Two Flavours of Mathematical Explanation

Mark Colyvan[*] John Cusbert[†] Kelvin J. McQueen[‡]

Abstract

A proof of a mathematical theorem tells us that the theorem is true (or should be accepted), but some proofs go further and tell us why the theorem is true (or should be accepted). That is, some, but not all, proofs are explanatory. Call this *intra-mathematical explanation* and it is to be contrasted with extra-mathematical explanation, where mathematics explains things external to mathematics. In this paper we focus on the intra-mathematical case. We consider a couple of examples of explanatory proofs from contemporary mathematics. We determine whether these proofs share some common feature that may account for their explanatoriness. We conclude with two plausible, but competing, accounts of mathematical explanation and suggest that there might be more than one kind of explanation at work in mathematics.

1 Introduction

Explanation in mathematics is puzzling. Mathematicians tell us that some proofs are explanatory while others are not.[1] That is, all proofs establish the theorem in question but some proofs go further and explain why the theorem holds.[2] But what kind of thing is an explanatory proof? Some of the usual candidates for explanation in science

[*]Department of Philosophy, University of Sydney, Sydney, NSW, 2006, Australia. *Email:* `mark.colyvan@sydney.edu.au`

[†]Faculty of Philosophy, University of Oxford, Oxford, OX2 6GG, UK. *Email:* `john.cusbert@philosophy.ox.ac.uk`

[‡]Department of Philosophy, Chapman University, Orange, California, 92866, USA. *Email:* `mcqueen@chapman.edu`

[1]For example, see [18, p. 879].

[2]Although we occasionally use the less clumsy realist language of mathematical "truths" and "facts", in this paper we wish to sidestep realism–anti-realism issues. If you're a mathematical realist, explanatory proofs tell us why the theorem is *true*. If you're a mathematical anti-realist you may not believe that the theorem in question is true. You might, instead, think that the theorem is "true-in-the-fiction of mathematics" or some such. In any case, you can, and should, still countenance the distinction between explanatory proofs and non-explanatory ones. The former, may, for example, provide an intra-fiction explanation of the fictional result, just as there are explanations in literary fiction of why some fictional character behaved as she did.

do not seem to work for mathematics. For example, some take explanation to be closely related to causal history but there is no place for causation in mathematics. Similar difficulties arise for counterfactual and interventionist accounts of explanation; mathematics, if true, is a body of necessary truths, so there does not seem to be any room for counterfactuals or intervening.[3]

If we focus on proofs as the locus of explanation in mathematics,[4] one rather natural thought is that mathematical explanations have something to do with the structure of the proof—the explanatory proofs have some especially desirable structure that reveals the reason for the theorem holding.[5] Although we will not argue against this view here,[6] we find it implausible that explanation can be characterised entirely in terms of the structure of the proof. In any case, in this paper we will dig a little deeper—below the level of the structure of the proofs.

To be clear about our target, it's worth distinguishing the kind of explanation we're interested in here from another that's prominent in the literature. *Intra-mathematical explanation* is the explanation of one mathematical fact in terms other mathematical facts. This is to be contrasted with *extra-mathematical explanation*, which is the explanation of some physical phenomenon via appeal to mathematical facts. The existence of such extra-mathematical explanation is still somewhat controversial.[7] We will be firmly focussed on intra-mathematical explanation. More specifically, our interest is in the intra-mathematical explanation found in proofs of theorems.[8]

2 A Few Words About Methodology

In this paper we will look, in some detail, at two different proofs of an important result in group theory: The Free Group Theorem. Neither the theorem nor the two proofs are straightforward but we make no apology for this. It is, in our view, important to tackle examples from more advanced mathematics. It would be all too easy to be misled by focussing on elementary examples from high school mathematics. What is required is a systematic study of proofs from various areas

[3] Although see [7] for some moves in this direction.

[4] It's not clear that proofs are the only place where explanation arises. For example, it might be argued that we find explanation in domain extensions [12, chap. 5].

[5] See for example an exchange between Alan Baker [4] and Marc Lange [23] on the explanatoriness of proofs by mathematical induction.

[6] See [12, chap. 5] for such an argument. For example, in some cases reductio proofs can be transformed into constructive proofs and, in such cases, it seems implausible that the former are not explanatory while the latter are. In such cases, either they are both explanatory or neither are. Either way, there's more to it than merely the structure of the proof.

[7] See [2, 3, 5, 6, 9, 10, 11, 27], for examples of extra-mathematical explanations.

[8] In the past this has received less attention in the philosophical literature on explanation [34, 36, 37], although that seems to be changing, with a number of recent contributions to this topic [12, 13, 19, 17, 24, 25, 29, 30, 32, 33].

of contemporary mathematics—analysis, abstract algebra, topology, number theory and so on. The examples also need to go beyond high-school mathematics.[9] Of course, there will be limits to how advanced the mathematics can be in order for philosophers of mathematics—who are, after all, typically not professional mathematicians—to understand it and be able to draw reliable philosophical morals.[10] Still, those wishing to take our word on the technical details of the proofs and our interpretations of them, can skip to the discussion for the philosophical upshot.

Ideally, we need the judgements of mathematicians on which proofs are, and which are not, explanatory. But mathematicians are notorious for covering their tracks in their written work and rarely commit to print judgements of the explanatory powers of proofs. But as anyone who has spent time with mathematicians knows, such judgements are forthcoming in the tea room, in the pub, and even in the class room. In order to get started on this project we need to scour the literature for the few places where mathematicians do offer judgements on whether the proofs in question are explanatory.[11] Beyond this, talking to, or formally surveying, mathematicians are the obvious ways forward. We decided to informally survey mathematicians on discussion forums, where some, at least, are inclined to give their opinions about such matters.[12] The forum discussion led to our investigation of the Free Group Theorem, in part because the mathematical community seemed to be divided on which proofs of this theorem are explanatory. It's often more fruitful to start with easy cases, but we were intrigued by this theorem and the dispute over its proofs.[13]

To anticipate our conclusions and help see where we are heading with the proofs and subsequent discussion, we suggest that the two proofs in question have different and competing claims for explanatory virtue. The first proof—the so-called constructive proof[14]—delivers the theorem in question via a detailed construction of the group in question and can be thought to be aligned with a model of reductive explanation in science. The second proof—the abstract proof—delivers the theorem by showing how it is one of a more general class of such theorems and as such, this proof can be thought to be aligned with a unificatory model of explanation. Indeed, the fact that this theorem has two such proofs is one of the reasons we chose

[9]Marc Lange has already started this project in his paper [24] in which he discusses some more advanced examples. The present paper can be seen as another step in that direction, although the conclusions we draw from our example are not the same as Lange's conclusions.

