On the Difficulty of Discovering Mathematical Proofs

Andrew Arana, Will Stafford

ON THE DIFFICULTY OF DISCOVERING MATHEMATICAL PROOFS

ANDREW ARANA, ANDREW ARANA, AND WILL STAFFORD

In memory of Mic Detlefsen

Abstract. An account of mathematical understanding should account for the differences between theorems whose proofs are "easy" to discover, and those whose proofs are difficult to discover. Though Hilbert seems to have created proof theory with the idea that it would address this kind of "discovermental complexity", much more attention has been paid to the lengths of proofs, a measure of the difficulty of *verifying* of a *given* formal object that it is a proof of a given formula in a given formal system. In this paper we will shift attention back to discovermental complexity, by addressing a "topological" measure of proof complexity recently highlighted by Alessandra Carbone (2009). Though we will contend that Carbone's measure fails as a measure of discovermental complexity, it forefronts numerous important formal and epistemological issues that we will discuss, including the structure of proofs and the question of whether impure proofs are systematically simpler than pure proofs.

1. Introduction

It is a truism that every mathematician wants simple proofs. One need only attend to the recent controversies over the abc conjecture and its alleged proof by Mochizuki to grasp the troubles incurred by complex proofs.[1] Mathematicians have always sought not only to prove new theorems but also to do so in simple ways.

A celebrated instance of this is Descartes' analytic geometry. Descartes canonised a procedure for solving geometrical problems as follows: first express the problem by algebraic equations, then solve these equations by algebraic manipulations, and finish by translating these algebraic solutions back into geometrical terms. He lauded this method for making it "easy" [*aisé*] to find constructions, though he noted that sometimes the method requires "dexterity" [*adresse*] in order to find "short and simple" [*courtes et simples*] constructions.[2]

The authors would like to thank Ryota Akiyoshi, John Baldwin, Anna Bellomo, Adrien Champougny, Walter Dean, Serena Delli, Mic Detlefsen, Jeremy Heis, Brendon Larvor, Koji Mineshima, Mitsu Okada, Tabea Rohr, Jeffrey Schatz, Sean Walsh, Kai Wehmeier, Richard Zach, and two anonymous referees for their helpful comments on drafts of this paper. They would also like to thank the audiences at Keio University, the Logic Colloquium 2017, the Midwest PhilMath workshop 2020, and the Applied Proof Theory conference at the University of Melbourne in 2020 for their questions.

[1] The importance of this issue can be seen by its appearance is the popular (and semi popular) press; for example `https://www.nature.com/articles/d41586-020-00998-2`.

[2] Cf. [Descartes, 1637], p. 351, though statements of this sort are found throughout *La géométrie*. For more on the simplicity of the Cartesian method in geometry, cf. [Arana, 2016], §2, and [Maronne, 2010].

Note that Descartes distinguishes here between two types of simplicity: the simplicity of the construction itself, and the simplicity of *discovering* a construction to solve a problem. This distinction, between the simplicity of a proof itself, and the simplicity of *discovering* a proof of a theorem, has been stressed by Michael Detlefsen, who writes:

> There are, of course, various complexity metrics that have found their way into the proof-theoretic literature, and the recent literature in theoretical computer science has produced even more. Yet all of these complexity metrics seem to be designed to measure a general type of complexity that might be called 'verificational complexity'; that is, the type of complexity that is encountered in determining of a given syntactical entity whether or not it is a proof in a given system of proofs. (Cf. [Detlefsen, 1990], p. 376f24; also [Detlefsen, 1996], p. 87)

Such complexity is the subject of a great deal of work today in automated proof verification and the adjoint area of automated software verification. The comparison of the length of proofs, valuable as it is, requires that we have proofs to compare in the first place. As Michael Potter has put it, "it is not much help that a short proof exists if we cannot find it" (cf. [Potter, 2004], p. 236).

Detlefsen contrasts verificational complexity measures with what he here calls "inventional complexity" measures: "the type of complexity that is encountered in coming up with a proof in the first place". We will, as Detlefsen himself did in other places, call such measures *discovermental complexity* measures.[3] It is unsurprising that mathematicians have long reflected upon this type of complexity. As we mentioned above, Descartes viewed his analytic geometry as an advance over classical synthetic geometry in virtue of its superior discovermental power. Leibniz too viewed his differential and integral calculus as easing the search for new theorems, writing that "what is better and more useful in my new calculus is that it yields truths by means of a kind of analysis, and without any effort of the imagination, which often works as by chance, and it gives us the same advantages over Archimedes, which Viète and Descartes gave us over Apollonius".[4] In his *Encyclopédie* entry on the application of algebra to geometry, d'Alembert too lauded the gain in discovermental power afforded by analysis in geometry, remarking that its methods enable us to "arrive nearly automatically at results giving the theorem or the problem that we

[3]This term first appears in print in [Detlefsen and Arana, 2011]. But Detlefsen gave a talk entitled "Discovermental Complexity & the Evaluation of Hilbert's Program" at the University of Nancy in 2008.

[4]Letter to Huygens, 21 September 1691. "Il est vray, Mons. comme vous jugés fort bien, que, ce qu'il y a de meilleur et de plus commode dans mon nouveau calcul c'est qu'il offre des verités par une espece d'analyse, et sans aucun effort d'imagination, qui souvent ne reussit que par hazard, et il nous donne sur Archimede tous les avantages que Viete et Des Cartes nous avoient donnés sur Apollonius." Cf. [Huygens, 1950], Volume X, p. 157.

sought, which otherwise we would not have gotten or would only have gotten with much effort."[5]

It would be easy to multiply such comments from mathematicians throughout its history. Yet these comments do not include any systematic reflection on the notion of proof complexity itself, neither of the verificational nor of the discovermental type. Descartes, Leibniz and d'Alembert made their remarks drawing on their impressions as practitioners of geometry, rather than on a systematic investigation of the nature of proof itself. Such a program only becomes thinkable with the advent of Hilbert's proof theory. Indeed, Detlefsen has suggested that Hilbert had envisioned proof theory as an investigation of both verificational and discovermental complexity, drawing attention to the following passage:

> For this formula game is carried out according to certain definite rules, in which the *technique of our thinking* is expressed. These rules form a closed system that can be discovered and definitively stated. The fundamental idea of my proof theory is none other than to describe the activity of our understanding, to make a protocol of the rules according to which our thinking actually proceeds. Thinking, it so happens, parallels speaking and writing: we form statements and place them one behind another. If any totality of observations and phenomena deserves to be made the object of a serious and thorough investigation, it is this one. (Cf. [Hilbert, 1927], p. 475)

In this passage, Hilbert suggests a focused descriptive study of the laws of thought, where "thought" includes the ways in which we "form statements and place them one behind another". Hilbert seems to have thought that our reasoning occurs in a sequential way, and that an isomorphic "protocol of the rules" of this sequential reasoning ought to be a central goal of proof theory. As Detlefsen reads Hilbert, proof theory should reflect the ways in which we in practice form chains of mathematical reasoning. That is, it should study the ways in which we discover proofs.

Proof theory has not fully followed through on Hilbert's suggestion, Detlefsen laments. Rather little progress has been made on identifying formal measures of discovermental complexity, which could yield a precise analysis of what propositions and what proofs are more difficult to discover than others. We thus turn our attention in this paper to a candidate for such a measure, the genus measure of proof complexity developed by Alessandra Carbone (2009), building on the unpublished Ph.D. thesis of Richard Statman (1974). The genus of a graph is the least genus of the surfaces the graph can be drawn on without any lines crossing, and the genus of a surface is the maximum number of cuts that can be made to it before the surface becomes disconnected. Both Statman and Carbone show how to extract graphs from proofs by representing their

[5]Cf. [Diderot and d'Alembert, 1751], vol. 1, p. 551. The original reads, "arrive presque machinalement à un résultat qui donne le théorème ou le problème que l'on cherchoit, & auquel sans cela l'on ne seroit point parvenu, ou l'on ne seroit arrivé qu'avec beaucoup de peine."

logical structure, and measure the complexity of a proof by the genus of the graph so extracted from it.

In this paper we will evaluate the genus measure of proof complexity as a measure of discovermental complexity, rather than of verificational complexity. A proof whose interferential structure is "convoluted" is intuitively harder to discover than one whose inferential structure is linear. Whether a proof is convoluted in this way seems to be related to its genus, for higher genus means ineliminable crossings among the edges of the graph representing its inferential structure. Intuitively, then, genus complexity is a good candidate for a measure of discovermental complexity because it captures the idea that a "convoluted" proof is hard to discover. And if the context of discovery proves to be too opaque for formal measurements, convolutedness might still represent an interesting measure of difficulty of understanding as distinct from difficulty of verification. Just as size is not the only roadblock to discovery, it is also not necessarily the best measure of understandability.

Though neither Statman nor Carbone mention discovermental complexity in their works, each formulates their motivations for their work on genus in ways that can be read as bearing on discovermental complexity. Statman claims that the genus of a proof is a measure of the *global* structural complexity of a proof, as opposed to the local structural complexity of a proof as determined by the logical complexity of its formulas. Representing the inferential structure of a proof as a configuration of formulas linked by edges, that is, by a graph, he sets out to study "the global structure of these configurations i.e., how the individual inferences fit together" ([Statman, 1974, p. vi]). He writes that "if this [global structural complexity] is not at a manageable level a proof will not even begin to be understood" ([Statman, 1974, p. v]). If a measure of global structural complexity is a good measure of discovermental complexity, then he will have produced a measure of discovermental complexity.

Carbone opens her 2009 article with the mission statement, "We shall not ask *why* we prove a statement, nor *how* to show a statement, but how *difficult* it is to prove it" (p. 139). The term "difficult" here is ambiguous; it can be read as applying to the verification of the proof's validity, as in verificational complexity, or to its discoverability. Since our project is not an exegesis of Carbone's work, but more generally on measuring the discovermental complexity of proof, we note simply that Carbone's aims are consistent with studying genus as a candidate for such a measure.

Our plan for the paper is as follows. In Section 2 we will develop Carbone and Statman's genus measures of proof complexity. In Section 3 we will argue that Carbone's measure fails as a measure of discovermental complexity, showing that in a certain way Statman's measure is more successful. In Section 4 we will address the related claim by Carbone and others that "impure" methods lower discovermental complexity, when the latter is measured by genus. Finally, in

Section 5 we will address a general concern about the relevance of formal considerations to the study of proofs as carried out by mathematicians in practice.

2. Introducing genus as a measure of discovermental complexity

In this section, we will first explain the proof formalisms used by Carbone and Statman. Next, we will explain the graph theory on which Carbone and Statman's complexity measures are based. Finally, we will turn to their main result.

2.1. **Carbone's formalism.**

Carbone works in a sequent calculus for propositional logic. Lines in a sequent calculus proof, called sequents, are written $A_1, \ldots, A_n \Rightarrow B_1, \ldots, B_m$ where $A_1, \ldots, A_n$ is called the antecedent and $B_1, \ldots, B_m$ is called the succedent. The sequent can be interpreted as $A_1 \wedge \cdots \wedge A_n \rightarrow B_1 \vee \cdots \vee B_m$. So on the assumption of A_1 to A_n it follows that *one of* B_1 to B_m holds.

