by definition 2 $\langle m,n\rangle \in Squares$ and $\langle v,w\rangle \in Squares$ and $(m = v \wedge (n = w+1 \vee w = n+1))$ or $(n = w \wedge (m = v+1 \vee v = m+1))$.
Now precisely one of the following cases holds:
Case 1: $m = v$ and $((n = w+1 \vee w = n+1)$. Then $v = m$ and $((w = n+1 \vee n = w+1))$. Then by definition 2, $Adj(v,w,m,n)$.
Case 2: $n = w$ and $(m = v+1 \vee v = m+1)$. Then $w = n$ and $((v = m+1 \vee m = v+1))$. Then by Definition 2, $Adj(v,w,m,n)$.
So in both cases $Adj(v,w,m,n)$. Qed.

Definition 3: Define $Black(v,w)$ iff $\langle v,w\rangle \in Squares$ and $v+w$ is even.

Definition 4: Define $White(v,w)$ iff $\langle v,w\rangle \in Squares$ and $v+w$ is odd.

Definition 5: Define $B$ to be $\{\langle v,w\rangle|\ Black(v,w)\}$.

Definition 6: Define $W$ to be $\{\langle v,w\rangle|\ White(v,w)\}$.

Lemma 2: $B \subseteq Squares$.
Proof:
Fix $v,w$. Suppose $\langle v,w\rangle \in B$. By Definition 5, $Black(v,w)$. By Definition 3, $\langle v,w\rangle \in Squares$. Qed.

Lemma 3: $W \subseteq Squares$.
Proof:
Fix $v,w$. Suppose $\langle v,w\rangle \in W$. By Definition 5, $White(v,w)$. By Definition 3, $\langle v,w\rangle \in Squares$. Qed.

Lemma 4: If $Adj(m,n,v,w)$ then ($\langle m,n\rangle \in B$ iff $\langle v,w\rangle \in W$).
Proof.
Let $Adj(m,n,v,w)$. By Definition 2, $(m = v \wedge (n = w+1 \vee w = n+1))$ or $(n = w \wedge (m = v+1 \vee v = m+1))$.
Suppose that $\langle m,n\rangle \in B$. Then by Definition 5, $Black(m,n)$. Then by Definition 3, $m+n$ is even.
By Definition 2, $(m = v \wedge (n = w+1 \vee w = n+1))$ or $(n = w \wedge (m = v+1 \vee v = m+1))$.
Now precisely one of the following cases holds:
Case 1: $m = v$ and $n = w+1$. So by Lemma (Addition Properties), $m+n = v+w+1$. Then $v+w$ is odd.
Case 2: $m = v$ and $w = n+1$. So by Lemma (Addition Properties), $m+n+1 = v+w$. Then $v+w$ is odd.
Case 3: $n = w$ and $m = v+1$. So by Lemma (Addition Properties), $m+n = v+w+1$. Then $v+w$ is odd.

Case 4: $n = w$ and $v = m+1$. So by Lemma (Addition Properties), $m+n+1 = v+w$. Then $v+w$ is odd.
So in all cases $v+w$ is odd. So by Definition 4, $White(v,w)$. So by Definition 6, $\langle v,w\rangle \in W$.
Thus $\langle m,n\rangle \in B$ implies $\langle v,w\rangle \in W$.
Suppose that $\langle v,w\rangle \in W$. Then by Definition 6, $White(v,w)$. Then by Definition 4, $v+w$ is odd.
Now precisely one of the following cases holds:
Case 1: $m = v$ and $n = w+1$. So by Lemma (Addition Properties), $m+n = v+w+1$. Then $n+m$ is even.
Case 2: $m = v$ and $w = n+1$. So by Lemma (Addition Properties), $m+n+1 = v+w$. Then $n+m$ is even.
Case 3: $n = w$ and $m = v+1$. So by Lemma (Addition Properties), $m+n = v+w+1$. Then $n+m$ is even.
Case 4: $n = w$ and $v = m+1$. So by Lemma (Addition Properties), $m+n+1 = v+w$. Then $n+m$ is even.
So in all cases $n+m$ is even. So by Definition 3, $Black(m,n)$. So by Definition 5, $\langle m,n\rangle \in B$.
Thus $\langle v,w\rangle \in W$ implies $\langle m,n\rangle \in B$.
Hence $\langle m,n\rangle \in B$ iff $\langle v,w\rangle \in W$. Qed.

Definition 7: Define $Colouring$ to be $\{B,W\}$.

Lemma 5: $Colouring$ covers $Squares$.
Proof. Fix $v,w$. Let $\langle v,w\rangle \in Squares$.
Now precisely one of the following cases holds:
Case 1: $v+w$ is even. Then by Definition 3, $Black(v,w)$. Then by Definition 5, $\langle v,w\rangle \in B$.
Case 2: $v+w$ is odd. Then by Definition 4, $White(v,w)$. Then by Definition 6, $\langle v,w\rangle \in W$.
So in both cases $\langle v,w\rangle \in B$ or $\langle v,w\rangle \in W$.
Hence by Definition (Covers), $Colouring$ covers $Squares$. Qed.

Lemma 6: $Colouring$ is pairwise disjoint.
Proof. Fix $v,w$. Suppose that $\langle v,w\rangle \in Squares$ and $\langle v,w\rangle \in B$ and $\langle v,w\rangle \in W$.
Then by Definition 5, $Black(v,w)$. Then by Definition 3, $v+w$ is even. By Definition 6, $White(v,w)$. Then by Definition 4, $v+w$ is odd. By Theorem (Not odd and even), contradiction.
Thus there are no $v,w$ such that $\langle v,w\rangle \in Squares$ and $\langle v,w\rangle \in B$ and $\langle v,w\rangle \in W$.
Hence by definition (Pairwise Disjoint) $Colouring$ is pairwise disjoint. Qed.

Theorem 1: $Colouring$ partitions $Squares$.
Proof.
By Lemma 5, $Colouring$ covers $Squares$.
By Lemma 6, $Colouring$ is pairwise disjoint.
Hence by Definition (Partitions), $Colouring$ partitions $Squares$. Qed.

Definition 8: Define $f$ from $B$ to $W \cup \{\langle 0,7\rangle, \langle 7,0\rangle\}$ such that $f(\langle m,n\rangle)$ is $\langle m, n+1(mod8)\rangle$.