[10]Moreover getting on top of proofs from several different areas of contemporary mathematics can be challenging, even for professional mathematicians.

[11]For example: [1], [14] and [20].

[12]See [21] for some interesting formal survey-based work getting at mathematicians judgments about the virtues of various mathematical proofs.

[13]We intend to follow up this present paper with further examples to see if our rather speculative conclusions hold up elsewhere in mathematics.

[14]This name is not meant to suggest that the proof is intuitionistically valid; "constructive" is being used in the non-technical sense here.

to focus on it as our case study.[15]

Another reason for focussing on the Free Group Theorem is that it is an important result; it is a central result in group theory, especially with respect to the presentation of groups, but it is also important for other, related areas of mathematics (e.g. hyperbolic geometry). Moreover, the result and the proofs we discuss are interesting in their own right. Enough about methodology, let's get into the mathematics.

3 The Free Group Theorem

First, recall the definition of a group: A *group* $(G, \cdot)$ is a set of elements G together with a binary operation that together satisfy the four fundamental properties of closure, associativity, the identity property, and the inverse property.

1. *Closure*: If a and $b \in G$, then $a \cdot b$ is also in G.
2. *Associativity*: The group operation is associative, i.e., for all $a, b, c \in G$, $(a \cdot b) \cdot c = a \cdot (b \cdot c)$.
3. *Identity*: There is an identity element $e \in G$ such that $e \cdot a = a \cdot e = a$, for every $a \in G$.
4. *Inverse*: There must be an inverse of each element: for each element $a \in G$, the set contains an element a^{-1} such that $a \cdot a^{-1} = a^{-1} \cdot a = e$.

Definition of *Free Group*: Let X be a set. Group F is free on X if there is a map $f : X \to F$ and for any group K and map $k : X \to K$ there is a unique group homomorphism $\Phi : F \to K$ such that $k = \Phi \circ f$, that is, so that the following diagram commutes.

$$\begin{tikzcd} & X \arrow[dl, "f"'] \arrow[dr, "k"] & \\ F \arrow[rr, "\Phi"'] & & K \end{tikzcd}$$

(This is sometimes expressed in terms of a *universal property*, where the property in question is that which characterises free groups up to isomorphism. It is the property of being such that the above diagram commutes.)

The Free Group Theorem asserts the existence of free groups. More formally, it states that for any set A, there exists a free group on A. (Or equivalently: Given a set A, there exists a free group with basis A.)

[15] It is important to note that the salient difference between the two proofs is not simply that the abstract proof delivers mere existence whereas the constructive proof constructs an example. There are several interesting differences between the two proofs and this is why we run through the proofs in some detail. We do not wish to give a superficial gloss on the two proofs but the differences highlighted in the main text of this paragraph do strike us as central.

4 Free Group Theorem: The "Constructive" Proof

Here we sketch a constructive proof of this result.[16] The proof has two phases: we first use A to construct a group, and define a function from A to that group; and we then prove that together this group and function form a free group on A. (In fact the second phase involves further construction, as we'll see.)

Given an arbitrary set A, we first use A to construct a group. The members of our group will be certain kinds of "words", whose "letters" are built up from members of A. We define the *alphabet* on A as the product $A \times \{-1, 1\}$. Each *letter* in our alphabet is thus an ordered pair $\langle a, \epsilon \rangle$ where $a \in A$ and $\epsilon = \pm 1$. For convenience, we abbreviate $\langle a, 1 \rangle$ as a and $\langle a, -1 \rangle$ as a^{-1}, and we call a and a^{-1} *rivals*.[17] (At this stage we avoid the term 'inverses' since it presupposes a group operation, surrounding which there will be some complications.) (Example: if our set A is $\{a, b\}$, then the alphabet over A is $\{a, b, a^{-1}, b^{-1}\}$.) We then define a *word* on A as a string of alphabet letters of finite length. (Example: $aba^{-1}b$ is a word of length 4.) The *empty word*, written 1, has no letters and length zero. We define the *rival of a word* w, written w^{-1}, as the word obtained by taking the rival of each letter of w and reversing their order. (Example: the rival of aab^{-1} is $ba^{-1}a^{-1}$.) The *concatenation* of words v and w is written vw. This is the word obtained by affixing the head of w to the tail of v. (Example: if $v = ab$ and $w = ab^{-1}$ then $vw = abab^{-1}$.)

Now, it would be nice if we could take our group on A to be the set of words on A, equipped with the operation of concatenation. But this won't work. While concatenation is an associative binary operation on words, and the empty word 1 will serve nicely as an identity element, the problem lies with inverses: nonempty words have no inverses under concatenation. (Example: there is no word that yields 1 when concatenated with ab.) The set of words on A is not a group under concatenation.

To address this problem, we define a special class of words. Call a word *reduced* if it contains no adjacent rival letters. (Example: aba^{-1} is reduced but $a^{-1}ab$ is not.) Note that the empty word 1 is reduced. The set of reduced words on A, written W, will be the base set of our group.[18]

Again though, things are not as straightforward as we'd like. To make W into a group, we'll need to specify a binary operation on W. But concatenation is not a binary operation on W, because the concatenation of two reduced words need not be reduced. (Example:

[16] Our proof sketch relies heavily on [35, 343–5], where further details can be found.

[17] These abbreviations assume that we don't already have $a, a^{-1} \in A$. If we did, then we'd have distinct letters $\langle a^{-1}, 1 \rangle$ and $\langle a, -1 \rangle$ both abbreviated as a^{-1}. In this unfortunate case we can either choose an alternative notation for $\langle a, -1 \rangle$ (perhaps a') or maintain the ordered pair notation.

[18] In general, a base set is a kind of building block. Here we mean that W will be the set from which we are able to build the group in question.

ab and $b^{-1}a$.)