Carbone's sequent calculus has one axiom $A \Rightarrow A$. The remaining rules are separated into two categories. Logical rules introduce logical connectives: $\neg, \wedge, \vee$. If they do this on the left of the sequent, they are left introduction rules and if they do this on the right, they are right introduction rules. The logical rules in Carbone's system are as follows:

$$\frac{\Gamma, A \Rightarrow \Delta}{\Gamma, A \wedge B \Rightarrow \Delta} \qquad \frac{\Gamma_1 \Rightarrow A, \Delta_1 \quad \Gamma_2 \Rightarrow B, \Delta_2}{\Gamma_1, \Gamma_2 \Rightarrow A \wedge B, \Delta_1, \Delta_2}$$

$$\frac{\Gamma_1 \Rightarrow A, \Delta_1 \quad \Gamma_2 \Rightarrow B, \Delta_2}{\Gamma_1, \Gamma_2, A \vee B \Rightarrow \Delta_1, \Delta_2} \qquad \frac{\Gamma, A \Rightarrow \Delta}{\Gamma \Rightarrow A \vee B, \Delta}$$

$$\frac{\Gamma, A \Rightarrow \Delta}{\Gamma \Rightarrow \neg A, \Delta} \qquad \frac{\Gamma \Rightarrow A, \Delta}{\Gamma, \neg A \Rightarrow \Delta}$$

Carbone's system does not contain $\rightarrow$.

The second type of rules are structural rules. Logical rules provide ways to infer information from the premises. Structural rules, however, can be thought of as restructuring the current information. Carbone's system has contraction rules, which reduce two occurrences of the same formula in either the antecedent or the succedent to one formula occurrence, and cut.

The cut rule is:

$$\frac{\Gamma \Rightarrow \Delta, A \quad A, \Gamma \Rightarrow \Delta}{\Gamma \Rightarrow \Delta}$$

Here A is called the cut formula, because it appears in both premises but is 'cut' from the conclusion. Reasoning with cut can be compared to reasoning with lemmas in informal proofs. With cut we break up a proof of the conclusion by first providing a proof of A (the lemma) and then showing that the conclusion follows from this lemma A.

It is worth pointing out that logical rules can have features that parallel the structural rules. Carbone's system only has the structural rules of contraction and cut. But that does not mean

her system is weaker than systems with additional structural rules.[6] Anything we want to do with the structural rules can be done with features of the logical rules that parallel the structural rules. For example, the left $\wedge$ rule above implicitly involves weakening as we are allowed to include the formula B in $A \wedge B$ despite it not occurring in the antecedent of the premise. When we look at Statman's natural deduction calculus there will be no explicit structural rules, but logical rules will still have features that parallel the structural rules.

2.2. **Statman's formalism.** Statman uses a natural deduction system rather than a sequent calculus. Unlike the sequent calculus, a line in the natural deduction calculus is just a formula. However, as well as axioms and rules, we can have assumptions. An assumption is a formula we can introduce at the beginning of a proof without applying any rules. While the sequent calculus has left and right introduction rules but no way to remove (or eliminate) a logical connective, the natural deduction calculus has rules for eliminating connectives and rules for introducing them. As the antecedent of a sequent can be though of as the assumptions from which the succedent is proven, left introduction rules can be thought of as parallel to elimination rules. In natural deduction we break down assumptions and then build up the conclusion, while in the sequent calculus both the conclusion and the assumptions are built up from atomic formulas.

Statman's system consists of the following rules:

$$\text{Introduction} \qquad \frac{\begin{array}{c}[F]\\ \vdots\\ G\end{array}}{F \to G} \qquad \frac{F \quad G}{F \wedge G} \qquad \frac{F}{F \vee G}$$

$$\text{Elimination} \qquad \frac{F \to G \quad F}{G} \qquad \frac{F \wedge G}{G} \qquad \frac{F \vee G \qquad \begin{array}{c}[F]\\ \vdots\\ C\end{array} \qquad \begin{array}{c}[G]\\ \vdots\\ C\end{array}}{C}$$

Note that $\neg$ is not in Statman's system.

Assumptions can come in two varieties: open assumptions which indicate exactly what needs to be assumed for the conclusion to follow, and closed or discharged assumptions which are assumptions that have had $\to$ introduction or $\vee$ elimination applied to them. In the schematic proof rules above, these are the formulas that have [and] around them, e.g. $[F]$. When we have a proof in a natural deduction system that ends with A and has open assumptions $B_1, \ldots, B_n$, we can write $B_1, \ldots, B_n \vdash A$ to represent this. What this tells us is that A follows on the assumption

[6] E.g. weakening rules allow the addition of an arbitrary formula to either the antecedent or the succedent. Merger rules take multiple sequents as premises and return a sequent with an antecedent composed of all the premises' antecedents and a succedent composed of all the premises' succedents.

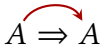
of $B_1, \ldots, B_n$. Note that when an assumption is discharged it no longer needs to be assumed for the conclusion to follow.

As mentioned earlier, there are no structural rules in this system. However, there are features that parallel the structural rules. When we discharge assumptions by an application of $\rightarrow$ introduction or $\vee$ elimination, we are allowed to discharge multiple occurrences of the same formula. We are also allowed to apply the rules when we have no assumption to discharge or to not discharge assumptions we could have. This freedom of discharging assumptions gives us features that parallel the structural rules which work similarly to contraction in the case of multiple discharges and weakening when nothing is discharged.

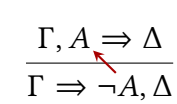
Another point of difference between the two approaches is that Carbone offers a multi-conclusion proof system for classical logic while Statman offers a proof system for minimal logic. However, Carbone's formal results on graphs can be transferred to a single-conclusion restriction of her sequent calculus. This would be done by using the left and right negation rules to "store" multiple conclusions in their negated form. This would result in an intuitionistic system.

2.3. **From proofs to graphs.** Carbone and Statman apply graph theory to proofs for their proof complexity measures. Beginning with a proof, each has a means of extracting a graph from the proof. The complexity of a proof can then be measured by the complexity of the resulting graph. Just as each has their own proof formalism, each has their own means of graph extraction. We discuss each in turn.

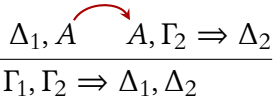
Carbone uses the "logical flow graph" of a proof, a means of extracting a graph from a proof that was introduced in [Buss, 1991] in the course of proving that it is undecidable whether a formula has a proof of k or fewer lines. If we look at the rules of the sequent calculus we see tokens of the same type of formula occurring in the premises and the conclusion. A logical flow graph is a graph that connects the atomic formulas in these tokens together. For an example of a rule, see Figure 1b; for an example of a logical flow graph, see Figure 2. (Logical flow graphs are orientated which is why those depicted have arrows on their edges, but this plays no role in our discussion and so we ignore it.) There are two special rules for drawing logical flow graphs shown in Figure 1a and 1c.



(A) Logical flow graph for an axiom.



(B) Flow graph for $R\neg$



(C) Flow graph for cut

FIGURE 1

Since Statman's proof system is different from Carbone's, he uses a different method of extracting a graph from a proof. His "derivation graphs" are composed of connections between

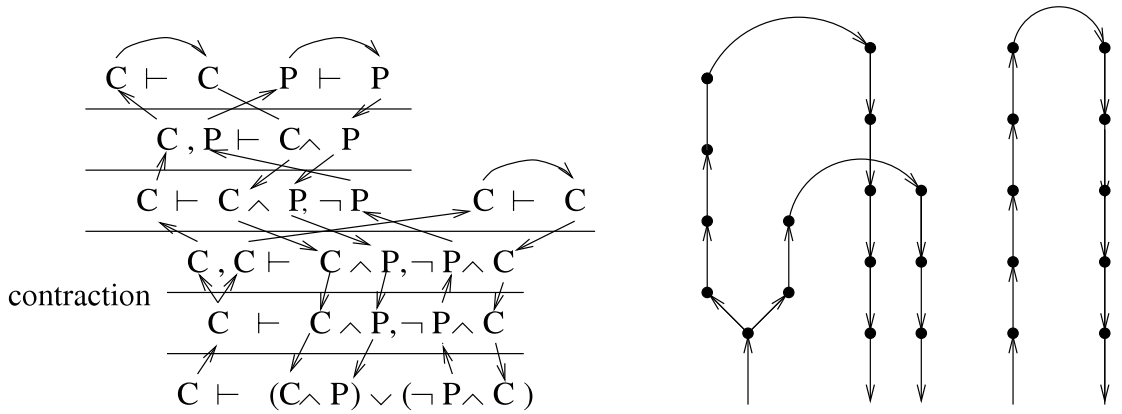


FIGURE 2. An instance of a logical flow graph, from [Carbone, 2009]

complex formulas, unlike logical flow graphs which only connect the atomic formulas. Statman's graphs are generated by first taking the tree obviously generated by the proof and then adding lines between closed assumptions and the formula that discharges them. These proofs are illustrated by Figure 3 and 4, with the rules listed in the table below.

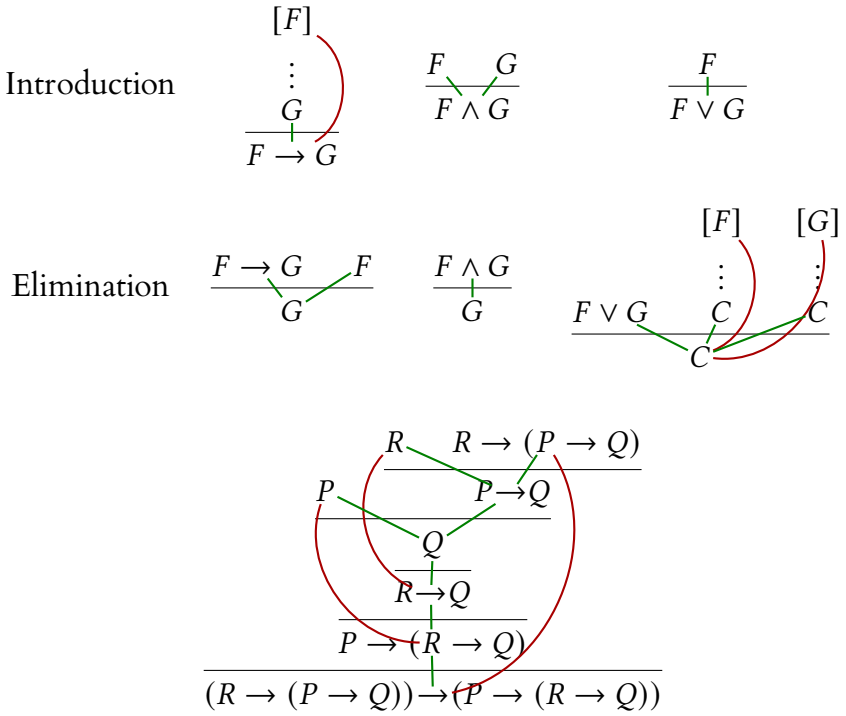


FIGURE 3. An example of a natural deduction proof of genus 0

2.4. **From graphs to proof complexity.** Now that the measures have been described, we can outline Carbone's formal result on cut-free proof complexity. An important piece of context here is Statman's result that in the propositional sequent calculus, the length of cut-free proofs