Lemma 7: $f$ is one-one.
Proof.
Fix $m, n, v, w$.
Suppose $f(\langle m,n\rangle) = f(\langle v,w\rangle)$. Then by Definition 8, $\langle m, n+1(mod8)\rangle = \langle v, w+1(mod8)\rangle$.
Then by Lemma (Ordered Pairs), $m = v$ and $n+1(mod8) = w+1(mod8)$.
By Definition 8, $\langle m,n\rangle \in B$. Then by Lemma 2, $\langle m,n\rangle \in Squares$. Then by Definition 1 $n \in 8$. Hence $n < 8$.
By Definition 8, $\langle v,w\rangle \in B$. Then by Lemma 2, $\langle v,w\rangle \in Squares$. Then by Definition 1 $w \in 8$. Hence $w < 8$.
By Theorem (Modular Arithmetic), $n = w$.
Thus $\langle m,n\rangle = \langle v,w\rangle$. Qed.

Lemma 8: $f$ is onto.
Proof. Fix $m, v, w$.
Suppose $\langle v,w\rangle \in W \cup \{\langle 0,7\rangle, \langle 7,0\rangle\}$.
By Lemma 3, $W \subseteq Squares$. By Lemma (Subsets), $\langle v,w\rangle \in Squares \cup \{\langle 0,7\rangle, \langle 7,0\rangle\}$.
By Definition 1, $Squares = (8 \times 8) - \{\langle 0,7\rangle, \langle 7,0\rangle\}$. Then $\langle v,w\rangle \in (8 \times 8)$.
Then by Definition (Cartesian Products), $v \in 8$ and $w \in 8$.
So there is an $m$ such that $w = m + 1(mod8)$. So by definition 8, there is an $m$ such that $f(\langle v,m\rangle) = \langle v,w\rangle$. Qed.

Lemma 9: $f$ is a bijection.
Proof.
By Lemma 7, $f$ is one-one.
By Lemma 8, $f$ is onto.
Thus by Definition (Bijection) $f$ is a bijection. Qed.

Theorem 2: $Card(B) = Card(W \cup \{\langle 0,7\rangle, \langle 7,0\rangle\})$.
Proof.
By Lemma 9, $f$ is a bijection.
By Definition 8, $Dom(f) = B$. By Definition 8, $CoDom(f) = W \cup \{\langle 0,7\rangle, \langle 7,0\rangle\}$.

So by Definition (Cardinality), $Card(B) = Card(W \cup \{\langle 0,7\rangle, \langle 7,0\rangle\})$. Qed.

Theorem 3: $Card(W) < Card(B)$.
Proof.
Lemma: $W \subset W \cup \{\langle 0,7\rangle, \langle 7,0\rangle\}$.
Proof.
By Definition 4, it is not the case that $White(0,7)$. So by Definition 6, $\langle 0,7\rangle \notin W$. Qed.
Lemma: $W \cup \{\langle 0,7\rangle, \langle 7,0\rangle\}$ is finite.
By Lemma 3, $W \subseteq Squares$. By Definition 1, $Squares \subseteq (8 \times 8)$. By Theorem (Subsets Transitivity), $W \subseteq (8 \times 8)$.
By Definition (Cartesian Products), $\{\langle 0,7\rangle, \langle 7,0\rangle\} \subseteq (8 \times 8)$.
By Lemma (Subset Unions), $W \cup \{\langle 0,7\rangle, \langle 7,0\rangle\} \subseteq (8 \times 8)$
By Lemma (Finite Products), $(8 \times 8)$ is finite.
By Lemma (Finite Subsets), $W \cup \{\langle 0,7\rangle, \langle 7,0\rangle\}$ is finite. Qed.
So by Theorem (Finite Cardinalities), $Card(W) < Card(W \cup \{\langle 0,7\rangle, \langle 7,0\rangle\})$.
By Theorem 2, $Card(B) = Card(W \cup \{\langle 0,7\rangle, \langle 7,0\rangle\})$.
Thus $Card(W) < Card(B)$. Qed.

Definition 9: Define $x$ to domino-cover $Squares$ iff $x$ partitions $Squares$ and for all $y$, if $y \in x$ then there exist $n, m, v, w$ such that ($y = \{\langle n,m\rangle, \langle v,w\rangle\}$ and $Adj(n,m,v,w)$).

Theorem 10: There is no $x$ such that $x$ domino-covers $Squares$.
Proof.
Assume for a constradiction that $x$ domino-covers $Squares$.
By Definition 9, for all $y$, if $y \in x$ then there exist $n, m, v, w$ such that $y = \{\langle n,m\rangle, \langle v,w\rangle\}$ and $Adj(n,m,v,w)$.
Define $g$ from $B$ to $W$ such that $g(\langle m,n\rangle) = \langle v,w\rangle$ iff $y = \{\langle n,m\rangle, \langle v,w\rangle\}$ and $y \in x$.
Lemma: $g$ is one-one.
Proof. Suppose $g(\langle m,n\rangle) = g(\langle v,w\rangle)$. So $g(\langle m,n\rangle) \in W$ and $g(\langle v,w\rangle) \in W$. Then by Lemma 3, $g(\langle m,n\rangle) = g(\langle v,w\rangle) \in Squares$.
By Definition 9, $x$ is a partition of $Squares$. So by Definition (Partition) $x$ is pairwise disjoint.
So there is a $y$ such that $y \in x$ and $g(\langle m,n\rangle) = g(\langle v,w\rangle) \in y$.
Let $y \in x$ and $g(\langle m,n\rangle) = g(\langle v,w\rangle) \in y$.
Then $\langle m,n\rangle \in y$ and $\langle v,w\rangle \in y$.
By Lemma 4, $\langle m,n\rangle \in B$ and $\langle v,w\rangle \in B$.
By Theorem 1, $Colouring$ paritions $Squares$.
So by Definition (Pairs), $\langle m,n\rangle = \langle v,w\rangle$.