Consequently, we define a second binary operation on W, called *juxtaposition* and written $*$, as follows. Let $v, w \in W$ be reduced words. Let u be the longest tail of v whose rival u^{-1} is a head of w. (There's always some such u: even in the case where vw is reduced, we have $u = 1$.) It follows that there exists a head v' of v and a tail w' of w such that $vw = v'uu^{-1}w'$. (Furthermore, we know that u, u^{-1}, v' and w' are all reduced, because v and w are.) Deleting central rivals gives us $v'w'$, which is *guaranteed to be reduced*. (If it weren't, then u wouldn't have been the longest tail of v such that u^{-1} is a head of u: we could have extended u by at least one letter.) We thus have the *Sandwich Lemma*: for any reduced words $v, w \in W$, there exist reduced words u, v' and w' such that (i) $v = v'u$ (ii) $w = u^{-1}w'$ and (iii) $v'w'$ is reduced. This allows us to define the juxtaposition of v and w by $v * w = v'w'$. Intuitively, juxtaposition amounts to concatenation with cancelling of central rivals. (Example: If $v = aab$ and $w = b^{-1}a^{-1}bb$, then $u = ab$ and $u^{-1} = b^{-1}a^{-1}$, so we have $v' = a$ and $w' = bb$, and so $v * w = abb$.) Juxtaposition (unlike concatenation) *is* a binary operation on W, since a juxtaposition of reduced words is always reduced.

We've now constructed our putative group on A: the set of reduced words on A, equipped with concatenation, or $(W, *)$. (We're yet to *prove* that it's a group: we'll get to that shortly.) Next, we construct our function from A to W. Define $f : A \to W$ such that $f(a) = a$ for all $a \in A$. Thus f simply maps each letter in A to its corresponding one-letter word in W.

Now for the second phase of the proof: showing that $(W, *)$ and f form a free group on A. One might naturally begin by showing that $(W, *)$ is a group. And indeed this seems within reach. First, we've seen that juxtaposition is a binary operation on W. Second, since $w * 1 = w = 1 * w$ for all $w \in W$, the empty word serves an identity element. And third (by contrast with concatenation), each $w \in W$ has an inverse in W under juxtaposition, namely its rival w^{-1}: it's easy to show that for each $w \in W$ we have $w * w^{-1} = 1 = w^{-1} * w$. (We can henceforth dispense with talk of rivals and safely speak of w^{-1} as the inverse of w in W.)

Unfortunately (again by contrast with concatenation), proving the associativity of juxtaposition is tedious. There are various cases to consider, since cancellation of central inverses can proceed differently depending upon how words are grouped. Rather than enumerating the various cases and laboriously producing a separate proof of associativity for each, we will prove that $(W, *)$ and f form a free group on A using the so-called *van der Waerden trick*.

The basic idea is to consider the members of W, not as reduced words, but as permutations of reduced words. To facilitate this, we construct a kind of "scale model" of A, using permutations instead of alphabet letters. We use this scale model of A to construct a scale model of $(W, *)$, using compositions of permutations instead of words. And then we construct a scale model of f that relates the respective scale models of A and W in the same way that f relates A and W. Finally, we prove that the scale models of $(W, *)$ and f form a free group on the scale model of A; and we infer from this that our "original" $(W, *)$ and f form a free group on our original A.

We begin with the general case and then illustrate with a simple example. To each letter $a \in A$ and each $\epsilon = \pm 1$ there corresponds a *single-letter prefixing function* $[a^\epsilon] : W \to W$, defined by $[a^\epsilon](w) = a^\epsilon * w$ for all $w \in W$. The function $[a^\epsilon]$ takes any reduced word $w \in W$ and juxtaposes it with a^ϵ. (Since juxtaposition is a binary operation on W, the word $[a^\epsilon](w)$ is guaranteed to be reduced, and so $[a]$ is a well-defined function from W to itself.) We then prove that $[a^\epsilon] \circ [a^{-\epsilon}] = 1_W = [a^{-\epsilon}] \circ [a^\epsilon]$ for each $a \in A$; and so it follows that each $[a^\epsilon]$ is a permutation on W with inverse $[a^{-\epsilon}]$. (Thus $[a^\epsilon]$ and $[a^{-\epsilon}]$ "undo" each other, and we have $[a^\epsilon]^{-1} = [a^{-\epsilon}]$.) We define $[A]$ as the set of all single-letter prefixing functions $[a]$ where $a \in A$. (This is our "scale model" of A.) Now, we know that the set of *all* permutations on W is a group under composition of functions; and we define $[W]$ as the subgroup of this group generated by $[A]$.[19] (This set $[W]$ under composition of functions is our "scale model" of W under juxtaposition.) Thus $[W]$ is the set of permutations of W of the form $[a_1^{\epsilon_1}] \circ \cdots \circ [a_n^{\epsilon_n}]$ where $a_i \in A$ and $\epsilon_i = \pm 1$. The members of $[W]$ are the permutations of reduced words obtainable by successive prefixing of alphabet letters and/or inverses of alphabet letters. We can think of these as prefixing functions more generally (including both single-letter and multi-letter prefixing functions).

Example: Where $A = \{a, b\}$, our single-letter prefixing functions are $[a]$, $[b]$, $[a^{-1}]$ and $[b^{-1}]$. (Note that $[a]$ and $[a^{-1}]$ are inverse functions, as are $[b]$ and $[b^{-1}]$.) Each of these can be applied to any reduced word: for example we have $[a](bba) = abba$ and $[b^{-1}](bba) = ba$. We thus have $[A] = \{[a], [b]\}$. And so $[W]$ contains all the prefixing functions generated by $[A]$, that is, all possible compositions of $[a]$, $[b]$, $[a^{-1}]$ and $[b^{-1}]$ (with repetitions allowed). For example we have $[a] \circ [a] \circ [b^{-1}] \in [A]$, which amounts to successive juxtaposition with b^{-1}, a and a, so that we have for instance $([a] \circ [a] \circ [b^{-1}])(ba^{-1}b) = ab$.

Note that in $[W]$ we do *not* have unique factorisation into single-letter prefixings: different products of single-letter prefixings can yield the same overall function. (Example: $[a] \circ [a^{-1}] \circ [b] = [b] \circ [b^{-1}] \circ [b]$.) However, if we require that the product resulting from factorisation correspond to a *reduced* word, we do get uniqueness: for each $\sigma \in [W]$ there is a unique reduced word $a_1^{\epsilon_1} \ldots a_n^{\epsilon_n}$ such that $\sigma = [a_1^{\epsilon_1}] \circ \cdots \circ [a_n^{\epsilon_n}]$. We call this factorisation the *reduced form* of σ. (Example: $[b]$ is the reduced form of $[a] \circ [a^{-1}] \circ [b]$.) The uniqueness of reduced forms will be important later.

We then define $[f] : [A] \to [W]$ such that $[f]([a]) = [a]$ for all $[a] \in [A]$. (This is our "scale model" of f.) Thus $[f]$ simply maps each single-letter prefixing function $[a] \in A$ to itself, considered as a prefixing function in $[W]$.