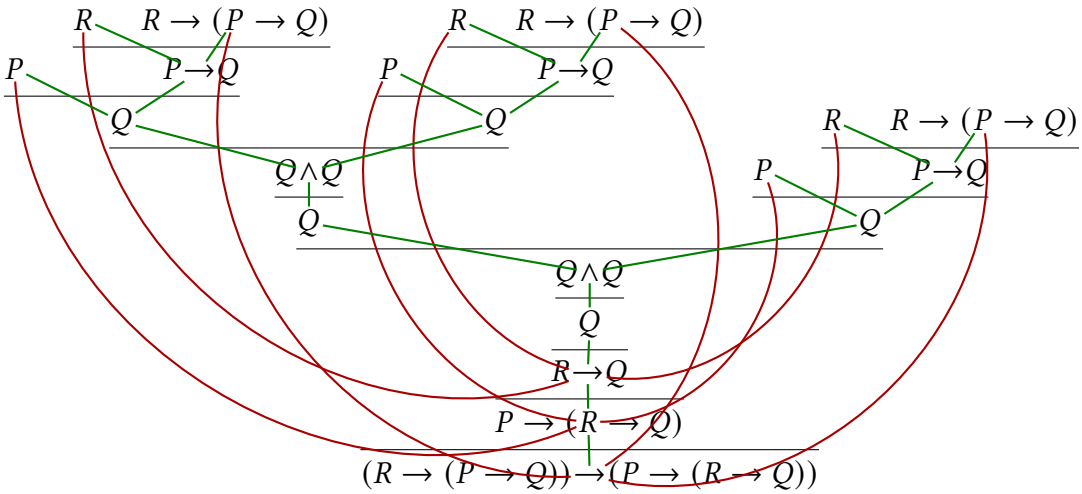


FIGURE 4. An example of a natural deduction proof of genus 1

may be significantly larger than that of proofs with cut. More precisely, he showed that there are sequents whose cut-free proofs are exponentially longer than their proofs with cut.[7] Carbone observes that while proofs with cuts may be shorter than cut-free proofs, proofs with cuts seem, a priori, difficult to discover, because precisely which cut formulas (lemmas) should be used is typically not obvious. Nevertheless, she points out, in practice we use cuts anyway. The "topological genus" of the logical flow graph of a proof, therefore offers another measure of complexity with which to consider this problem. Similarly, Statman focuses on the topological genus of the derivation graph of a proof. We thus turn next to the crucial notion of the genus of a graph.

2.5. **Measuring graph complexity by genus.** In order to describe the topological genus of a graph, or for short its genus, we begin with the notion of a "crossing". The graph depicted in the left of Figure 5 has an edge crossing, while the graph on the right does not. But these two graphs are *isomorphic*: there is a one-to-one correspondence between the points of these two graphs which preserves adjacency of points by edges. Thus, graphs with crossings are sometimes isomorphic to graphs without crossings, and we say that such graphs can be drawn without crossings (even if a representation of that graph has a crossing). We will talk abstractly about a graph G as a collection of points and a relation that says which edges connect which points. This abstract description of a graph relates to the more familiar drawing of a graph by being what isomorphic drawings of a graph have in common. We call a drawing of a graph an embedding of the underlying graph on the surface.

A graph is *planar* if it can be drawn in the plane in such a way that no edges cross. The graphs in Figure 5 are planar graphs. Not all graphs are planar, however. Consider for example the bipartite graph called $K_{3,3}$, shown in Figure 6, which has two pairs of three vertices such that

[7] Cf. [Statman, 1978], and [Buss, 1988] (pp. 2-1 to 2-5) for an excellent presentation of these results.

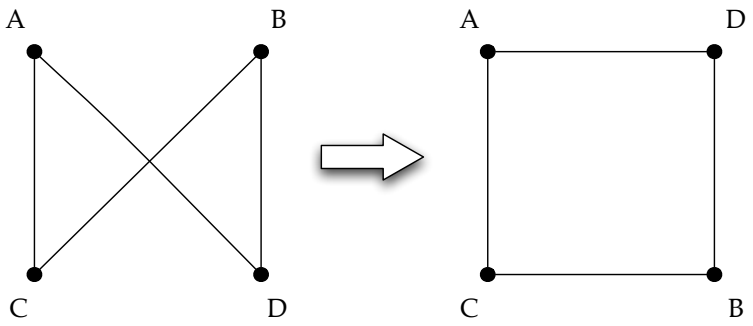


FIGURE 5. Planar graphs

each vertex in the first pair is connected to each vertex in the second pair. As shown in Figure 6, we can try to "unravel" the crossings in the original representation of $K_{3,3}$ on the left, but no matter what we do we seem to be stuck with a crossing. In fact it can be proved that one crossing is necessary, using Euler's polyhedron formula. This formula says that for a planar graph, $V - E + F = 2$, where V denotes the number of vertices in the graph, E the number of edges, and F the number of "faces", that is, regions bounded by edges (cf. [Harary, 1969], pp. 103–104). For $K_{3,3}$ to be planar, it would need a representation with 3 faces. But $K_{3,3}$ always has at least 4 faces and so does not satisfy Euler's polyhedron formula.

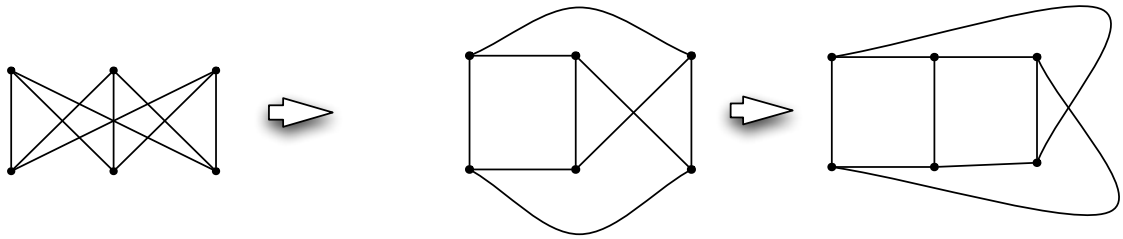


FIGURE 6. A non-planar graph, $K_{3,3}$

However, $K_{3,3}$ can be drawn on a torus without edge crossings, as depicted in Figure 7. What surface a graph can be drawn on yields a way of measuring the complexity of the graph. The *topological genus* of a graph is, roughly, the least number of handles that need to be added to a sphere in order to permit that graph to be drawn on that surface without any edges of the graph crossing each other (cf. [Harary, 1969], pp. 102, 116). Planar graphs thus have genus 0, while $K_{3,3}$ has genus 1 (since the torus can be thought of as a sphere with one handle added).

Of importance for Carbone's work is the *complete graph* K_n of n vertices, in which every pair of vertices is connected by an edge. One such example, K_5, the complete graph of 5 vertices, is depicted in Figure 8. In general, K_n has $\binom{n}{2} = \frac{n(n-1)}{2}$ edges. Fortunately, unlike genus in general, there is an easy way to calculate the genus of complete graphs. It was shown by Ringel

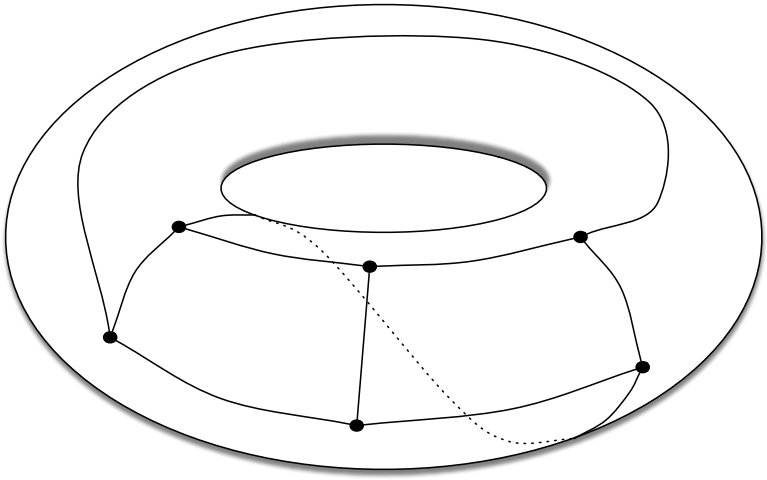


FIGURE 7. $K_{3,3}$ on a torus

and Youngs that the genus of K_n is $\lceil \frac{(n-3)(n-4)}{12} \rceil$, so that the values of K_n for $n = 5$ to 14 are 1, 1, 1, 2, 3, 4, 5, 6, 8, 10 (cf. [Harary, 1969], p. 118).

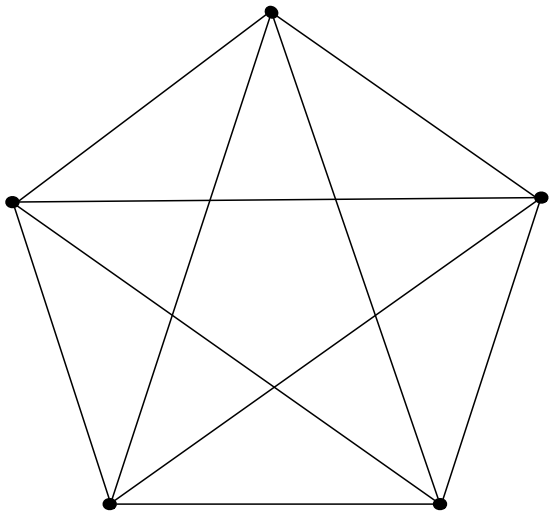


FIGURE 8. The complete graph K_5

2.6. **Carbone's main result.** We can now present Carbone's main result, which is also proven in [Statman, 1974]. (Carbone's proof is purely graph theoretic.) She shows that for any topological genus n, there is a *cut-free proof* with that genus. She shows this by constructing, for each $n \geq 3$, cut-free proofs into whose logical flow graphs she has "embedded" the complete graph K_{2n} of $2n$ vertices. The proof is constructed via an "acyclic optical graph". Carbone has shown that for any such graph there is a formal proof in the sequent calculus plus a rule

for function composition[8] whose logical flow graph has the same topological structure as the acyclic optical graph. A further condition can be placed on the graph to make sure the resulting proof is cut-free ([Carbone, 2009, p. 144]).

3. Evaluating genus as a measure of discovermental complexity

The goal of this section is to assess whether discovermental complexity can be measured by the genus of proof graphs. We will carry this out in three steps. Firstly, we will show that genus is a measure of the structural complexity of *graphs*. Secondly, we will argue that the genus of a logical flow graph is not a measure of the structural complexity of a *proof* but that the genus of Statman's derivation graphs may be. Finally, we will consider whether a measure of structural complexity is a good measure of discovermental complexity.

3.1. Genus as a measure of the structural complexity of graphs.

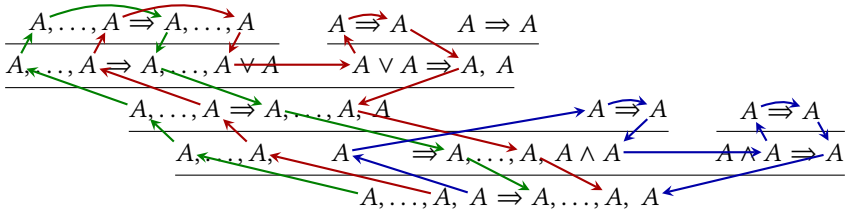
Carbone suggests that genus measures a graph's combinatorial complexity ([Carbone, 2009, p. 139]). Statman similarly argues that genus is a measure of the structural complexity of surfaces and analogously of proof graphs ([Statman, 1974, p. vi]). We can however give more concrete reasons for accepting this view.