`Thus` $g$ `is one-one. Qed.`
`Lemma:` $g$ `is onto.`
`Proof. Suppose` $\langle m,n\rangle \in W$. `By Lemma 3,` $W \subseteq Squares$. `So by Lemma (Subset Transitivity),` $\langle m,n\rangle \in Squares$.
`By Definition 9,` $x$ `partitions` $Squares$.
`So there is a` $y$ `such that` $y \in x$ `and there are` $t,u$ `such that` $y = \{\langle m,n\rangle, \langle t,u\rangle\}$ `and` $\langle t,u\rangle \in B$.
`So there are` $t,u$ `such that` $\langle m,n\rangle = g(\langle t,u\rangle)$.
`Thus` $g$ `is onto. Qed.`
`Hence` $g$ `is a bijection.`
`So by Definition (Cardinality),` $Card(B) = Card(W)$.
`By Theorem 3, Contradiction.`
`Thus there is no` $x$ `such that` $x$ `domino-covers` $Squares$. `Qed.`

## 5.4 Decisions of the Formaliser

Let's look at what sort of conclusions can be drawn from the above example and compare it to another formalisation of the same problem, but in Mizar, given in [7]. The process of formalisation happened in several stages. Before I could even begin formalising I needed to know what exactly it was I was going to attempt to formalise. There was a fair bit of thought involved in getting a simple but effective representation of the board and its relevant properties, leading me to believe I would not have succeeded had I attempted to go straight from the informal proof in [58] to the Naproche proof. Robinson's version of the proof, of course, was presented in a way that emphasised informality, so it does not follow that the first step would be necessary for all formalisations: in my experience, many informal proofs are already sufficiently explicit and exact to begin the formalisation process, but it is worth emphasising that they need not be. On the other hand, though, different proofs may be further from this type of formality (think of, say, the Proofs Without Words mentioned in section 2.1) and therefore require more effort at the earlier stage of extracting the mathematics.

An important point I would like to draw out of the example is the large number of decisions I made at both stages.

Firstly, if we look back at the original problem and compare it to the mathematics involved in the final formalisation, I claim that to a certain extent I actually decided in the extraction phase what the mathematical content of the proof was going to be. Although there are certain aspects of the original informal proof that were going to have to be included for the formalisation to be fairly described as a formalisation of that original proof, there are certainly points at which we could imagine someone else attempting the same process choosing an entirely different focus for the formal proof. For example, I focused heavily on partitions of the squares of the chess board into the tilings and domino-coverings, while ignoring the area and other facts about the two-dimensional representation of the board. A different formaliser may have chosen to go the other way on this entirely. This can be seen by looking at the formalisation in [7] of the same problem but in Mizar. Although structurally similar in many respects (notably, the board is defined in the same way) it is different in others: the colour system made use of modular arithmetic and the structure of the second half of the proof is more general in stating the need for the same number of squares of each colour.

Secondly, there was the choice of background axiomatic system to formalise into. Without too much consideration (and probably due personal academic background) I went ahead with a set-theoretic orientation. It would certainly be possible to attempt a formalisation of the same proof in any number of settings from category theory to Euclidean geometry. The language of set theory and the assumed background of existing sets (or numbers and pairs) meant the chess board was quickly subsumed as a theorem about finite partitions of sets under a particular relation (which I called adjacency) but in another setting this may have formalised into something that looked very different. For example, the given problem is clearly combinatorial and focusing on that aspect of the proof may have yielded a rather different formal proof.

Thirdly, there was the specific choice of exactly how to define each concept that was being used in the original. I chose to abstract the board to just square-coordinates and the adjacency relation, with a tiling as partitions and revealing that a domino-covering was also a bijection between the two sides of the tiling, thus their cardinality must be the same, reaching a contradiction when it is given that there are not the same number of them. Although the informal proof implicitly relies on the tiling being a partition, the rest did require me to make decisions. For this particular proof, these definitions worked well, but had I been working on the actual mathematics of chess and dominoes, a different formalisation may well have been preferred, even if I had still chosen to use numbers and sets as a basis. Had somebody else been formalising it may well have been that they chose entirely different definitions from which the defining features that I used follow and vice versa.

All of these mean that the resulting formal proof that I constructed in Naproche was just one among many ways of formalising the informal proof. Furthermore, had I chosen to use Mizar or one of the other formal checkers instead I would have produced another proof altogether, because even after extracting the mathematics, getting this to fit with the particulars of the formal language of Naproche meant that more decisions were involved.

Another key point is that both main stages of formalisation required understanding of both the informal proof and the formal proof system. In this case the informal proof was so simple that the understanding was likewise simple, but understanding the workings of Naproche took some time and effort, even for such a relatively straightforward system designed to be easy to get the hang of. One implication of this is that the informal proof could not have been transformed into the formal proof by any straightforward method of clarifying all assumptions in some algorithmic way: the formalisation was a very human process.

I should also point out the benefits of doing this formalisation. The ultimate ideal would be for the formal proof to be checkable by Naproche, but as pointed out above this is not yet possible because of the current limitations of Naproche. However, the formalisation did require extra mathematics to be done. For example, the definitions of the adjacency relation and bijections were extra pieces to fill in the gaps left without the reliance on our ability to reason about the chess board that was present in the informal proof. The mutilated chess board is easy, but in more advanced proofs such additional mathematics could be a worthwhile pusuit in itself; for instance, in the formalisation of the Four Colour Theorem discussed in [28] they mention several original pieces of mathematics that add to the understanding of the combinatorics involved in the theorem. Finally, I believe that while the original informal proof was very specific to a single mutilated chess board, the process of formalisation does demonstrate a number of directions for generalisation and further inquiry. For example, in discussing the problem and its formalisation, we found that domino-coverability and having an equal number of black and white squares is not equivalent. With a first formalisation under the belt we may now gain insight on how to formalise other harder problems in combinatorics or related fields.

# Chapter 6

# Explication and the Formalisability Thesis

In this concluding chapter I will return to the Formalisability Thesis in light of the lessons learned in Chapter 5 by working through the mutilated chess board example of formalisation. Based on these lessons I will firstly argue in favour of the Weak Formalisability Thesis and secondly attack the Strong Formalisability Thesis.

## 6.1 Carnap on Explication

Following the discussion of the development of Carnap's ideas presented in [10], although it took until 1945 for Carnap to introduce the term 'explication' in [17] and another five years after that before the methodology received a full treatment at the start of [18], the roots of the idea were firmly implanted in Carnap's writings from much earlier, such as in his 'rational reconstruction' found in [16]. The usage of such methods of investigation are very closely related to the project of this thesis since they try to link informal concepts to formal equivalents, as is part of what is done in proof formalisation.