[19] A set of generators $\{g_1, ..., g_n\}$ is a set of group elements such that possibly repeated application of the generators on themselves and each other is capable of producing all the elements in the group. The set of generators is said to generate the relevant group.

Next we show that $[W]$ and $[f]$ form a free group on $[A]$. (From this result regarding the "scale models" we'll easily infer that W and f form a free group on A.) We see immediately that $[W]$ is a group under composition of functions: it's a subgroup of the group of all permutations on W. (In particular, associativity is obvious, and we circumvent the tedious proof mentioned above.)

It remains to prove that for every group $(G, \cdot)$ and every function $g : [A] \to G$, there is a unique homomorphism $\phi : [W] \to G$ such that $g = \phi \circ [f]$. We proceed as follows. Let $(G, \cdot)$ be a group and $g : A \to G$ be a function. Define $\phi : [W] \to G$ such that for each $\sigma \in [W]$ we have:

$$\phi(\sigma) = g([a_1])^{\epsilon_1} \cdot g([a_2])^{\epsilon_2} \cdot \ldots \cdot g([a_n])^{\epsilon_n}$$

where $[a_1^{\epsilon_1}] \circ [a_2^{\epsilon_2}] \circ \cdots \circ [a_n^{\epsilon_n}]$ is the reduced form of σ. To apply ϕ to $\sigma \in [W]$, we first factorise σ into reduced form, then apply g to each factor individually, finally multiplying the results together in G. (The uniqueness of the reduced form ensures that ϕ is a well-defined function on $[W]$.) It follows easily enough that $g = \phi \circ [f]$, since if $[a] \in [A]$, then by the definition of $[f]$ we have $[f]([a]) = [a]$, and so $(\phi \circ [f])([a]) = \phi([a]) = g([a])$, by the definition of ϕ. Showing that ϕ is a homomorphism is more involved. For $\sigma_1, \sigma_2 \in [W]$, with corresponding reduced words $w_1, w_2 \in W$, there are two cases: either the concatenated word $w_1 w_2$ is reduced, or it isn't. If it is, then it follows quickly that $\phi(\sigma_1 \circ \sigma_2) = \phi(\sigma_1) \cdot \phi(\sigma_2)$. If not, then we use the Sandwich Lemma to write $w_1 w_2 = w_1{}' u u^{-1} w_2{}'$ where $w_1{}' w_2{}'$ is reduced. We can therefore apply the same reasoning as in the reduced case to show that $\phi(\sigma_1 \circ \sigma_2) = \phi(\sigma_1) \cdot \phi(\sigma_2)$. We then prove uniqueness: since any homomorphism ψ such that $(\psi \circ [f])([a])$ must agree with ϕ on the generating set $[A]$, it must also agree with ϕ on the whole of $[W]$. We therefore show that $[W]$ and $[f]$ form a free group on $[A]$.

Finally, exploiting the structural similarity between our "scale models" and our "originals", we infer that W and f form a free group on A. Because each prefixing function has a unique reduced product, there's a bijective correspondence between prefixing functions and reduced words; and so the relationship between A, W and f and mirrors that between $[A]$, $[W]$ and $[f]$. We thus see that W and f form a free group on A, as required. ■

5 Free Group Theorem: The "Abstract" Proof

Here we provide a different, more abstract, proof of the existence of free groups.[20] Recall the definition of a *Free Group*: Let X be a set. Group F is free on X if there is a map $f : X \to F$ and for any group K and map $k : X \to K$ there is a unique group homomorphism $\Phi : F \to K$ such that $k = \Phi \circ f$, that is, so that the following diagram commutes.

[20] This proof is due to Michael Barr [8].

$$
\begin{tikzcd}
& X \arrow[dl, "f"'] \arrow[dr, "k"] & \\
F \arrow[rr, "\Phi"'] & & K
\end{tikzcd}
$$

To prove the existence of free group F we will define two other groups on X, G_B and G_α, with respective maps g_B and g_α. For now, think of G_B as (roughly) the group composed of all groups on X (B for 'Big'), and think of G_α as one G_B's components. With some minor qualifications we will show that F can be defined in terms of G_B so that there is a homomorphism (hom) from F to G_B, another homomorphism from G_B to G_α and then another homomophism from G_α to K (for any group K on X). These homomorphisms compose a composite homomorphism from F to K (for any group K on X). This will establish the existence of the homomorphism we are looking for. Effectively, then, we aim to show that the more complex diagram commutes:

$$
\begin{tikzcd}
& & X \arrow[dll, "f"'] \arrow[ddl, "g_B"'] \arrow[ddr, "g_\alpha"] \arrow[drr, "k"] & & \\
F \arrow[dr, "j"'] & & & & K \\
& G_B \arrow[rr, "\pi_\alpha"'] & & G_\alpha \arrow[ur, "\Psi"'] &
\end{tikzcd}
$$

Here j is an inclusion map from F to G_B, π_α is a projection map from G_B to G_α, and Ψ is an isomorphism from G_α to K. The proof aims to show that given the definitions of G_B and G_α, these 3 maps are homomorphisms. And given that homomorphisms compose, it follows that there is a homomorphism from F to K. We therefore set out to prove that $\Phi = \Psi \circ \pi_\alpha \circ j$ such that $k = \Phi \circ f$.

It is easily proved that homomorphisms compose:

Composition Theorem: Let $\beta : G_1 \rightarrow G_2$ and $\alpha : G_2 \rightarrow G_3$ be group homomorphisms. Then the composite map $\alpha \circ \beta : G_1 \rightarrow G_3$ is a homomorphism.

Proof: We show that $\alpha \circ \beta(a \cdot b) = \alpha \circ \beta(a) \cdot \alpha \circ \beta(b)$ for any $a, b, \in G_1$

(where $\cdot$ is the respective group operation):

$$\begin{aligned}
\alpha \circ \beta(a \cdot b) &= \alpha(\beta(a \cdot b)) && [\text{def. of } \alpha \circ \beta] \\
&= \alpha(\beta(a) \cdot \beta(b)) && [\beta \text{ is a hom.}] \\
&= \alpha(\beta(a)) \cdot \alpha\beta(b)) && [\alpha \text{ is a hom.}] \\
&= \alpha \circ \beta(a) \cdot \alpha \circ \beta(b) && [\text{def. of } \alpha \circ \beta]
\end{aligned}$$

Note that the composite diagram breaks down into 3 triangles where the base of each triangle is one of our three component homomorphisms. This enables us to simplify the discussion by establishing that each triangle commutes, before putting the three triangles together to prove that the composite diagram commutes. Let us therefore begin with the first triangle whose base is the inclusion map.