First, note genus' connection to more standard complexity measures such as size, which in analogy to length as a measure of proof complexity, might be thought of as a simple measure of graph complexity. As genus increases, so will the minimum size of graphs. This follows from the Euler characteristic of an embedded graph G, which is $\chi(G) = V - E + F$, where V is the number of vertices, E the number of edges, and F the number of faces of the embedding. It follows that the genus $g = \frac{2-(V-E+F)}{2}$ ([Wilson, 2013, p. 39]). Note that there is a relationship between the maximum number of faces and the number of vertices and edges. For example, with four vertices and six edges the most faces one can have is four. And as we only let one edge hold between any two vertices, the number of vertices constrains the number of edges. It follows that given any number V there is a maximum possible E and F and so a maximum

[8] The rule for function composition is as follows:

$$\frac{F(x_1), \ldots, F(x_n) \Rightarrow F(x_1), \ldots, F(x_n)}{F(x_1), \ldots, F(x_n) \Rightarrow F(x_1, \ldots, x_n)}$$

This rule can be removed by iterating the following construction till there are n occurrences of A in each cedent and then applying contraction on the right.



possible genus. Thus a high genus requires a large graph, and so given any reasonable method of producing proof graphs, a large proof.

However, genus is a more subtle measure of graph complexity than size alone. A high genus requires a large graph, but a large graph does not ensure a high genus. Every tree graph is planar, so there are arbitrarily large graphs of genus 0. This demonstrates that genus excludes some very large graphs with simple structures. And when one looks at graphs of higher genus, there appear to be complex relations (edges) between the points. High genus graphs have edges connected to points in such a way that the only way to prevent crossing is to add holes for the edges to pass through. As such, a higher genus seems to capture greater interrelatedness among the points. This seems to correctly capture an important facet of structural complexity.

Statman calls genus a measure of *global* structural complexity. This is appropriate because the genus of a graph cannot be calculated from the genus of its parts. Genus is a property of the structure as a whole. The genus of a graph $G = G_1 \cup G_2$ in which G_1 and G_2 share 3 vertices may be arbitrarily larger than the addition of the genus of G_1 and G_2 ([Archdeacon, 1986]). This tells us that if genus is a measure of complexity it has properties that interestingly distinguish it from measures of length. Because we will discuss structural proof rules later, we will use Statman's term of global structure to refer to the overall structure of the graph or proof.

So genus appears to be a good choice for measuring the global structural complexity of graphs. But, for genus to measure the global structure of the *proof*, the proof graphs must encode the global structure of the proof. This is not the case for logical flow graphs, as we will now show.

3.2. **Genus as a measure of the structural complexity of proofs.**

3.2.1. *Logical flow graphs.* Buss's logical flow graphs aim to "develop a theory of how the influence of a formula spreads through a proof" ([Buss, 1991, p. 85]). While formal proofs are static syntactic objects, they represent the dynamic and temporal process of reasoning. In the natural deduction calculus we can think of each inference rule as one step that might be taken in reasoning to the conclusion. However, things are more complicated in the sequent calculus. If the proof of the sequent $\Delta \Rightarrow \Gamma$ is to represent our reasoning, it must represent reasoning that starts with Δ as assumptions and reasons to something in Γ as the conclusion. A logical flow graph can be thought of as tracing the role played by a single atomic formula in the reasoning represented by a proof in the sequent calculus. Carbone largely agrees with Buss's assessment of logical flow graphs.[9]

As we discussed in the previous section, logical flow graphs have many useful applications. But we will now argue that measuring the global complexity of proofs is not one of them. The following argument relies on two claims. First, Carbone's method of producing proofs of

[9]See [Carbone and Semmes, 2000, p. 484]; [Carbone, 2005, p. 3]; [Carbone and Semmes, 1997, p. 153]; [CAR, 1997]; [Carbone, 1999], [Carbone, 2000], [Carbone, 2001], [Carbone, 2002].

higher genus relies only on structural rules.[10] Second, graphs encoding only the structural rules do not capture the global structure of the proof.

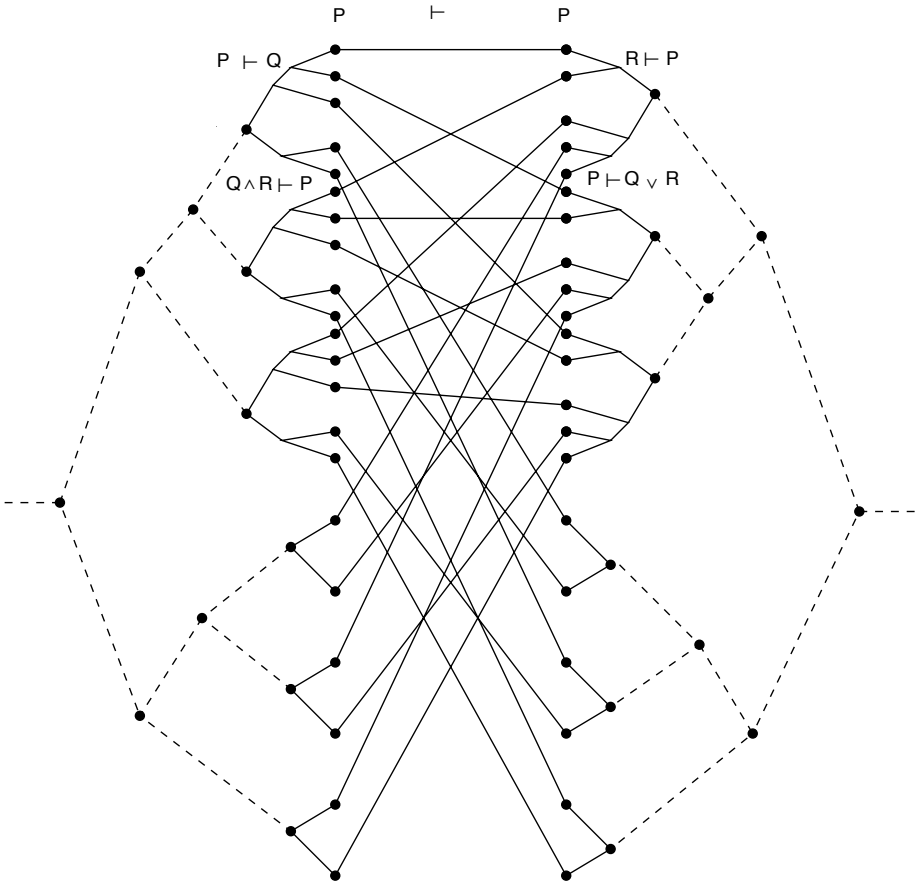
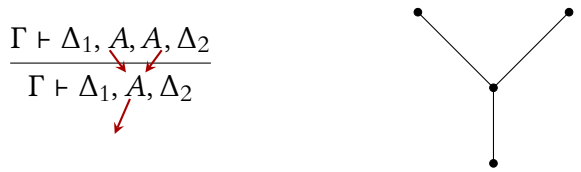


FIGURE 9. A piece of a proof with non-planar genus



(A) Flow graph on proof (B) Flow graph isolated

FIGURE 10. Flow graph for contraction

3.2.2. *Logical flow graphs ignore the structure of the logical rules.* The following observation is at the crux of our argument. All graphs with a genus greater than 0 must have points with three or more edges attached. If this is not the case then the graph is either a line or a cycle, both of

[10]Recall, that structural rules are rules that restructure the information in a sequent and are distinct from questions of the global structure of the proof.

which have genus 0. As one can see by inspection of the definition of the logical flow graph, only contraction (Figure 10) produces a point with 3 edges attached. It follows that all logical flow graphs of proofs that do not contain contraction will have genus 0 as they will be lines or cycles.

By the above observation, it follows that if genus is a measure of the global structural complexity of a proof, then the logical rules on their own cannot produce proofs with anything but the simplest global structure. But for the sake of argument let us allow that contraction is necessary for complex proofs. It will now be argued that the logical connectives still contribute nothing to the complexity of the graph because they can be replaced by structural rules.

Consider left and right $\neg$. These are both one premise rules. Given a proof $\mathfrak{D}$ and a graph G, if we apply one of these rules to the conclusion of $\mathfrak{D}$ to get a new proof $\mathfrak{D}'$ and graph G', then G' is structurally identical to G. So the application of these rules does not affect the genus. For left $\wedge$ and right $\vee$, we introduce a new formula. Say we add a formula with m atomic formulas; then our graph G' will look just like the case of $\neg$ but with m unconnected points. This is the same effect that weakening has on a proof.

The rules of left $\vee$ and right $\wedge$ laid out in Section 2 might look like counterexamples to our point. These are both two-premise rules and as such produce a combined graph composed of the graphs associated with the derivations of the premises. Consider right $\wedge$, let G_1 be the graph associated with the derivation of $\Gamma_1 \Rightarrow A, \Delta_1$ and G_2 be the graph associated with $\Gamma_2 \Rightarrow B, \Delta_2$. Then the graph associated with $\Gamma_1, \Gamma_2 \Rightarrow A \wedge B, \Delta_1, \Delta_2$ will be $G_1 \cup G_2$ where G_1 and G_2 share no vertices. What impact can this have on the genus? None at all, the genus of $G_1 \cup G_2$ will just be the addition of the genus of G_1 and G_2 ([Battle et al., 1962]). However, the application of the rule does allow contraction to be applied later to formula occurrences which could not happen before, if one occurrence is in Γ_1/Δ_1 and another in Γ_2/Δ_2. So derivatively an application of right $\wedge$ introduction or left $\vee$ introduction might contribute to the genus of the proof. It was discussed in section 2 that as well as the divide between logical and structural rules, we could identify the features of logical rules that parallel the structural rules. This is relevant here because it allows us to identify the features of right $\wedge$ introduction and left $\vee$ introduction that parallel the structural rules. We see that Carbone's rendering of these rules parallels the structural rules in that they merge the antecedents and the succedents. This is the same effect that the rule of merger has. Any impact on the genus brought about by these rules is a result of the contingent fact that they also merge the antecedents and the succedents.

What the above considerations point to is that logical flow graphs capture the features of rules that parallel the structural rules without capturing the structure of the logical rules. Do not be confused by the expression 'features that parallel the structural rules'. Recall that what this implies is that these rules are concerned not with building up formulas from subformulas as the logical rules are but rather with where the formulas are in the sequent. We are investigating

whether the global structure of a proof is captured. And the global structure is not confined to the features that parallel the structural rules. This is because the application of a rule like right $\wedge$ has a structural effect on the proof. It binds together two proofs, one for each of the conjuncts. Yet this structure is entirely missing from logical flow graphs.

3.2.3. *Possible replies.* In proving her results Carbone adds one rule to *LK* which does affect genus and isn't a structural rule. Does adding this rule mitigate the argument of the last subsection? The rule in question is function composition:

$$\frac{F(t_0) \Rightarrow F(t_0) \quad \ldots \quad F(t_n) \Rightarrow F(t_n)}{F(t_0), \ldots, F(t_n) \Rightarrow F(f(t_0, \ldots, t_n))}$$

This rule also allows for the generation of nodes of degree at least three and so can affect the genus of a proof. Note that now only structural rules and function composition affect genus. If the complaint was that many complex proofs do not include contraction, there will be many that also do not include function composition. What is more, the use of function composition is merely a convenience, as we can replace instances of function composition with proofs containing only rules in *LK* (see footnote 8). As such, function composition is not a vital feature of this complexity measure.