In Carnap's terminology, explication is

> the transformation of an inexact, prescientific concept, the explicandum, into a new exact concept, the explicatum. [18, p. 3]

The purpose of explication is to gain new understanding and insight on this explicandum, as well as to replace an informal or natural language concept with a scientific one which is useful in more academic research. By talking about replacement it emerges that the explicatum is something that we introduce and that it is something different to the explicandum. The change in meaning does imply that there will be cases where we cannot substitute the explicatum for the explicandum, but this is not the role of the explicatum;

rather, its purpose is to serve as a more fruitful and precise concept for the more exacting scientific realms.

A prerequisite for being successful at explicating is to know what it is we want to find out more about, which is to say that while the explicandum is inexact, this does not mean anything goes. Before a concept is explicated, Carnap emphasises that there is the need to be as clear as possible about that concept. This is not explication, but a part of formulating the problem: we do have methods for reaching a reasonable clarity of even inexact concepts and can work to have general agreement as to what falls under the concept and what does not. For example, I hope to have achieved this much clarity in section 2.1 regarding the concept of informal proof.

Once a target explicandum is settled and clear to the best of our abilities, we wish to go about the transformation of explication. Carnap's process of explication sets four requirements for an explicatum which I shall briefly discuss in turn: similarity to the explicandum, exactness, fruitfulness and simplicity. The similarity requirement is what ensures we stay close to our original explicandum. How is this meant? Well, it could be meaning that our explicandum and explicatum coincide as much as possible in their extensions. But Carnap rejects this requirement as too strict, prefering to allow the similarity requirement to be malleable around the other three requirements. A principled replacement of items in the extension of the explicandum or being far stricter in what is falls under the explicatum (in comparison to the explicandum) may serve the scientific purposes motivating the explication better, thus similarity should be a sufficiently weak condition to allow such decisions to be made on the part of the explicator.

The second criterion of explication is fruitfulness:

> A scientific concept is the more fruitful the more it can be brought into connection with other concepts on the basis of observed facts; in other words, the more it can be used for the formulation of laws. [18, p. 6]

Carnap also states that if the concept being explicated is a logical concept (as opposed to an empirical one) then the explicatum should be useful in formulating logical laws. The fruitfulness of a new concept is demonstrated by it fitting well into the web of other concepts in the scientific discourse.

The third requirement is that of exactness, that the explicatum is explicitly defined such that

> the rules of its use (for instance, in the form of a definition), is to be given in an exact form, so as to introduce the explicatum into a well-connected system of scientific concepts. [18, p. 7]

The explicatum should be defined precisely such that we can accurately determine what falls under the concept and when it can be used in scientific discourse.

The final requirement is that of simplicity, although this secondary to the other three requirements. If there are multiple explications that are possible that offer equal degrees of exactness and fruitfulness, it is generally preferable to choose the more parsimonious concept as one's explicatum. This requirement does not compel us to reject complicated explications, unless there is an equally good explication which is simpler.

Having been through the nature of explication, one important point made by Carnap that is particularly relevant here is that

> ...if a solution for a problem of explication is proposed, we cannot decide in an exact way whether it is right or wrong. Strictly speaking, the question whether the solution is right or wrong makes no good sense because there is no clear-cut answer. The question should rather be whether the proposed solution is satisfactory, whether it is more satisfactory than another one, and the like. [18, p. 4]

There is no reason to say that one explication is somehow correct, because the very point of explication is that one concept is replaced by another in certain areas of discourse. It may well be that one explication is by far the most suitable for a particular area, but this does not make it the right explication only the best.

## 6.2 Proof Formalisation and Explication

In this section I will argue that for some systems $\Lambda$ the concept of formal proof in $\Lambda$ explicates the concept of informal proof. This has already been noted by Sjögren in [61] as the main thrust of his paper. The difference between here and there is that Sjögren treats the matter mainly as an exercise in explication, with no consideration of the consequences of this assertion, while I will continue from here into the consequences for proof formalisation.

To show that formal proof is the explicatum to informal proof's explicandum we must check that the requirements listed in section 6.1. First, I note an opponent to this thesis in Leitgeb's [43] who declares that these conditions fail to obtain:

> [W]hy not think of "formally provable(-in-T)" (for some instantiation of "T") as a Carnapian explication of "informally provable"? The answer is simple: because it is not. According to Carnap, whatever explicates an explicandum must be as similar as possible to the latter, but as our comparison... has shown, formal provability and informal provability are just too dissimilar to satisfy this criterion. There is no reason to believe that if one could explicate informal provability at all, then this could not be done while preserving more of its essential features than any explication in terms of "formally provable(-in-T)" would ever achieve. [43, p. 272]

Leitgeb's claims here could be charitably read as based on taking the instantiations of "T" in a traditional, proof-theoretic way (as, it would appear, does Sjögren), say, as formal proofs in natural deduction or sequent calculus. In this particular case, it is almost tempting to agree with Leitgeb for such proofs do feel very different from informal proof. Yet having now investigated Formal Mathematics and the controlled natural language of Naproche, this position is less appealing, as in these cases the formal proofs do appear much more similar to the informal counter-parts.

Other than those of form, the dissimilarities that refers to are along the same lines as those I have considered above and mostly consist of the fact that informal proofs are vague and unspecific in a number of respects, while formal proofs are explicit and precise about everything. But these dissimilarities are not the sort to prevent explication: if anything they sound exactly like the pre-conditions for explication! Replacing an inexact concept with an exact one is what explication is for. Disregarding this, even, Carnap's requirement that they be as similar as possible is qualified with the extra liberality that one may, within reason, put fruitfulness and exactness first. Leitgeb seems to be casting aspersions on these as well, but we already saw in our example that such an explication is very fruitful, on top of which formal proofs are by their very nature exact.

When it comes to fruifulness, the discussions of sections 3.2 and 3.3 only focused on some local advantages, like mechanical checkability, the potential for new mathematics, encouraging the formaliser to get to understand the mathematics more thoroughly, the scope for generalisation etc. Add to this the general advantages of formal provability:

> This explication has made it possible not only to give precise formulations to vague questions concerning e.g. axiomatisability, consistency, decidability, proof strength, etc., but also in providing mathematical methods for answering them. On the whole, one must say that the explication is in fact extremely successful. [61, p. 455]

Leitgeb suggests that we may find a more similar explication of informal provability: formally provable (-in-Naproche) or formally provable (-in-Mizar) appear to fulfill this role.