We define F as the subgroup of G_B that is generated by g_B. But before we define G_B and g_B we prove a general theorem that relates any group G to a subgroup H by the inclusion map.

Inclusion Lemma (INC): Let G be a group with map $g : X \to G$. Then there is a subgroup H of G and map $h : X \to H$ such that h generates H and $g = j \circ h$ where j is the inclusion map.

Definition of *generates*:
$h : X \to H$ *generates* $H \equiv$ the image $h(X)$ generates H.
Set A generates group $H \equiv$ no proper subgroup of H contains A.

Proof of INC: Let H be the intersection of all subgroups of G that contain $g(X)$. No proper subgroup of H contains $g(X)$ so $g(X)$ generates H.

Given INC we say: H is the subgroup of G generated by $g : X \to G$. We define F as the subgroup of G_B that is generated by g_B. So the first triangle commutes.

$$\begin{tikzcd}
 & & X \arrow[dll, "f"'] \arrow[dl, "g_B"] \\
F \arrow[r, "j"'] & G_B &
\end{tikzcd}$$

The inclusion map $j : F \to G_B$ is a homomorphism since $j(x) \cdot j(y) = x \cdot y$ defines inclusion maps.

We now move to the middle triangle. We need no theorem to introduce the projection map. Its existence falls out of our definitions of G_B and G_α. Consider a collection of groups on X that are each generated by their corresponding maps. We then have a collection of pairs (G_α, g_α) where each map $g_\alpha : X \to G_\alpha$ generates its corresponding G_α. We now define G_B as the Cartesian product of such groups.

$$\begin{tikzcd}
 & X \arrow[dl, "g_B"'] \arrow[dr, "g_\alpha"] & \\
G_B \arrow[rr, "\pi_\alpha"'] & & G_\alpha
\end{tikzcd}$$

Definiton: $G_B = \prod G_\alpha$
Example: Let $G_1 = \{a_1, b_1\}$ and $G_2 = \{a_2, b_2\}$, then
$G_1 \times G_2 = \{(a_1, a_2), (a_1, b_2), (b_1, a_2), (b_1, b_2)\}$ G_B is a group whose operation entails: $(a_1, a_2) \cdot (b_1, b_2) = (a_1 b_1, a_2 b_2)$, where $a_i b_i$ is the product in G_i.

Definition: $g_B = \prod g_\alpha$
Example: Let $g_1 : X \to G_1$ and $g_2 : X \to G_2$ be maps, then $g_1 \times g_2 : X \to (G_1 \times G_2)$, such that: $(g_1 \times g_2)(x) = (g_1(x), g_2(x))$. So $g_B : X \to G_B$ is a map.

Now we define our projection map. Let $\Pi_\alpha : G_B \to G_\alpha$ be a projection map. Example: $\Pi_1 : G_B \to G_1$ (obviously a homomorphism). So it is clear that: $g_\alpha = \Pi_\alpha \circ g_B$ and the middle triangle commutes.

We now consider the third triangle. Recall: given set X there is a collection of pairs (G_α, g_α) where each G_α is a group and $g_\alpha : X \to G_\alpha$ generates G_α.

$$\begin{tikzcd}
X \arrow[rr, "k"] \arrow[dr, "g_\alpha"'] & & K \\
 & G_\alpha \arrow[ur, "\Psi"'] &
\end{tikzcd}$$

Isomorphism Lemma (ISO): if K is any group and $k : X \to K$ generates K, then for some α there is an isomorphism Ψ on G_α onto K such that $k = \Psi \circ g_\alpha$.
Intuitive Proof of ISO: Simply take all pairs (G_α, g_α) where G_α is a group and g_α generates G_α. Then K is included by definition. That's the

intuitive idea but the problem is that "all" leads to "the usual logical paradoxes". So we turn to a more rigourous poof but first we need some lemmas.

Sub-lemmas for Rigorous Proof of ISO:

Sub-lemma SL1: Let $k : X \to K$ generate K. Then $|K| \leqslant \max(|X|, \aleph_0)$. (proved on p. 366–7 of [8])

Sub-lemma SL2: Y is a set. Then the collection of groups G generated by Y has cardinality less than or equal to $|Y|^{|Y|^2}$. (Just consider the possible ways of filling out the relevant multiplication table.)

Rigorous proof of ISO: For each cardinal s with $s \leqslant \max(|X|, \aleph_0)$ choose a set Y_s with $|Y_s| = s$. Consider the set of all groups G with underlying set Y_s; by SL2 Y_s exists. Consider all pairs (G_α, g_α) where $g_\alpha : X \to G_\alpha$ generates G_α. ISO follows from this construction and SL1.

We now have our three commuting triangles. We put them together to see that the large diagram commutes. But we must consider two separate cases, the second of which requires a special qualification.

Recall the definition of F: Subgroup of G_B generated by g_B. By INC: $g_B = j \circ f$.

Case 1 (already depicted in the previous large diagram): Assume $k : X \to K$ generates K. By ISO there is an α and an isomorphism Ψ such that $k = \Psi \circ g_\alpha$ (third triangle). Note: $g_\alpha = \pi_\alpha \circ j \circ f$ (combining first and second triangles). Then by substitution: $k = \Psi \circ \pi_\alpha \circ j \circ f$ (combining all three). But the composite $\psi \circ \pi_\alpha \circ j$ is a homomorphism in virtue of its components being homomorphisms. So let $\Phi = \Psi \circ \pi_\alpha \circ j$. Case 1 is thus closed.