Not all proof systems have structural rules. There is a system equivalent to *LK* which does not have any structural rules ([Troelstra and Schwichtenberg, 2000, § 3.5]). If we were to move to such a system would we avoid the issue raised here? If we look at the rules in the modified system, we see they all have context sharing. Take the example of right $\wedge$-introduction:

$$\frac{\Gamma \Rightarrow \Delta, A \quad \Gamma \Rightarrow \Delta, B}{\Gamma \Rightarrow \Delta, A \wedge B}$$

Recall that the principle formula in the lower sequent of an inference is the formula where a connective was introduced. Here we see something quite like contraction happening to the non-principal formulas and again contributing nothing to the genus by the principal formula. Still, it might seem like the situation here is better, as the inference rules for some of the logical connectives allow the forming of nodes of degree three. But it is noticeable that this is due to every formula other than the ones we want to affect the complexity, namely, the principal formulas. As such it seems correct to say that, in this system, inference rules have features that parallel the structural and non-structural rules, and it remains the case that the features that parallel the structural rules are those alone that contribute to the genus.

To sum up, the source of the above criticism is the fact that the logical flow graphs ignore the structure introduced by the logical connectives. This is not a criticism of logical flow graphs: as we saw their purpose is to track the movement of a single formula through a proof. The problem is that logical flow graphs simply forget the logical inference rules, and so do not capture the

global structure of proofs. To put it another way, genus should measure the structure of the proof, but logical flow graphs do not capture this structure satisfactorily.

3.2.4. *Global structural complexity of proofs and derivation graphs.* Recall that Statman takes the genus of a proof graph to be a measure of the global structural complexity of the associated proof. He maintains that lowering the global structural complexity of a proof is of practical importance, because 'if this [global structural complexity] is not at a manageable level a proof will not even begin to be understood' ([Statman, 1974, p. v]). But if a measure of global structural complexity is a good measure of discovermental complexity and Statman is correct that his proof graphs capture the global structure of proofs, then he will have produced a measure of discovermental complexity.

Statman's argument that he has captured the global structure of proofs is that derivation graphs account for all the relationships between any two parts of the proof. Statman ([Statman, 1974, p. 2]) argues that by representing proofs as trees, Gentzen did not represent all relations between formulas in the proof. He claims that the relationship between an assumption and the conclusion of the inference that discharges it is missing. Should we think of the discharge of assumptions as part of the global structure of the proof? Two features of the discharge of assumptions suggest it is a structural property of proofs. Firstly, assumption discharge is restricted by the tree structure of the proof. For example, an assumption cannot be discharged by the application of a rule on another branch. Secondly, without this the tree does not tell us which assumptions are open or closed, which we need to know to know what the proof is a proof of.

By adding the relationship between an assumption and where it is cancelled, Statman additionally claims that *all* structural relations between formulas in the proof have been accounted for. This claim is supported by an examination of how schematic proof rules are displayed. When we specify proof rules, we need only specify discharged formulas, premises, and conclusions and the deduction graph connects all these formulas.[11] As such in the natural deduction system, it seems correct that the two relationships that hold between inferences in a proof are premise-to-conclusion and assumption-to-discharge. As a result of this, it can be concluded that Statman's graphs show a representation of the global structure of a proof.

Further, Statman's graphs avoid the difficulty that logical flow graphs have. The genus of the graph is affected by the logical rules. This point is slightly more complicated than simply pointing out that it is not only structural rules that lead to proofs of higher genera, because the natural deduction system that Statman is using does not have structural rules distinct from the logical rules. Rather, in natural deduction, the structural rules are hidden in the inference rules.

[11] One might wonder if the inclusion of quantifiers and the restrictions on variable occurrences might make that case more complicated.

For example, if we have an inference such as:

$$\dfrac{\dfrac{[A] \quad [A]}{A \wedge A}}{A \to (A \wedge A)}$$

where both hypotheses are cancelled by implication introduction, then the equivalent proof in the sequent calculus would be as follows:

$$\dfrac{\dfrac{\dfrac{A \Rightarrow A \quad A \Rightarrow A}{A, A \Rightarrow A \wedge A}}{A \Rightarrow A \wedge A}}{\Rightarrow A \to (A \wedge A)}$$

But now we see that a structural rule is needed to combine the two separate assumptions of A in the sequent calculus. And this is a hidden feature of the natural deduction proof that parallels the structural rules.[12] So, both Statman's method and Carbone's have features that parallel the structural rules and these features impact genus. But the logical structure of the proof has an impact on the genus in Statman's method because we include the proof tree, which tracks how the premises are combined to get the conclusion. This is illustrated in Figure 3 and 4 where the difference in genus between the two graphs would not occur if we were tracking the atomic formulas rather than the formulas themselves. It is crucial for the increased genus in the second proof that formulas are introduced and then eliminated, thereby increasing the connectivity of the graph. In contrast, logical flow graphs do not track how premises are combined, because they only track atomic formulas. The combining of two formulas into a larger one does not affect logical flow graphs. A further difference between the derivation graphs and logical flow graphs is that contraction is not required to generate derivation graphs of genus greater than 0, as it can be shown (as in Figure 11) that there are proofs of genus greater than 0 which do not use contraction (discharging multiple assumptions at once). If the above argument is accepted, then we have shown that genus is a measure of the global structural complexity of graphs and that Statman's derivation graphs represent the global structure of proofs.

3.3. **Global structural complexity of proofs as discovermental complexity.** The last point to be discussed is why a measure of global structural complexity would be a measure of discovermental complexity. Both Carbone and Statman agree that a high global structural complexity means that the proof has many highly interconnected ideas of which the discoverer or reader must keep track. This seems correct. The higher genus represents a structure that

[12]This is a common observation. [Troelstra and Schwichtenberg, 2000] ([Troelstra and Schwichtenberg, 2000, p. 68]) note that natural deduction is closed under contraction and weakening. [von, , p. 46] tells us that the discharge of multiple assumptions is parallel to contraction and vacuous discharge is parallel to weakening. Restall ([Restall, 2014, p. 1157]) states that Gentzen's natural deduction system has contraction unless the discharge of multiple premises is not permitted.

$$\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{1_a \quad 1_b}{1_a \wedge 1_b(1)} \quad 4_a}{1 \wedge 4_a(4)} \quad 2_a}{4 \wedge 2_a(2)}}{1_a \to 2(5)}}{4_a \to 5(3)}}{2_a \to 3(6_a)}}{1_b \to 6_a(6)}$$

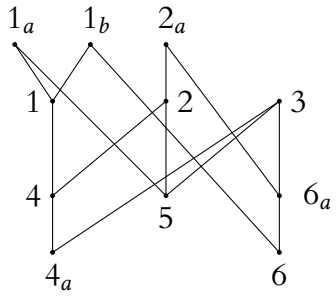


FIGURE 11. A proof without contraction and its graph which embeds $K_{3,3}$

not only requires more resources due to being larger or having more connections, but also a structure that cannot be "lain flat" and surveyed as such, without attending to a more tightly interwoven network of edges. The prover must keep this inferential network straight, without confusing what inference leads to what inferred formula. As the genus increases, this becomes harder. It becomes harder to represent all the pieces of the proof together, and this should correspond to an increased difficulty in assembling all the pieces of the proof together.

Interestingly, both Carbone ([Carbone, 2009, p. 139]) and Statman ([Statman, 1974, p. v]) have concerns about the complexity of proofs with cut and the inclusion of lemmas. We will focus on derivation graphs here, but our discussion should hold for any measure which is closely tied to the global structure of the proof. While it will be discussed in much more detail in section 4, note that an open assumption is a formula one assumes for the purposes of proof. If one were to then prove this assumption, it now behaves like a lemma. Because of the relevance of this point to the discussion on purity it is worth pausing to consider how genus behaves on proofs with open and closed assumptions.

Consider the following three proofs:

$$\varphi \quad \begin{matrix}\varphi \\ \mathcal{D} \\ \chi\end{matrix} \qquad \psi \quad \begin{matrix}\mathcal{D}' \\ \varphi\end{matrix} \qquad \begin{matrix}\mathcal{D}' \\ \varphi \\ \mathcal{D} \\ \chi\end{matrix} \quad \begin{matrix}\mathcal{D}'' \\ \psi\end{matrix} \qquad \dfrac{[\varphi] \quad [\varphi] \quad [\psi]}{\varphi \to \psi \to \chi}\ \begin{matrix}\mathcal{D} \\ \chi\end{matrix}$$

If the genus of a proof like $\mathcal{D}$ is g and it has $\sum_{i\leq n} m_i$ assumptions consisting of m_i copies of each φ_i, then the genus of the proof resulting in case 2 will be at least g and not more than $g + \sum_{i\leq n} m_i g_i$ where g_i is the genus of the proof of each φ_i ([Decker et al., 1981]). Whereas, the genus of the third case will be at least g and not more than $g + \sum_{i\leq n} m_i$.

The moral of this is that with the exception of planar graphs, the upper-bound on genus for proofs with lemmas (case 2) is greater than that for proofs with assumptions discharged (case 3).

This suggests that the measure could reveal interesting relationships between conditional proof and lemmas.

However, it remains a concern that only proofs which include discharged assumptions can have genus greater than zero. This is because proofs with no assumptions discharged are trees and all trees have genus zero ([Chen, 2013, p. 746]). It may be that there is a story to tell here about proofs without discharge of assumptions being trees and so structurally far simpler than some proofs in which assumptions are discharged. However, one might worry that assumptions can be quite complicated and the proofs that follow from them may not be easy to find. Similarly, it is not clear why upon discharging assumptions the complexity would suddenly jump up in some cases.

One option is simply to admit that the complexity of the proof tree should impact proof complexity.[13] Examples of such a measure include: number of branches, height, and width. The measure of complexity could then be the genus of the proof graph plus the complexity measure on the proof tree. As measures on trees are likely to be measures of size such an addition would suit those who think that size is not merely necessary for higher complexity but also sufficient.[14]

One might further worry that global structural simplicity does not contribute to ease of discovery but rather to how easy a proof is to understand.[15] As was emphasised in the introduction, mathematicians prize proofs that are (pretheoretically) simple and it is not uncommon that a reproof of a theorem is considered the simpler proof. For example, it is Henkin's construction, rather than Gödel's original proof, that is usually used in the teaching of the completeness of predicate logic because of the perceived pedagogical value of the simplicity of the construction. But for this intuition to be made into an objection to our view, we would need to be able to distinguish not just discovermental complexity from verificational complexity but a third measure of understanding distinct from the two. However, as discussed above measures of simplicity of understanding are standardly measures of how difficult a proof is to verify correct.

4. Genus and purity

In the introduction to [Carbone, 2009], Carbone observes that the traditional measure of proof complexity, length of proof, does not account adequately for the differences between cut-free proofs and proofs with cuts. She notes that cut-free proofs are usually longer than proofs with cuts, but geometrically simpler. We have seen how she purports to measure the geometric simplicity of a proof, via the topological genus of the proof's logical flow graph. We have compared this geometric measure with Statman's, which measures the genus of a different combinatorial structure within proofs. Both Carbone and Statman obtain results comparing

[13] We would like to thank an anonymous reviewer both for pushing us to say more on this point and for making this suggestion.