## 6.3 Explication and the Formalisability Thesis

Although it is interesting to consider whether the concept of informal proof explicates to the concept of formal proof (in some system), this thesis is focused on what is for that explication at the object level: moving from individual informal proofs to individual formal proofs. Here, though, I think that something similar is still the case. While on first glance explication happens on the conceptual level, which individual formal and informal proofs fall under, this does not necessarily block the way to saying that proof formalisation can also be explicative.

So how can proof formalisation be explicative? Well, even though an informal proof itself will fall under the general concept of informal proof that we have been discussing, we also know that such a proof is not a discrete singularity. What I mean is, an informal proof in mathematics will make use of many other mathematical concepts. In the process of formalising informal proofs we will formalise these concepts to correspond to some formal equivalents. Doing so will certainly give more exact concepts by its very nature, and I have already discussed the fruitfulness of formal mathematics. Simplicity decreases in the one sense because formal versions of concepts are almost invariably less intuitive, but in another this makes them far simpler by replacing concepts that Rav can claim have "irreducible semantic content" [55, p. 11] with concepts having only a few basic and algorithmic rules. Lastly, similarity is a less clear case, for usually formalising a proof involves switching to formal *objects*, as well as concepts. The need for a large extensional overlap might well not be satisfied, say, if we replace each object with it's formal equivalent.

But the need for extensional overlap is all well and good when Carnap exemplifies the explication of the concept 'fish' as in the example of [18, p. 6], but the actual similarity requirement is given as:

> The explicatum is to be similar to the explicandum in such a way that, in most cases in which the explicandum has so far been used, the explicatum can be used; [18, p. 7]

When it comes to mathematical objects, this may still obtain if the formal representatives of the objects have the correct properties. The systematic replacement of objects described as a part of explication in section 6.1 will in this case be exact, fruitful and simple, thus warranted by the Carnapian mode of explication.

The formalisation of informal proofs, though, is clearly not just explication of it's various mathematical concepts. An additional factor is in play: the explication is *uniform*. Each relevant mathematical concept is not merely explicated in some way; rather, each concept is explicated in the same way. For one thing, a single formal language is present within which each of the concepts must be expressed. Secondly, the many mathematical concepts in a proof will interact and interrelate so it must be ensured that the explications stay coherent. As a result, the systematic replacement of the subject matter of your proof (numbers, points, etc.) with formal counterparts will be reflected in the explication of the higher concepts ranging over these objects, while the explication of concept must be in line with the explication of the concepts which it is defined in terms of.

At this point, though, the picture is still no quite complete. What is also required is the reworking of inferences and dependencies. Where before there were many ways of arguing, still appealing to understanding and intuition, now there are specific rules of manipulation which strictly enumerate the ways in which one can reason. The new formal objects will also have axioms governing their use, exiling meanings to the meta-language. In part, the replacement of interrelations and consequences will be dictated by the choice of formal objects and the explication of the mathematical concepts, but additionally we also (uniformly) explicate the reasoning structure to conform to the axioms and inference rules of our system. This naturally will involve the expansion of many pieces of reasoning to conform to the explicit nature of the formal deduction rules, but there will be similarity here because the formalisation process must maintain a degree of similarity to properly be said to be a formalisation of the original informal proof.

Thus, on this picture, the formalisation process from informal to formal proofs involves the *systematic replacement* of objects and the *uniform explication* of concepts and reasoning patterns. Is it still correct to call this explication? I believe so, for one can still maintain many aspects of the original informal proof, enough to maintain similarity or else to not be able to call the formal proof a formalisation of the informal proof. Certainly the formal proof is exact, with clear rules for each concept and their interrelations. Formalisation is also fruitful, as has been noted, in that it leads to mechanical checkability, the discovery of new mathematics, generalisation, insights into the intricate web of mathematical dependencies and allows for interesting meta-mathematical results. Simplicity follows from the fact that even if as a whole the proof becomes less readable and intuitive, each step is purely mechanical, meaning that the formal proof is sufficiently simple to not require any understanding of the mathematical content.

For example, in the mutilated chess board example I systematically replaced the objects, squares on the board, with set-theoretic ordered pairs which gave their co-ordinates on the board. From here we could uniformly explicate concepts such as adjacency, colourings and domino-coverability into formal counterparts. The reasoning steps could then be explicated such that the same overarching reason for the correctness is maintained, while also expanding the justification for many of the more implicit, intuitive steps.

In section 5.4, it was seen that the process of formalising the mutilated chess board problem did involve decisions on the part of the formaliser. Formalisation, I noted, requires understanding of the problem being worked on and the formal system you want to formalise the informal proof in. This coincides neatly with Carnap's explication which involves intelligent analysis and understanding. But, more importantly, this works towards the Weak Formalisability Thesis with its agent-centric view on proof formalisation. If the agent can make enough decisions and has the relevant understanding of the informal reasoning and a good formal system, then of course they can give a formalisation of that proof. This has been shown by the success of Formal Mathematics and our own example (recalling that the example was chosen as one that was seen as particularly informal).

What is there to stop an agent with sufficient understanding of the informal proof from carrying out a formalisation? In section 3.6 we rejected that the length of formal proofs is a sufficient problem now, given the success of systems like Mizar, Coq, HOL Light and Isabelle. For any given proof it seems that the formaliser will have to note the relevant properties that the explicated concepts need to have (this part of the very purpose of the process) so even if a concept were semantically irreducible in general, there does not seem to be a problem in listing its salient properties. It may be that diagrammatics are more problematic, but once again the chess board included important visual, albeit it imagined, information that was not too challenging to convert into formal reasoning. To my mind, it is hard to even conceive of a diagram which would contain uninterpretable mathematical content to challenge the idea that every informal proof can be formalised.