Case 2 (depicted below): Assume $k : X \to K$ does not generate K. By INC there is a subgroup K' of K such that $k' : X \to K'$ generates K' and $k = j' \circ k'$, where $j' : K' \to K$ is an inclusion map. We know there is a homomorphism $\Phi' : F \to K'$, such that $k' = \Phi' \circ f$ (see case 1). So let $\Phi = j' \circ \Phi'$ since $j' : K' \to K$ is a homomorphism. Then $\Phi : F \to K$ is a homomorphism and $k = \Phi \circ f$. Case 2 is also closed.

$$
\begin{tikzcd}
& & X \arrow[rr, "k"] \arrow[dll, "f"'] \arrow[ddl, "g_B"'] \arrow[ddr, "g_\alpha"] \arrow[drr, "k'"] & & K \\
F \arrow[dr, "j"'] & & & & K' \arrow[u, "j'"'] \\
& G_B \arrow[rr, "\pi_\alpha"'] & & G_\alpha \arrow[ur, "\Psi"'] &
\end{tikzcd}
$$

Uniqueness of Φ

Let: $\Phi_1 : F \to K$ and $\Phi_2 : F \to K$ be homomorphisms, where $\Phi_1 \circ f = \Phi_2 \circ f = k$. Let F_0 be the set of x in F such that $\Phi_1(x) = \Phi_2(x)$. But then F_0 is a subgroup of F. Now we prove that $f(X) \subseteq F_0$ For all $x \in X : \Phi_1 \circ f(x) = \Phi_2 \circ f(x)$ (by def. of Φ_i). $\Phi_1(f(x)) = \Phi_2(f(x))$)(by def. of $\circ$). But $f(X)$ generates F (because F is free). So $F_0 = F$, hence $\Phi_1 = \Phi_2$. The proof of the theorem is complete. ■

6 The Explanatory Value of the Proofs

6.1 The Constructive Proof and Local Dependence-based Explanatory Value

The constructive proof explains the existence of free groups by building them and showing how the intrinsic structure of what's built guarantees the universal property. For example, the proof builds "words" and "letters" from members of an arbitrary set, before defining a suitable group operation. The proof then shows how the universal property depends on features of this construction. This structural feature of the proof appears to fit the dependence-based model of explanation in the philosophy of science. In that case, assuming that a proof has explanatory value if it fits this model of explanation, then the constructive proof has a distinctive kind of explanatory value.

The dependence-based model of explanation really breaks into two distinct but analogous models. On the one hand there are dynamic causal theories on which (roughly) a phenomenon is explained by describing the cause that the phenomenon causally depends on.[21] On the other hand there are synchronic reductive theories on which (roughly) a phenomenon is explained by describing the underlying structure or process that the phenomenon reduces to, or metaphysically depends on [15]. The reductive theories are somewhat closer to what we are after. Such theories try to model what is happening, for example, in the reduction of the laws of thermodynamics to statistical mechanics, and in the reduction of rigid body mechanics to particle mechanics.

In the former case, hypothetical statistical mechanical systems are constructed, and we are shown how the principles of thermodynamics fall out of these constructions plus the statistical mechanical laws. To explain the Boyle-Charles Law, for example, one constructs an idealized gas and describes it in terms of Newton's laws. One then shows (by deduction) that the mean kinetic energy of the gas particles gives rise to the Boyle-Charles Law, since it can be deduced by identifying temperature with mean kinetic energy [15, § 2]. In the latter case, one constructs a hypothetical idealized microphysical system, and shows how the principles of rigid-body mechanics fall out of these constructions plus the microphysical laws. To explain the principle of mass additivity, for example, one constructs an idealized microphysical system and describes it in terms of Newton's laws. One then shows (by

[21] For example, see [26].

deduction) that the mass of its composite is the sum of the masses of its elementary components [31].

The constructive proof of free groups does something very similar, and so should therefore be similarly described by reductive models of explanation. For in this proof, a group is constructed out of a set, and we are shown how the universal property definitive of free groups falls out of this construction plus principles of group theory.[22] We believe that this is at least similar to Steiner's [36] idea that (i) to explain the behaviour of an entity, one deduces the behaviour from the essence or nature of the entity and (ii) mathematical proofs exhibit this deductive structure.[23] Thus, already existing models of explanation give us good reason to think the constructive proof is explanatory, since the constructive proof satisfies the key requirements of such models.[24]

The abstract proof does not do this. Although we have a construction given in terms of the Cartesian product of all groups on a set, we are given no information about the intrinsic structure of these component groups. Instead the abstract proof works with abstract relationships among groups to show how those relationships guarantee that (the subgroup of) this Cartesian product satisfies the universal property. But we are left wondering what it is about the intrinsic structure that guarantees this.[25] And so if a proof can only have explanatory value if it is modelled by a dependence account, then the abstract proof may be seen to be unexplanatory. But this is too quick. There are explanatory virtues in the abstract proof, but they are apparently of a different kind.

6.2 The abstract proof and global unification-based explanatory value

Consider a different kind of explanatory virtue based on the unificationist account of explanation found in the philosophy of science [16, 22]. On the unification approach in the philosophy of science, an

[22]Perhaps talk of "falls out" sounds a bit loose and does not get at the core distinction between a proof and an explanatory proof. But here we don't mean simply that the universal property logically follows from the construction in question, rather, we mean that the universal property naturally arises from the core properties of the construction in question.

[23]This is also similar to Colyvan's [12, chap. 5] suggestion that relevance (in the technical sense) might be a way of spelling out this local/intrinsic notion of explanation in mathematics.

[24]Here we're arguing by analogy. We're appealing to accepted similarities between two things (a reductive explanation and a constructive proof) to support the conclusion that some further similarity between them exists (namely, explanatory value). Of course Steiner's account of mathematical explanation has its critics (e.g. see [29, 30]) but to be clear, we are not suggesting that Steiner's account is correct or problem free. We are merely noting that there is an explanatory virtue found in the constructive proof that might be fruitfully spelled out along similar lines to at least some parts of Steiner's account.

[25]This line of thought was expressed by some mathematicians and physicists in our informal discussions on the Physics Forum.

event is explained by deriving the occurrence of the event using a theory that unifies many diverse phenomena, and thereby showing that the event is part of a very general, perhaps utterly pervasive, pattern of events in the universe. In the best-developed unification account of explanation, due to Philip Kitcher, an event is explained by deducing it using the theory that unifies the phenomena better than any other.[26]

One can straightforwardly adapt this philosophy of science account to the philosophy of mathematics by replacing "event" and "occurrence of event" with "theorem", meanwhile "theory" can be replaced by "proof". For example, on the unificationist approach in the philosophy of mathematics, a theorem is explained by deriving the theorem using a proof that unifies many diverse theorems, and thereby showing that the theorem is part of a very general, perhaps utterly pervasive, pattern of theorems in mathematics. With this in mind, consider what Michael Barr says about the abstract proof:

> The proof is modelled after that of the general adjoint functor theorem of category theory and, as such, is readily adapted to solving any universal mapping problem in the category of groups, such as the existence of free products. It also works in any category consisting of all the algebras and algebra homomorphisms of any algebraic theory. [...] Thus included are all such categories as sets, sets with a base point (and base-point preserving functions), groups, abelian groups, rings, commutative rings, Lie rings, Jordan rings, algebras of these types, etc., each considered as a category with the evident definition of homomorphism. [8, p. 364]

We can therefore say that the free group theorem is explained by the abstract proof because the abstract proof unifies many diverse free object existence theorems, and thereby shows that the free group theorem is part of a very general, persuasive, pattern of theorems in mathematics (free object theorems).