[14] Though there are reasons to think size is not sufficient. For example, it is plausible that the obvious proof of $\bigwedge_n p_n$ has complexity 0 for all n despite the linear growth of proof size.

[15] We want to thank an anonymous reviewer for pressing this point and the reply.

the geometric simplicity of cut-free proofs and of proofs with cuts. In particular, they show that for any topological genus *n*, there is a *cut-free proof* with that genus.

Since Gentzen we have recognized that cuts are comparable to lemmas in informal proof. Just as a lemma may draw on resources that are not used elsewhere in the proof, in a cut inference the cut formula occurs in the upper sequent but not in the lower sequent and hence is not a subformula of the conclusion. To infer $\Gamma \Rightarrow \Delta$ (say, concerning circles and lines), a cut may invoke formulas in Γ and Δ as well as other formulas (say, concerning right angles, also) that are not subformulas of $\Gamma \Rightarrow \Delta$. By contrast, in a cut-free proof every formula is a subformula of the conclusion. This points to Gentzen's observation that all of the formulas occurring in cut-free proofs are subformulas of the conclusion. Gentzen described the import of this "subformula property" as follows:

> The final result is, as it were, gradually built up from its constituent elements. The proof represented by the derivation is not roundabout in that it contains only concepts which recur in the final result.... No concepts enter into the proof other than those contained in its final result, and their use was therefore essential to the achievement of that result.[16]

Similarly, Takeuti observed that the subformula property shows that "any theorem in the predicate calculus can be proved without detours, so to speak."[17]

In saying that cut-free proofs are "not roundabout" and avoid "detours", Gentzen and Takeuti suggest viewing cut-free proofs as "pure proofs", that is, proofs realizing the ideal of purity of methods so important to the inventor of proof theory, David Hilbert. Roughly speaking, a proof is pure if it draws only on what is "close" or "intrinsic" to what is being proved. As Hilbert put it, the aim of the search for purity is "to prove theorems if possible using means that are suggested by the content of the theorem" (cf. [Hilbert, 2004], pp. 315–6), rather than means that are extraneous, distant, remote, alien, or foreign to it. Purity as an ideal of proof goes back to Aristotle and remains today important to many mathematicians, even if impurity also is held as an ideal of proof by many mathematicians as well. A fuller analysis of purity and its epistemic value can be found in [Detlefsen and Arana, 2011]; in this paper it suffices to recall the importance of purity and impurity in mathematical practice, as we focus on showing how Statman and Carbone's work bears on purity.

In short, Statman and Carbone's work bears on the question of whether impure proofs, understood as proofs with cuts, are simpler than pure proofs, understood as cut-free proofs. In Section 2 of [Arana, 2017], this question was discussed in a historical context. It was observed that Newton, for instance, judged the use of algebra in proving geometric theorems to be an

[16]This combines two passages from [Gentzen, 1935], p. 88, 69.
[17]Cf. [Takeuti, 1987], p. 21–2, and again on p. 29.

impurity.[18] In analysis as well, a distinction was drawn between pure and impure proofs of propositions of real analysis on the basis of their use of complex numbers. Famously, Jacques Hadamard remarked that "the shortest and best way between two truths of the real domain often passes through the imaginary one" (cf. [Hadamard, 1945], p. 123).

This way of thinking remains widespread today among mathematicians. A notable recent advocate of this view is Carlo Cellucci, who has claimed that "the use of 'impure' methods leads to a marked improvement in efficiency" (cf. [Cellucci, 1985], p. 173). Using the parallel between cuts and lemmas remarked upon above, he notes that lemmas are always redundant in practice, since every use of a lemma in a proof can be replaced with a proof of that lemma. Nevertheless, he remarks, "in mathematical practice we feel better off if we manage with such redundancies than without them" (*Ibid.*, p. 174). To explain this, he suggests that "this circumstance may be accounted for by the fact that redundancies generally lead to a significant gain in efficiency." The thought seems to be that by proving a lemma just once in the course of proving a theorem, we can draw on that lemma repeatedly, and as a result we can compress the proof relative to a cut-free proof of that theorem. Cellucci recalls Statman's result that in propositional sequent calculus the length of cut-free proofs may be significantly larger than that of proofs without cut. More precisely, as mentioned in the introduction, Statman showed that there are sequents whose cut-free proofs are exponentially longer than their proofs with cut.[19] Cellucci takes Statman's result to support his contention that impure proofs yield a gain in simplicity.

In reply to Cellucci, we firstly recall the findings of [Arana, 2017]. That work investigated conservative extensions of PRA by elements that yield, it is argued, impure proofs for theorems of PRA. These theories, Π^1_2-axiomatizable extensions of RCA_0, add sets and principles governing sets to the purely arithmetical theory of PRA: RCA_0, WKL_0 and WKL^+_0, familiar from reverse mathematics (cf. [Simpson, 2009]). Proofs in these theories of purely arithmetic theories, making use of sets, are thus arguably impure. The article in question then compared the simplicity of proofs of theorems of PRA, measured by proof length, with proofs of those same theorems in the set-theoretic extensions. No general pattern of simplicity in moving from pure to impure proof was found; on the contrary, the addition of set-theoretic resources in the theories RCA_0, WKL_0 and WKL^+_0 yield only polynomial speed-up over PRA. Following the

[18] "Equations are Expressions of Arithmetical Computation, and properly have no Place in Geometry, except as far as Quantities truly Geometrical (that is, Lines, Surfaces, Solids, and Propositions) may be said to be some equal to others. Multiplications, Divisions, and such sort of Computations, are newly received into Geometry, and that unwarily, and contrary to the first Design of this Science.... Therefore these two Sciences ought not to be confounded. The Antients did so industriously distinguish them from one another, that they never introduced Arithmetical Terms into Geometry. And the moderns, by confounding both, have lost the Simplicity in which all the Elegancy of Geometry consists." Cf. [Newton, 1720], p. 119–20.

[19] Cf. [Statman, 1978], and [Buss, 1988], pp. 2-1 – 2-5, for an excellent presentation of these results.

tradition in computational complexity theory (cf. [Dean, 2016]), super-exponential speed-ups are considered to be significant gains in simplicity, while polynomial speed-ups are not.

Secondly, Cellucci's comment on the efficiency of impure methods refers to verificational rather than discovermental complexity. Statman's result in [Statman, 1978] concerns proof length, which we argued earlier measures the complexity of verifying that a proposition is a theorem, rather than the complexity of discovering a proof of that theorem. As we observed earlier, though, advocates of impurity on simplicity grounds seem to be thinking as much of its superior discovermental simplicity as its verificational simplicity. Consider again the passage quoted earlier from d'Alembert, in fuller context:

> We can say of the ancient geometrical works, that almost none of them have the ease that algebra gives in reducing their demonstrations to a few lines of calculation.... [I]f anyone would have solely the method of the ancients, it does not appear that, even with the greatest genius, one could make in geometry such great discoveries, or at least in as great a number, as one can with the help of analysis. (Cf. [Diderot and d'Alembert, 1751], vol. 1, p. 551)

Analytic methods provide for a significant shortening of proof, d'Alembert thought, and as a result they dramatically improve our ability to discover new results compared with purely synthetic methods. That is, he seems to have thought that the discovermental complexity of theorems of geometry is lower when permitting analytic methods than when permitting only synthetic methods, so that impurity renders proofs simpler to find.

Like Cellucci, Carbone wants to explain why in practice mathematicians use impure methods (as indicated by cuts/lemmas). Unlike Cellucci, she is particularly interested in "*how difficult* it is to prove" a given proposition (cf. [Carbone, 2009], p. 139). She contends that while impure proofs (proofs with cuts) may be generally *shorter* than pure proofs (cut-free proofs), proofs with cuts seem to be more difficult to *discover*, because precisely which cut formulas (lemmas) are good candidates to be used is typically not obvious. By contrast, in searching for a cut-free proof one may consider only subformulas of the conclusion. This would suggest that the gain in simplicity afforded by the relative shortening of length of proof via impurity is counterbalanced by the relative gain in difficulty of discovering impure proofs. Carbone echoes Cellucci in noting that in practice we search for impure proofs anyway. She seizes upon Statman's genus measure of proof complexity, rather than proof length, in order to explain this preference, which may seem irrational when measuring proof complexity by length.

Carbone thus poses again the question of whether impure proofs are simpler than pure proofs, this time measuring simplicity by proof genus. Her Theorem 3 (cf. [Carbone, 2009], p. 145) seems to be her response. As we discussed above, she shows that for any genus n, there is a cut-free proof with that genus. In the terms of this section, there are pure (cut-free) proofs of arbitrarily high genus. By contrast, it is presently unknown whether this is true for impure

proofs (that's to say, for proofs with cut). Thus her main result shows an asymmetry between purity and impurity. Discovering a cut-free proof, i.e. pure proof, *may* require us to find a proof of high genus complexity. But finding a proof with cuts, i.e. an impure proof, *may not* require us to find such a complex proof. In this precise sense, it is more difficult to find a pure proof than an impure proof. The complexity of a proof may be measured by the genus of its logical flow graph, and Carbone's Theorem 3 provides evidence that pure proofs are generally more complex, in this sense, than impure proofs.[20]

This evidence should be taken with caution, however. We have already objected to the claim that discovermental complexity can be measured by Carbone's genus measure. We did so on the grounds that logical flow graphs are inappropriate for measuring discovermental complexity. These objections also carry weight against the application of Theorem 3 to the discovermental complexity of pure and impure proof.[21] Here we will add two further objections toward this applicability. Firstly, as noted above there is at present no analogue of Theorem 3 for proofs with cut. It may be that one can embed graphs of arbitrarily high genus inside proofs with cut as well. If so, then the alleged gain in simplicity in moving from purity to impurity would vaporize. Secondly, the best way to answer whether impure proofs are generally simpler than pure proofs would be to compare the complexity of pure and impure proofs of a single proposition. That's to say, one should ask for a given proposition φ whether cut-free proofs of φ are systematically less genus complex than proofs with cut of φ. Theorem 3 does not answer this question. It says that for a given genus n, there is a cut-free proof with that genus. It does *not* say that for a given genus n and a given proposition φ, there is a cut-free proof of φ with genus n. That's because Carbone takes proofs as combinatorial objects in their own right, and does not distinguish any particular node of that object as a conclusion. Thus Theorem 3 gives no information on how proofs of a single given proposition vary.

We turn briefly to Statman's measure of genus complexity, which also yields a theorem like Carbone's Theorem 3 (namely, Proposition 3 of Chapter 1, §4, [Statman, 1974], p. 27). While there may be reasons to think that Statman's genus measure is better suited for measuring discovermental complexity, the other objections just assayed apply to Statman's measure as well.

We thus conclude that there is not sufficient evidence for the claim of a general pattern of genus simplicity of impure over pure proof. This coincides with the conclusion for simplicity measured by proof length. The claim that impurity affords gains in discovermental simplicity over purity, though observed by many mathematicians over centuries, is not supported by

[20] In so doing, she gives evidence in favor of Thesis 2 from [Arana, 2017]: "Impure proofs are generally simpler to discover than pure proofs of the same statement."