Returning now to two readings of the Formalisability Thesis, I believe that what we have seen of this explicative process and the chess board example of chapter 5 have served two major functions. Firstly, seeing formalisation as explicative in this way and observing the fact that even in our own example we made non-trivial decisions which lead to substantially different (i.e. not $\Lambda$-same, where $\Lambda$ is Naproche) formal proofs refutes the SFT by violating the uniqueness condition. Secondly, the chess board example and the argument for explication give reasonable inductive evidence in favour of the WFT, since the chess board example was chosen as something hard to formalise because of its informality, but nonetheless, taking on the role of agent myself the formalisation managed to proceed. Additionally, the fact that non-trivial decisions needed to be made shows that other formalisers may have reached substantially different formalisations of the problem, as is suggested by the WFT. In its original discussion, I declared the WFT to be an empirical claim which can be investigated; through the mutilated chess board example I hope to have begun such an investigation.

## 6.4 Informal Rigour

I would like to add to the refutation of the SFT claimed in the previous section a *coup de grâce* for the Formalists by dismissing their favourite defence: that rigour requires formality. I hereby add to the literature, such as [2], defending informal rigour.

The argument is that only formal deduction is free from the inherent fallibility of the intuitions invoked in informal proving. In my opinion, this is obviously not the case. Let me consider for a moment the distinction between *formal consequence* and *material consequence* as discussed, for example, by Read in [56]. Formal consequence involves a conclusion following from the premises because of the form of the argument. Read associated this with the interpretational account of validity:

> an argument is valid if there is no (possible) interpretation of the expressions (other than a reserved class of "logical" expressions) in the argument under which the premises come out true and the conclusion false. [56, p. 250]

On the other hand, material consequence links premises to a conclusion based on the meaning of terms contained in them, following the representational account of validity:

> an argument is valid if there is no possible situation where the premises are true and the conclusion false. [56, p. 250]

Read argues that the representational account better captures actual validity than the interpretational account; consequence should be material not formal. As an example of an argument which is materially but not formally valid, Read gives:

> (5) Iain is a bachelor
> So Iain is unmarried. [56, p. 249]

He goes on to consider the Suppressed Premise Strategy, open to those who would argue in response that (5) is only valid based on the additional suppressed premise that 'All bachelors are unmarried'.[1] But Read rejects this strategy:

> The extra premise is strictly redundant. For if the original argument were invalid, the added premise would not be logically true. Given that it is logically true, it follows that the unexpanded argument was already valid. Hence it was (logically) unnecessary to add the extra premise. [56, p. 249]

The point of this detour into consequence is that the same can be seen in informal proof. It is not hard to construct examples of the exact same phenomenon:

> (1) The natural number $n$ is even
> So the natural number $n$ is divisible by two.

As in Read's example, the Formalist could point to the suppressed premise 'All even natural numbers are divisible by two', but the same reply that this additional premise is redundant may be offered. Without needing to go as far as Rav and say that there is irreducible semantic content in mathematical terms, we can still say that there *is* semantic content. So, then, informal proofs in mathematics may be rigorous through their consequences being *materially valid* on the representational account of validity. Such a solution to the rigour and indefeasibility of mathematics is also a more natural solution, in line with the look and feel of mathematical practice and not requiring reduction, abbreviation or indication to a formal proof. Formalisation and formality are not required for mathematical rigour.[2]

(There is a slight worry for the use of material validity in mathematics that needs addressing. The problem is that true mathematical facts are commonly seen as *necessarily* true. Therefore, any mathematical fact follows from no premises or any premises by the representational account of validity. So it is possible to construct all kinds of consequences, e.g.

> (2) Every natural number is odd or even.
> Therefore, there are no positive integers $a,\ b,\ c,\ n$ such that $a^n + b^n = c^n$ and $n > 2$.

[1] This idea is not a new one, being found in Buridan:

> It seems to me that a material consequence is evident in inference only by its reduction to a formal one. Now it is reduced to a formal one by the addition of some necessary proposition or propositions whose addition to the given antecedent produces a formal consequence. [14, Ch. 4].

Here translated by Read himself and obviously present in his work.

[2] Note also that validities on the interpretational view are a subclass of validities on the representational view, meaning that if we take mathematics to be rigorous by material consequences, then it does still include the pieces of formal reasoning that maths obviously involves.

The response, though, is to bite the bullet on consequence, but to note that consequence alone is insufficient for proof. Firstly, the argument (2) fails on all acounts of the purpose of proof: it does not convince, explain, provide mathematical knowledge, or any of the other purposes of proof outlined in sections 1.2 or 2.1. For to be convinced of the validity of the inference, one would first need to be convinced that the conclusion is true, thus failing a key requirement of proof as laid out in chapters 1, 2 and 3.)

The Formalist response might be to still maintain that the above fails because mathematical terms only have semantic content in their guise as abbreviations of or indications to formal definitions. Key to Azzouni's arguments for the Derivation-Indicator view is the attempt to form a coherent defence of nominalism in the philosophy of mathematics. This response can be seen as trying to rescue the formality of mathematics by an appeal to nominalism. But in doing this the Formalist is in danger of circularity, for often the formality of mathematics is used to lead us to a Formalistic nominalism (see, for example [60, pp. 140-141] or [65]). The usual order for eliciting the intuitions of the Formalist is to convince us that mathematics is reducible (or somesuch) to formal manipulations and thereby reach the nominalism. I thus settle here with the burden on a Formalist to provide an independent justification of mathematical nominalism, one more convincing than the picture I have offered of material validity which fits mathematical practice so neatly.

## 6.5 Conclusion

By separating two different readings of the Formalisability Thesis, I have carved out a position that lies somewhere between those of Rav and Azzouni.

I have favoured the agent-centric weak reading of the Formalisability Thesis, in which formalisation is seen as a practical process that we can carry out as a fruitful analytic tool. Any proof can be subjected to this process, so long as the agent understands the proof and the target formal system. From an in-depth example of formalising a concrete proof, I showed that the process invariably requires the agent to make decisions; decisions that lead to substantially different (i.e., not $\Lambda$-same for the formal system $\Lambda$) formal proofs, which, if correct, gives good reason to hold the WFT. Drawing on Carnap's notion of explication, I have argued that these different formal proofs *can* all be legitimate formalisations of the original informal proof. Nonetheless, seeing it as possible to correctly formalise an informal proof in not just one way but many different ways, each providing different gains and benefits, goes against Rav's doubt that every proof can be formalised and negative assessment of the positive achievements of doing this.