We need not think of the unification theory as being *the* theory of explanation, just as we need not think of the dependence theory as being *the* theory of explanation. We only need to think of them as providing a means of spelling out a source of explanatory value. These sources need be neither necessary nor sufficient conditions for possessing explanatory power. If that's right, then since the abstract proof fits so nicely into the philosophy of mathematics version of the unification account, then arguably the abstract proof has distinctive explanatory value.[27]

Perhaps the abstract proof also has some claim to exhibiting the

[26] This particular formulation is due to Strevens [38].

[27] In particular, given that we are not committed to the unification account being *the* account of explanation, some of the noted shortcomings of unificationism need not concern us. Indeed, the problem cases for Kitcher's account do not undermine it completely but, rather, serve to highlight its limitations as the complete account of explanation. (See [39] for discussion on this issue.)

reduction-style virtues. This may well be but if so, such virtues are not prominent. The abstract proof does not build a specific group and prove that there must be a free group from those specifics. That's why it's not meant to be analogous to reductive explanations, which (e.g. in the rigid-body mechanics case) build a specific microphysical system and deduce a given property (mass additivity, say) from those specifics. Rather, the relevant property (freeness) is derived from certain very general properties of groups and this becomes explanatory insofar as the proof can enable one to see how it can be generalised to other domains, thereby unifying them. But it's plausible that these things come in degrees. Perhaps the abstract proof is somewhat reductive, even if that's not the feature of it that yields most of its explanatory value, and perhaps the constructive proof is somewhat unificatory even if that's not the feature of it that yields most of its explanatory value.

6.3 Which proof is more explanatory?

If the above is a correct characterisation of the situation, then the two proofs are hard to compare because their primary sources of explanatory value differ. For example, if both proofs were modelled only by the unificationist model, we could ask which proof has the more general scope, or which proves a range of theorems using the most minimal basic assumptions (or both). Determining which is more explanatory wouldn't exactly be straightforward but we would at least have a clear way forward. But it is not clear how to compare the relative explanatory strengths of proofs whose primary explanatory value come from such distinct sources.

What do mathematicians think? Some comments from Saunders Mac Lane on this issue are interesting. Mac Lane notes that one of the applications of the category theory Representability Theorem is that it facilitates a neat (category theory) proof of the Free Group Theorem "without entering into the usual (rather fussy) explicit construction of the elements of [the free group on X] as equivalence classes of words in letters of X" [28, p. 123]. It's not clear from Mac Lane's comments whether this claimed advantage for the abstract approach to proving the Free Group Theorem is supposed to be an *explanatory* advantage. At first blush the advantage looks pragmatic: the abstract proof avoids some tedious constructions and saves some ink. But the claimed advantage can also be thought of as an explanatory advantage: whereas the constructive proof bogs down in detail, the category theory proof rises above such details to reveal the real reasons for the existence of free groups. According to this line of thought, the real reason for the existence of free groups is found at the more abstract structural level. The existence of free groups is just a special case of a more general result, so focussing on the details in group theory is to miss the point, or so the suggestion goes.[28] A related suggestion is that

[28]This line of thought was also expressed by some of the mathematicians and physicists on the Physics Forum discussion.

a proof of free groups is explanatory to the extent that it helps justify/secure/illuminate the applications of free groups in group theory. If this is on the right track, it may provide a neutral way of comparing the explanatory power of proofs whose respective explanatory values come from different sources (e.g. local dependence versus global unification).

So we can ask, to what extent is the subgroup of G_B generated by g_B that which one has in mind when engaged in applications? Analogously: to what extent is the free group with group operation juxtaposition that which one has in mind when engaged in applications? Does it ever make a difference? These are not questions we can answer here but these are the kinds of questions that need to be addressed in advancing our understanding of mathematical explanation.[29]

Our conclusion may seem a little unsatisfying: there is good reason to suspect that there are two competing candidates for explanatory power in mathematics—two flavours of mathematical explanation, if you like—and it is difficult to make trade offs between the two. But as we said at the outset, mathematical explanation is puzzling—puzzling enough that we should be suspicious of any account that promises easy answers. In any case, we make no apologies for not offering easy answers. Instead, we offer a case study that we believe is helpful in shedding light on the nature(s) of mathematical explanation. We have argued that a given theorem admits two intuitively explanatory proofs, one which is structurally similar to reductive explanation, another which is structurally similar to unificationist explanations. We speculate that the explanatoriness derives from these structures.

Although it is common to talk of a proof being explanatory or not, and we too mostly follow this way of talking, it seems to us that it is more plausible that explanatory virtues come in degrees. Those proofs that exhibit an explanatory virtue to a high degree are those that we speak of as being explanatory. (Just as belief comes in degrees and if the degree is high enough we tend to treat that as full belief.) But accepting that explanatory virtue comes in degrees and that there is more than one kind of explanatory virtue does not trivialise the view. It does not, for example, mean that all proofs are explanatory because they are all explanatory to some degree in some explanatory virtue or other. The proofs in question need to exhibit the explanatory virtue(s) to a suitably high degree. What is a

[29] In this spirit, here are a couple of specific applications to think about:
(i) Often one proves a result about groups by first establishing the result for free groups and then showing how it holds for the quotient of these groups. When these groups are abelianized (mod out by commutators) this has important consequences for computing things like Ext and Tor.
(ii) Every group is a quotient of two free groups. (Let G be any group and Let F_G be the free group generated by the elements of G. The universal property of this free group provides a homomorphism $F_G \to G$ and let K denote its kernel. By the first isomorphism theorem it follows that $F_G/K = G$ and since subgroups of free groups are free, this establishes that every group is a quotient of two free groups.) This entails that every group has a presentation. (The generators are given by the generators of FG and the relations are given by the generators of K.)

high enough degree? This is probably context sensitive and perhaps also vague, but there will be clear cases on either side. There will also be some difficult comparisons—even among the clear cases of explanatory proofs. Indeed, the two proofs in this paper illustrate such difficulties: one proof is high in the unificatory stakes and low in the reductionist stakes (the abstract proof), the other is high in the reductionist stakes and low in the unificatory stakes (the constructive proof). But according to our account, both proofs are explanatory, albeit explanatory for different reasons. Each is explanatory because it exhibits one of the explanatory virtues to a high degree but it is not clear how to compare these two kinds of explanatory virtues so there is no straightforward way to say which proof is the more explanatory. Indeed, there may be no fact of the matter about such comparisons.