[21] It should be noted that any proof without cut can be trivially extended to a proof with cut. If the final sequent of a proof is $\varphi, \Gamma \vdash \Delta$, this can be done by adding a cut on $\varphi \vdash \varphi$ to the last line. The question here concerns proofs with non-trivial uses of cut.

the proof-theoretic methods currently available. It awaits further refinement of complexity measures, for both verificational and discovermental complexity.

5. Conclusion

Finally, we would like to consider an objection that can be raised at any attempt to draw conclusions about mathematical practice using proof theory. Briefly, the objection goes, the proof formalisations that are the concern of proof theory are of distant relevance, at best, to what mathematicians actually do when they give proofs. On the contrary, this line goes, what mathematicians write should give the "inner logic" of the proof, but not all its details. The latter are important for the validity of the proof, but not for what Michael Harris calls "the purpose of a proof", which is "to illuminate a concept rather than merely confirm a theorem" ([Gowers, 2008], p. 978). A consequence of this objection is that the formal measures of discovermental complexity studied in this paper are also only of distant relevance, at best, to proof discovery in actual mathematics.

The point is sometimes made even more strongly. John Baldwin calls *Tait's maxim* the observation that "the notion of formal proof was invented to study the *existence* of proofs, not methods of proof" ([Baldwin, 2018], p. 281). He adds John Burgess' observation that "For formal provability to be a good model of informal provability it is not necessary that formal proof should be a good model of informal proof" ([Burgess, 2010]). Yehuda Rav claims along similar lines that

> The study of proofs…and the proof-theoretical study of derivations and related problems belong respectively to different methodologies. We render therefore unto proof theory the things which are proof theory's, and let philosophy of mathematics deal with the nature and function of conceptual proofs as they occur in actual mathematical practice. (Cf. [Rav, 1999], p. 12.)

Fenner Tanswell has pointed out that the existence of many formal proofs allegedly formalizing any given informal proof makes problematic the relation between formal and informal proofs (cf. [Tanswell, 2015]). Lastly, Brendan Larvor has documented other attempts in this direction (cf. [Larvor, 2019], p. 2716n1,2), for instance by Bernd Buldt, Benedikt Löwe and Thomas Müller, who write that "the completion of enthymematic, semi-formal proofs to formal derivations almost never happens and hardly plays any rôle in the justification that mathematicians give for their theorems"; on the contrary, they ask whether more informal notions of proof, like those given in blackboard sketches, should "replace the unrealistic notion of formal derivation in our epistemology of mathematics" ([Buldt et al., 2008], p. 311). On these grounds, the observation goes, one can conclude that the study of formal proofs (as opposed to formal provability) is *irrelevant* to the study of proofs as made by mathematicians in their ordinary work. If this is

correct, the measures of proof complexity studied in this paper would be irrelevant to actual mathematical proof, bearing only on the simulacrum studied by proof theorists.

An instance of this alleged irrelevance concerns proof length. It is well known that proof length depends on choices of means of expression. For instance, Mathias has shown that the term expressing 1 in Bourbaki's 1954 set theory has approximately 10^{12} characters; but in the fourth edition (using Kuratowski's definition of ordered pairs rather than taking them as primitive) it grows to 10^{54} characters (cf. [Mathias, 2002], also [Potter, 2004], pp. 234–236). Simpson has stressed as well that the formalisations of ordinary proofs in subsystems of second-order arithmetic, as studied in reverse mathematics, are "sometimes much more complicated than the standard proof" (cf. [Simpson, 1988], p. 361). Avigad has concluded that while "length has something to do with explaining how infinitary methods can make a proof simpler and more comprehensible", the philosopher interested in the complexity of proof should focus instead on "the perspicuity and naturality of the notions involved, and using the number of symbols in an uninterpreted derivation as the sole measure of complexity is unlikely to provide useful insight" (cf. [Avigad, 2003], p. 276n18).

The thrust of these lines of reasoning is to call into question the connection of the formal results studied in this paper with the reductions of discovermental complexity studied by Descartes, Leibniz and d'Alembert and discussed earlier. We could put the point bluntly: no one has ever said, "proving things in primitive recursive arithmetic is hard, but is made so much easier by working in $I\Sigma_1$." But the claims about discovermental complexity from mathematical practice that we have seen *do* make claims like this.

We take the point of this line of reasoning to be that the epistemic features of proofs most important to mathematical practice are not captured by proof theory as it has been done until today. Instead, the objection emphasizes, proof theory studies the micro structure of proofs, at the level of individual inferences at logically fine levels of granularity, even at the level of propositional logic as studied by Carbone and Statman. The thrust of the line of argument considered so far in this section has been that this level of granularity is irrelevant to the study of the complexity of proofs as actually discovered and used by mathematicians.

In this section we want to defend the salience of the proof theory employed by Carbone and Statman to actual mathematical practice. We will pursue two lines of defense. The first will emphasize the importance of propositional logic for recent advances in a core area of contemporary mathematics, arithmetic combinatorics, while the second will argue for the continued need to attend to the complexity of "low-level", logical details that the objections above hold to be irrelevant to actual mathematical practice.

Our first line of defense turns, then, to arithmetic combinatorics. This is an active area of contemporary mathematics, indicated for instance by the Fields Medals earned by practitioners in the area (Roth, Bourgain, Gowers, Tao; cf. [Arana, 2015], Section 1). An example from

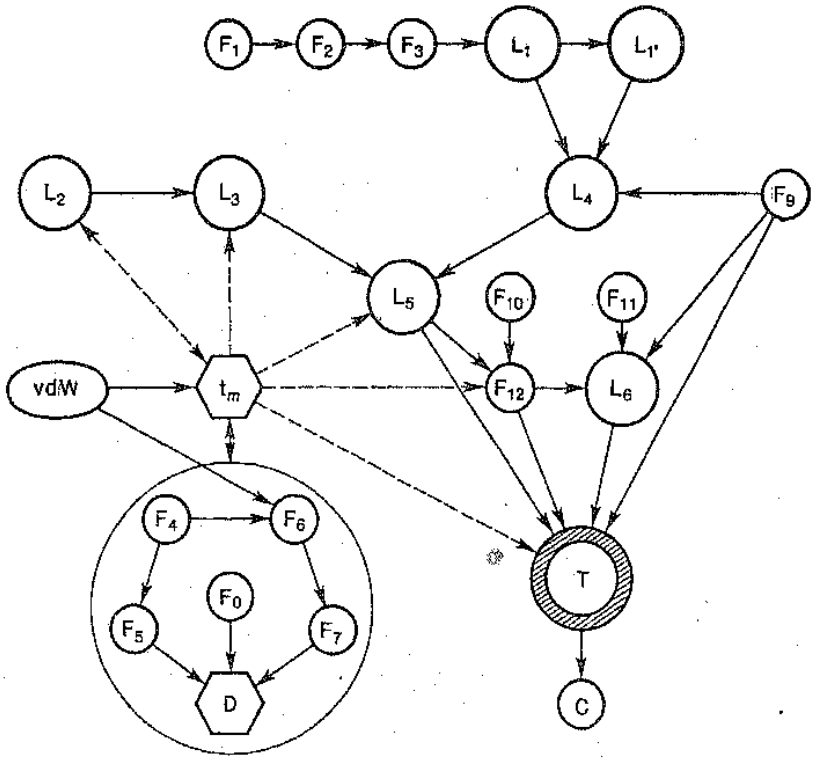
arithmetic combinatorics is the *Boolean Pythagorean Triples problem*, a problem for which Ronald Graham, in the style of Paul Erdős, offered a cash prize in the 1980s to its eventual solver (cf. [Lamb, 2016], p. 17). A Pythagorean triple is a collection of three natural numbers a, b, c, such that $a^2 + b^2 = c^2$. The question is whether the natural numbers can always be partitioned into two parts such that one of those parts contains a Pythagorean triple. For instance, we can partition the natural numbers into odd and even numbers. While the odd part contains no Pythagorean triple, since an odd number squared is odd and the sum of two odd numbers is even, the even part does, since, for example, $6^2 + 8^2 = 10^2$ (cf. [Heule and Kullmann, 2017], p. 72). The problem can be thought of instead in terms of colorings: the problem asks if each natural number can be colored one of two colors, say, red or blue, so that every Pythagorean triple is multicolored (e.g. if 3 and 4 were red, 5 would have to be blue). In this case neither partition would contain a Pythagorean triple.

Marijn Heule, Oliver Kullmann and Victor Marek answered the Boolean Pythagorean Triples problem by showing that for the set of natural numbers up to 7,824, there exist partitions into two parts avoiding Pythagorean triples, but for sets of natural numbers surpassing this threshold, at least one part of such partitions must contain a Pythagorean triple (cf. [Heule et al., 2016]). They earned Graham's prize money, moreover, by a novel application of SAT solvers, software that implements an algorithm to determine whether a given formula in propositional logic is satisfiable. They did so by expressing the Boolean Pythagorean Triples problem as a formula of propositional logic, and then using a SAT solver to determine the threshold of 7,825.

There is much that is philosophically tantalizing about this work: for instance, does the massive search space traversed by the SAT solver and the according length of the generated proof (taking 200 terabytes of storage) give reason to doubt what the epistemic value of such a proof is for agents of cognitive type like us (cf. [Detlefsen and Luker, 1980] and [Heule and Kullmann, 2017], pp. 77–8)? Here we want only to underline that this work shows the relevance of propositional logic to core contemporary mathematics, in that a long-standing open problem was solved by way of expressing it in propositional logic. We can conclude that the sorts of propositional formulas to which Carbone and Statman's complexity measures apply can indeed have non-trivial mathematical content themselves.

Our second line of defense against the irrelevance of the complexity of "low-level", logical details to actual mathematical practice, turns to the structure of proofs themselves. Proofs at any level of granularity have logical structure, even if (as for instance Poincaré and Brouwer argue) the quality of their evidence cannot be *reduced* to logical evidence, that is, the sort of evidence produced by attention to logical inference. Here we may draw on what Van Bendegem calls a "proof-outline": "A proof-outline is best understood as a summary of a proof: it lists the essential steps without filling in the details. It is perfectly comparable to the high-level structure of a computer program" (cf. [van Bendegem, 1988], p. 252). Figure 12 depicts such an outline from

[Szemerédi, 1975] (p. 202), which Szemerédi calls a "flow chart" of his proof of his eponymous theorem, by means of a planar graph.



The diagram represents an approximate flow chart for the accompanying proof of Szemerédi's theorem. The various symbols have the following meanings: $F_k \equiv$ Fact k, $L_k \equiv$ Lemma k, T $\equiv$ Theorem, C $\equiv$ Corollary, D $\equiv$ Definitions of B, S, P, a, β, etc., $t_m \equiv$ Definition of t_m, vdW $\equiv$ van der Waerden's theorem, $F_0 \equiv$ "If $f: \boldsymbol{R}^+ \to \boldsymbol{R}^+$ is subadditive then $\lim_{n\to\infty} \frac{f(n)}{n}$ exists".

FIGURE 12. Szemerédi's proof-outline

Such outlines have logical structure, even if that structure "differs substantially from a derivation" with all logical details presented, as Rav puts it (cf. [Rav, 1999], p. 29). It is that structure to which we can apply our graph-theoretic complexity measures.

Such structure is pervasive in informal proofs and there are standard phrases used to signal the structure to the reader. When an author writes "the proof follows by induction" or "...thus *A* or *B*. If *A*, then...If *B*, then...", they are signaling informally the inference rules to which they are appealing. These moves in informal proofs are not *analogous* to the formal inference rules, but rather are examples of them, not in the sequence calculus or in natural deduction, but in mathematical natural language.