In contrast, I have rejected the strong reading of the Formalisability Thesis, which proclaims the Formalist position: that informal proofs abbreviate, indicate or otherwise depend on formal counterparts. The demonstration that decisions are made in the process of formalising which give substantially different outcomes is directly against the uniqueness of the SFT. By furthermore arguing that material consequence, rather than formal consequence, is at the heart of the mathematical practice of proving, I hereby take mathematics to have semantic content independent of any formal reductions or indications, contra Azzouni and other Formalists and rule out the argument that mathematical rigour depends on formality.

Finally, I shall end on the titular moral of the story here presented: holding that all informal proofs can be formalised does not make you a Formalist.

# Bibliography


[1] Alibert, D. & Thomas, M. (1991): "Research on Mathematical Proof", in Tall, D. (ed.), (1991): *Advanced Mathematical Thinking*, Dordrecht: Springer Netherlands, pp. 215-230.

[2] Antonutti Marfori, M. (2010): "Informal Proofs and Mathematical Rigour", *Studia Logica*, Vol. 96 No. 2, pp. 261-272.

[3] Azzouni, J. (2004): "The Derivation-Indicator View of Mathematical Practice", *Philosophia Mathematica (3)* Vol. 12 No. 2, pp. 81-105.

[4] Azzouni, J. (2005): "How to Nominalize Formalism", *Philosophia Mathematica (3)* Vol. 13 No. 2, pp. 135-159.

[5] Azzouni, J. (2009): "Why Do Informal Proofs Conform to Formal Norms?", *Foundations of Science* Vol. 14 No. 1-2, pp. 9-26.

[6] Barwise, J. & Etchemendy, J. (1994): *Hyperproof*, Stanford: CSLI.

[7] Bancerek, G. (1995): "The Mutilated Chess Board Problem— checked by Mizar", in Matuszewski, R. (ed.): *The QED Workshop II, Technical Report No. L/1/95*, pp. 37-38.

[8] Baker, A. (2008): "Experimental Mathematics", *Erkenntnis* Vol. 68 No. 3, pp. 331-344.

[9] Beall, JC (1999): "From Full Blooded Platonism to Really Full Blooded Platonism", *Philosophia Mathematica (3)* Vol. 7 No. 3, pp. 322-325.

[10] Beaney, M. (2004): "Carnap's Conception of Explication: From Frege to Husserl?" in Awodey, S. & Klein, C. (eds.): *Carnap Brought Home: The View from Jena*, Chicago: Open Court, pp. 117-50.

[11] Bertot, Y. (2010): "Coq in a Hurry", Version 5, available at ⟨http://cel.archives-ouvertes.fr/inria-00001173⟩.

[12] Black, M. (1946): *Critical Thinking*, Prentice Hall.

[13] de Bruijn, N. G. (1999): "A Survey of the Project Automath", in Seldin, J. P. & Hindley, J. R. (eds.): *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, Orlando: Academic Press Inc., pp. 579-606.

[14] Buridan, J. (1976): *Iohannis Bvridani Tractatvs de Conseqventiis*, Hubien, H. (ed.), Read, S. (trans.), Louvain-Paris.

[15] Carl, M. & Koepke, P. (2010): "Interpreting Naproche— An Algorithmic Approach to the Derivation-Indicator View", in Pease, A., Guhe, M. & Smaill, A. (eds.) *Proceedings of the International Symposium on Mathematical Practice and Cognition*, pp. 7-10.

[16] Carnap, R. (1928): *Der logische Aufbau der Welt*, Berlin-Schlachtensee: Weltkreis-Verlag.

[17] Carnap, R. (1945): "The Two Concepts of Probability", *Philosophy and Phenomenological Research* Vol. 5 No. 4, pp. 513-532.

[18] Carnap, R. (1950): *Logical Foundations of Probability*, Chicago: University of Chicago Press.

[19] The Coq Development Team (2012): *Coq Reference Manual*, Version 8.4, available at ⟨http://coq.inria.fr/distrib/V8.4/refman/⟩.

[20] Dawson Jr., J. W. (2006): "Why Do Mathematicians Re-prove Theorems?", *Philosophia Mathematica (3)* Vol. 14, pp. 269-286.

[21] Došen, K. (2003): "Identity of Proofs Based on Normalization and Generality", *Bulletin of Symbolic Logic* Vol. 9 No. 4, pp. 477-503.

[22] Feferman, S. (2000): "Mathematical Intuition vs. Mathematical Monsters", *Synthese* Vol. 125 No. 3, pp. 317-332.

[23] Feferman, S. (2012): "And so on... : Reasoning with Infinite Diagrams", *Synthese* Vol. 186 No. 1, pp. 371-386.

[24] Frege, G. (1903): *Grundgesetze der Arithmetik 2*, Olms: Hildesheim.

[25] Gardner, M. (1957): "Mathematical Games" in *Scientific American* Iss. March 1957.

[26] Grabowski, A., Korniłowicz, A. & Naumowicz, A. (2010): "Mizar in a Nutshell" in *Journal of Formalized Reasoning* Vol. 3 No. 2, pp. 153-245.

[27] Goldrei, D. (2005): *Propositional and Predicate Calculus: A Model of Argument*, London: Springer.

[28] Gonthier, G. (2008): "Formal Proof— The Four Colour Theorem", *Notices of the American Mathematical Society* Vol. 55 No. 11, pp. 1382-1392.

[29] Harrison, J. (2012): "HOL Light Tutorial", Version 2.20, available at ⟨http://www.cl.cam.ac.uk/∼jrh13/hol-light/tutorial_220.pdf⟩.

[30] Hilbert, D. (1899): *Grundlagen der Geometrie*, Leipzig: Teubner; Townsend, E. (trans.): (1959) *Foundations of Geometry*, La Salle, Ill.: Open Court.

[31] Hilbert, D. (1935): *Gesammelte Abhandlungen, Dritter Band*, Berlin: Springer.

[32] Heine, E. (1872): "Die Elemente der Funktionslehre", *Crelle's Journal für die reine und angewandte Mathematik*, Vol. 74.

[33] Hersh, R. (1991): "Mathematics Has a Front and a Back", *Synthese* Vol. 88 No. 2, pp. 127-133.

[34] Hersh, R. (1993): "Proving is Convincing and Explaining", *Educational Studies in Mathematics* Vol. 24 no. 4, pp. 389-399.

[35] Hersh, R., (1997): "Prove— Once More and Again", *Philosophia Mathematica (3)* Vol. 5 No. 2, pp. 153-165.

[36] Kamp, H. & Reyle, U. (1993): *From Discourse to Logic: Introduction to Model-theoretic Semantics of Natural Language*, Dordrecht: Kluwer Academic Publishers.

[37] Kolev, N. (2008): *Generating Proof Representation Structures in the Project NAPROCHE*, M.A. Thesis at Rheinischen Friedrich-Wilhelms-Universität zu Bonn, available at ⟨http://naproche.net/inc/downloads.php⟩.

[38] van Kerkhove, B. & van Bendegem, J. P. (2008): "Pi on Earth, or Mathematics in the Real World", *Erkenntnis* Vol. 68 No. 3, pp. 421-435.

[39] Kühlwein, D., Cramer, M., Koepke, P. & Schröder, B. (2009): "The Naproche System", available at ⟨http://naproche.net/inc/downloads.php⟩.

[40] Lakatos, I., (1976): *Proofs and Refutation*, Cambridge: Cambridge University Press.

[41] Landau, E. (1930): *Grundlagen der Analysis*, Leipzig: Akademische Verlagsgesellschaft m. h. B.

[42] Larvor, B. (1998): *Lakatos: An Introduction*, New York: Routledge.

[43] Leitgeb, H. (2009): "Why Do We Prove Theorems?", in Bueno, O. & Linnebo, Ø. (eds.): *New Waves in Philosophy of Mathematics*, Basingstoke: Palgrave Macmillan, pp. 263-299.

[44] Löwe, B. & Müller, T. (2008): "Mathematical Knowledge is Context Dependent", *Grazer Philosophische Studien*, Vol. 76, pp. 91-107.

[45] Löwe, B. (2010): "Comparing Formal Frameworks of Narrative Structures" in Finlayson, M. (ed.): *Computational Models of Narrative: Papers from the 2010 AAAI Fall Symposium*, Menlo Park 2010, [AAAI Technical Report FS-10-04], pp. 45-46.

[46] Löwe, B. (2011): "Methodological Remarks about Comparing Formal Frameworks for Narratives", in Allo, P. & Primiero, G. (eds.): *Third Workshop in the Philosophy of Information, Contactforum van de Koninklijke Vlaamse Academie van België voor Wetenschappen en Kunsten*, Brussel 2011, pp. 10-28.

[47] Maddy, P. (2011): *Defending the Axioms*, Oxford: Oxford University Press.

[48] Müller-Hill, E. (2011): *Die epistemische Rolle formalisierbarer mathematischer Beweise*, PhD Thesis, Bonn: Rheinischen Friedrich-Wilhelms-Universität Bonn.

[49] Nakamura, Y., Watanabe, T., Tanaka, Y. & Kawamoto, P. (2002): *Mizar Lecture Notes* $4^{th}$ Edition, Nagano City, Japan: Shinshu University Department of Information Engineering.

[50] Nelsen, R. B. (1993): *Proofs Without Words: Exercises in Visual Thinking*, Washington: The Mathematical Association of America.

[51] Nelsen, R. B. (2000): *Proofs Without Words II: More Exercises in Visual Thinking*, Washington: The Mathematical Association of America.

[52] Nipkow, T., Paulson, L. P. & Wenzel, M. (2012): *A Proof Assistant for Higher-Order Logic* New York: Springer-Verlag available at ⟨http://www.cl.cam.ac.uk/research/hvg/Isabelle/dist/Isabelle2012/doc/tutorial.pdf⟩.

[53] Pelc, A. (2009): "Why Do We Believe Theorems?", *Philosophia Mathematica (3)* Vol. 17 No. 1, pp. 84-94.

[54] Priest, G. (2008): *An Introduction to Non-Classical Logic*, Second Edition, Cambridge: Cambridge University Press.

[55] Rav, Y. (1999): "Why Do We Prove Theorems?", *Philosophia Mathematica (3)* Vol. 7 No. 1, pp. 5-41.

[56] Read, S. (1994): "Formal and Material Consequence", *Journal of Philosophical Logic* Vol. 23 No. 3, pp. 247-265.

[57] Renz, P. (1981): "Mathematical Proof: What It Is and What It Ought To Be", *The College Mathematics Journal* Vol. 12 No. 2, pp. 83-103.

[58] Robinson, J. A. (1991): "Formal and Informal Proofs" in Boyer, R. S. (ed.): *Automated Reasoning*, Kluwer Academic Publishers: pp. 267-282.

[59] Robinson, J. A. (1997): "Informal Rigor and Mathematical Understanding" in Gottlob, G. et al. (eds.): *Computational Logic and Proof Theory*, *Lecture Notes in Computer Science* Vol. 1289, pp. 54-64.

[60] Shapiro, S. (2000): *Thinking About Mathematics*, Oxford: Oxford University Press.

[61] Sjögren, J. (2010): "A Note on the Relation Between Formal and Informal Proof", *Acta Anal* Vol. 25 No. 4, pp. 447-458.

[62] Tarski, A. (1969): "Truth and Proof", *Scientific American*, pp. 63-70 & pp. 75-77.

[63] Thomae, J. (1898): *Elementare Theorie der analytischen Functionen einer complexen Veränderlichen*, Second Edition, Halle.

[64] Schwichtenberg, H. & Troelstra, A. S. (2000): *Basic Proof Theory*, Second Edition, Cambridge: Cambridge University Press.

[65] Weir, A. (2011): "Formalism in the Philosophy of Mathematics", *The Stanford Encyclopedia of Philosophy (Fall 2011 Edition)*, Zalta, E. N. (ed.), available at ⟨http://plato.stanford.edu/archives/fall2011/entries/formalism-mathematics/⟩.

[66] Wiedijk, F.: "The 'De Bruijn Factor' ", available at ⟨http://www.cs.ru.nl/∼freek/factor/index.html⟩.

[67] Wiedijk, F.: "Formalizing 100 Theorems", available at ⟨http://www.cs.ru.nl/∼freek/100/index.html⟩.

[68] Wiles, A. (1995): "Modular elliptic curves and Fermat's Last Theorem", *Annals of Mathematics* Vol. 142 No. 3, pp. 443-551.