It is unnecessary that this logical structure be fully analyzed; we are not taking the view that proof theory only applies to fully-articulated proofs, nor even that such full articulation is

possible. It may be, as Rav says, that proofs are infinitary objects in the sense that they can be analyzed further and further (cf. [Rav, 1999], p. 15; see also [Kreisel, 1970], p. 511n22); but then the graphs of these further analyzed proofs will only be yet more complex. But so long as a proof, or its outline, has logical structure, our discovermental complexity metrics can be applied to it. They are thus not irrelevant to actual mathematical proving.

References

[von,] A sequent calculus isomorphic to gentzen's natural deduction. 4.

[CAR, 1997] (1997). Interpolants, cut elimination and flow graphs for the propositional calculus. *Annals of Pure and Applied Logic*, 83(3):249 – 299.

[Arana, 2015] Arana, A. (2015). On the depth of Szemerédi's Theorem. *Philosophia Mathematica*, 23(2):163–176.

[Arana, 2016] Arana, A. (2016). Imagination in mathematics. In Kind, A., editor, *The Routledge Handbook of Philosophy of Imagination*, chapter 34, pages 463–477. Routledge, London and New York.

[Arana, 2017] Arana, A. (2017). *On the Alleged Simplicity of Impure Proof*, pages 205–226. Springer International Publishing, Cham.

[Archdeacon, 1986] Archdeacon, D. (1986). The orientable genus is nonadditive. *Journal of Graph Theory*, 10(3):385–401.

[Avigad, 2003] Avigad, J. (2003). Number theory and elementary arithmetic. *Philosophia Mathematica*, 11:257–284.

[Baldwin, 2018] Baldwin, J. (2018). *Model Theory and the Philosophy of Mathematical Practice*. Cambridge University Press, Cambridge.

[Battle et al., 1962] Battle, J., Harary, F., and Kodama, Y. (1962). Additivity of the genus of a graph. *Bull. Am. Math. Soc.*, 68(6):565–568.

[Buldt et al., 2008] Buldt, B., Löwe, B., and Müller, T. (2008). Towards a new epistemology of mathematics. *Erkenntnis*, 68(3):309–329.

[Burgess, 2010] Burgess, J. P. (2010). Putting structuralism in its place. Preprint.

[Buss, 1988] Buss, S. R. (1988). Weak formal systems and connections to computational complexity. Student-written notes of a Berkeley Math 271 topics course.

[Buss, 1991] Buss, S. R. (1991). The undecidability of k-provability. *Annals of Pure and Applied Logic*, 53(1):75–102.

[Carbone, 1999] Carbone, A. (1999). Turning cycles into spirals. *Ann. Pure Appl. Logic*, 96(1-3):57–73. Festschrift on the Occasion of Professor Rohit Parikh's 60th Birthday.

[Carbone, 2000] Carbone, A. (2000). Cycling in proofs and feasibility. *Trans. Amer. Math. Soc.*, 352(5):2049–2075.

[Carbone, 2001] Carbone, A. (2001). Asymptotic cyclic expansion and bridge groups of formal proofs. *J. Algebra*, 242(1):109–145.

[Carbone, 2002] Carbone, A. (2002). The cost of a cycle is a square. *J. Symbolic Logic*, 67(1):35–60.

[Carbone, 2005] Carbone, A. (2005). Pathways of deduction. In *Geometries of Nature, Living Systems and Human Cognition*, pages 383–400. World Sci. Publ., Hackensack, NJ.

[Carbone, 2009] Carbone, A. (2009). Logical structures and genus of proofs. *Annals of Pure and Applied Logic*, 161(2):139–149.

[Carbone and Semmes, 1997] Carbone, A. and Semmes, S. (1997). Making proofs without modus ponens: an introduction to the combinatorics and complexity of cut elimination. *Bull. Amer. Math. Soc. (N.S.)*, 34(2):131–159.

[Carbone and Semmes, 2000] Carbone, A. and Semmes, S. (2000). *A Graphic Apology for Symmetry and Implicitness*. Oxford Mathematical Monographs. Oxford University Press, Oxford. Oxford Science Publications.

[Cellucci, 1985] Cellucci, C. (1985). Proof theory and complexity. *Synthese*, 62:173–189.

[Chen, 2013] Chen, J. (2013). Minimum genus and maximum genus. In Gross, J. L., Yellen, J., and Zhang, P., editors, *Handbook of Graph Theory, Second Edition*, pages 745–759.

[Dean, 2016] Dean, W. (2016). Computational complexity theory. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2016 edition.

[Decker et al., 1981] Decker, R. W., Glover, H. H., and Huneke, J. P. (1981). The genus of the 2-amalgamations of graphs. *J. Graph Theory*, 5(1):95–102.

[Descartes, 1637] Descartes, R. (1637). La géométrie. In *Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences*, pages 297–413. Jan Maire, Leiden.

[Detlefsen, 1990] Detlefsen, M. (1990). On an alleged refutation of Hilbert's program using Gödel's first incompleteness theorem. *Journal of Philosophical Logic*, 19(4):343–377.

[Detlefsen, 1996] Detlefsen, M. (1996). Philosophy of mathematics in the twentieth century. In *Philosophy of Science, Logic, and Mathematics*, volume 9 of *Routledge History of Philosophy*, pages 50–123. Routledge, London and New York. Edited by Stuart G. Shanker.

[Detlefsen and Arana, 2011] Detlefsen, M. and Arana, A. (2011). Purity of methods. *Philosophers' Imprint*, 11(2):1–20.

[Detlefsen and Luker, 1980] Detlefsen, M. and Luker, M. (1980). The four-color theorem and mathematical proof. *The Journal of Philosophy*, 77(12):803–820.

[Diderot and d'Alembert, 1751] Diderot, D. and d'Alembert, J. L. R. (1751). *Encyclopédie ou Dictionnaire raisonné des sciences, des arts et des métiers*, volume 1. Briasson, David, Le Breton, and Durand, Paris.

[Gentzen, 1935] Gentzen, G. (1934-1935). Untersuchungen über das logische schliessen. *Mathematische Zeitschrift*, 39:405–431. Translated as "Investigations into logical deduction" in *The collected papers of Gerhard Gentzen*, M.E. Szabo (ed.), North-Holland, 1969.

[Gowers, 2008] Gowers, T., editor (2008). *The Princeton companion to mathematics*. Princeton University Press, Princeton, NJ.

[Hadamard, 1945] Hadamard, J. (1945). *The Psychology of Invention in the Mathematical Field*. Princeton University Press, Princeton.

[Harary, 1969] Harary, F. (1969). *Graph theory*. Addison-Wesley Publishing Co., Reading, Mass.-Menlo Park, Calif.-London.

[Heule and Kullmann, 2017] Heule, M. J. H. and Kullmann, O. (2017). The science of brute force. *Communications of the ACM*, 60(8):70–79.

[Heule et al., 2016] Heule, M. J. H., Kullmann, O., and Marek, V. W. (2016). Solving and verifying the boolean pythagorean triples problem via cube-and-conquer. In Creignou, N. and Le Berre, D., editors, *Theory and Applications of Satisfiability Testing – SAT 2016*, pages 228–245, Cham. Springer.

[Hilbert, 1927] Hilbert, D. (1927). The foundations of mathematics. In van Heijenoort, J., editor, *From Frege to Gödel: A Source Book in Mathematical Logic, 1879–1931*, pages 464–479. Harvard University Press.

[Hilbert, 2004] Hilbert, D. (2004). *David Hilbert's Lectures on the Foundations of Geometry, 1891–1902*. Springer-Verlag, Berlin. Edited by Michael Hallett and Ulrich Majer.

[Huygens, 1950] Huygens, C. (1888–1950). *Oeuvres complètes de Christiaan Huygens*. Martinus Nijhoff, The Hague. 22 volumes.

[Kreisel, 1970] Kreisel, G. (1970). Principles of proof and ordinals implicit in given concepts. In A. Kino, J. M. and Vesley, R., editors, *Intuitionism and Proof Theory*, pages 489–516. North-Holland, Amsterdam.

[Lamb, 2016] Lamb, E. (2016). Maths proof smashes size record. *Nature*, 534:17–18.

[Larvor, 2019] Larvor, B. (2019). From euclidean geometry to knots and nets. *Synthese*, 196(7):2715–2736.

[Maronne, 2010] Maronne, S. (2010). Pascal versus Descartes on geometrical problem solving and the Sluse-Pascal correspondence. *Early Science and Medicine*, 15:537–565.

[Mathias, 2002] Mathias, A. (2002). A term of length 4,523,659,424,929. *Synthese*, pages 75–86.

[Newton, 1720] Newton, I. (1720). *Universal arithmetick*. J. Senex, W. Taylor, T. Warner, and J. Osborn, London. Reprinted in *The mathematical works of Isaac Newton*, Vol. II. Edited by Derek T. Whiteside, Johnson Reprint Corp., New York, 1967.

[Potter, 2004] Potter, M. (2004). *Set theory and its philosophy*. Oxford University Press, New York.

[Rav, 1999] Rav, Y. (1999). Why do we prove theorems? *Philosophia Mathematica*, 7(3):5–41.

[Restall, 2014] Restall, G. (2014). Normal proofs, cut free derivations and structural rules. *Studia Logica*, 102(6):1143–1166.

[Simpson, 1988] Simpson, S. G. (1988). Partial realizations of Hilbert's Program. *The Journal of Symbolic Logic*, 53(2):349–363.

[Simpson, 2009] Simpson, S. G. (2009). *Subsystems of Second Order Arithmetic*. Cambridge University Press, Cambridge, second edition.

[Statman, 1974] Statman, R. (1974). *Structural complexity of proofs*. PhD thesis, Stanford University.

[Statman, 1978] Statman, R. (1978). Bounds for proof-search and speed-up in the predicate calculus. *Annals of Mathematical Logic*, 15(3):225–287 (1979).

[Szemerédi, 1975] Szemerédi, E. (1975). On sets of integers containing no k elements in arithmetic progression. *Acta Arithmetica*, 27:199–245.

[Takeuti, 1987] Takeuti, G. (1987). *Proof theory*. North-Holland, Amsterdam, second edition.

[Tanswell, 2015] Tanswell, F. (2015). A problem with the dependence of informal proofs on formal proofs. *Philosophia Mathematica*, 23(3):295–310.

[Troelstra and Schwichtenberg, 2000] Troelstra, A. S. and Schwichtenberg, H. (2000). *Basic Proof Theory*, volume 43 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, Cambridge, second edition.

[van Bendegem, 1988] van Bendegem, J. P. (1988). Non-formal properties of real mathematical proofs. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1988:249–254.

[Wilson, 2013] Wilson, R. J. (2013). History of graph theory. In Gross, J. L., Yellen, J., and Zhang, P., editors, *Handbook of Graph Theory, Second Edition*, pages 31–51.

Université de Lorraine, Archives Poincaré (UMR 7117), 91 avenue de la Libération. 54001 Nancy, France

Email address: andrew.arana@univ-lorraine.fr

Department of Mathematics, University of Bristol, Fry Building, Bristol, BS8 1UG, United Kingdom

Email address: will.stafford@bristol.ac.uk