As we have already noted, further philosophical work needs to be done on understanding the broader roles of free groups in mathematics to see which of the proofs of the theorems in question best support these roles. We also need to look at proofs of theorems from a variety of areas of mathematics to see if the same issues arise.[30] Finally, we need greater collaboration between mathematicians and philosophers on this project. This is not something philosophers can do alone. Most philosophers' intuitions about explanatory power in mathematics run out fairly quickly and, in any case, are unlikely to be reliable. Our case study of the Free Group Theorem is just a small step towards a better understanding of the intricacies of the explanatory virtues of different proofs.[31]

References

[1] Aigner, M. and Ziegler, G. M. 2010. *Proofs from the Book*, 4th edition, Heidelberg: Springer.

[2] Baker, A. 2005. 'Are There Genuine Mathematical Explanations of Physical Phenomena?', *Mind*, 114(454): 223–238.

[3] Baker, A. 2009. 'Mathematical Explanation in Science', *British Journal for the Philosophy of Science*, 60(3): 611–633.

[4] Baker, A. 2010. 'Mathematical Induction and Explanation', *Analysis*, 70: 681–689.

[30] We intend to take up this issue elsewhere.

[31] We are grateful to Sam Baron, Rachael Briggs, Clio Cresswell, Ed Mares, Daniel Nolan, Jeff Pelletier, Graham Priest, Dave Ripley and Jamie Tappenden for discussions about the material covered in this paper. We are also grateful to several mathematicians and physicists who contributed to our discussions on Physics Forum. Their insights about explanation in mathematics were extremely helpful, as was their suggestion of looking at the Free Group Theorem. Material from this paper was presented to the 2014 Australasian Association of Logic Conference at the University of Sydney. We are grateful to the audience in attendance for their very helpful comments and suggestions. We'd also like to acknowledge Manya Raman-Sundström and the anonymous referees for this volume for many helpful comments on earlier versions of this paper. This work was funded by an Australian Research Council Future Fellowship grant to Mark Colyvan (grant number FT110100909).

[5] Baron, S. 2014. 'Optimization and Mathematical Explanation: Doing the Lévy Walk', *Synthese*, 191(3): 459–479.

[6] Baron, S. and Colyvan, M. 2016. 'Time Enough for Mathematical Explanation', *Journal of Philosophy*, 113(2): 61–88.

[7] Baron, S., Colyvan, M. and Ripley, D. Forthcoming. 'How Mathematics Can Make a Difference', *Philosophers' Imprint*.

[8] Barr, M. 1972. 'The Existence of Free Groups', *American Mathematical Monthly*, 79(4): 364–367.

[9] Colyvan, M. 2001. *The Indispensability of Mathematics*, New York: Oxford University Press.

[10] Colyvan, M. 2002. 'Mathematics and Aesthetic Considerations in Science', *Mind*, 111(441): 69–74.

[11] Colyvan, M. 2010. 'There is No Easy Road to Nominalism', *Mind*, 119(474): 285–306.

[12] Colyvan, M. 2012. *An Introduction to the Philosophy of Mathematics*, Cambridge: Cambridge University Press.

[13] Colyvan, M. Forthcoming. 'The Ins and Outs of Mathematical Explanation', *Mathematical Intelligencer*.

[14] Davis, P. J. and Hersh, R. 1981. *The Mathematical Experience*, Boston: Birkhäuser.

[15] Dizadji-Bahmani, F., Frigg, R., and Hartmann, S. 2010. 'Who's Afraid of Nagelian Reduction?', *Erkenntnis*, 72(3): 303–322.

[16] Friedman, M. 1974. 'Explanation and Scientific Understanding', *Journal of Philosophy*, 71(1): 5–19.

[17] Giaquinto, M. 2016. 'Mathematical Proofs: The Beautiful and the Explanatory', *Journal of Humanistic Mathematics*, 6(1): 52–72.

[18] Gowers, T. and Neilson, M. 2009. 'Massively Collaborative Mathematics', *Nature*, 461 (15 October 2009), 879–881.

[19] Hafner, J. and Mancosu. P. 2005. 'The Varieties of Mathematical Explanation', in P. Mancosu, K. F. Jørgensen and S. A. Pedersen (eds.), *Visualization, Explanation and Reasoning Styles in Mathematics*, Dordrecht: Springer, pp. 215–250.

[20] Hardy, G. H. 1967. *A Mathematician's Apology*, Cambridge: Cambridge University Press.

[21] Ingliss, M. and Aberdeen, A. 2015. 'Beauty Is Not Simplicity: An Analysis of Mathematicians' Proof Appraisals', *Philosophia Mathematica*, 23(1): 87–109.

[22] Kitcher, P. 1981. 'Explanatory Unification', *Philosophy of Science*, 48(4): 507–531.

[23] Lange, M. 2009. 'Why Proofs by Mathematical Induction are Generally Not Explanatory', *Analysis*, 69(2): 203–211.

[24] Lange, M. 2014. 'Aspects of Mathematical Explanation: Symmetry, Unity, and Salience', *Philosophical Review*, 123(4): 485–531.

[25] Lange, M. 2016. 'Explanatory Proofs and Beautiful Proofs', *Journal of Humanistic Mathematics*, 6(1): 8–51.

[26] Lewis, D. 1986. 'Causal Explanation', in *Philosophical Papers Vol. II*. Oxford University Press, New York, pp. 214–240.

[27] Lyon, A. and Colyvan, M. 2008. 'The Explanatory Power of Phase Spaces', *Philosophia Mathematica*, 16(2): 227–243.

[28] Mac Lane, S. 1998. *Categories for the Working Mathematician*, second edition, New York: Springer.

[29] Mancosu, P. 2008a. 'Mathematical Explanation: Why it Matters', in P. Mancosu (ed.), *The Philosophy of Mathematical Practice*, Oxford: Oxford University Press, pp. 134–149.

[30] Mancosu, P. 2008b. 'Explanation in Mathematics', in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2008 Edition), URL = http://plato.stanford.edu/archives/fall2008/entries/mathematics-explanation/.

[31] McQueen, K. J. 2015. 'Mass Additivity and A Priori Entailment', *Synthese*, 192(5): 1373–1392.

[32] Pincock, C. 2015. 'The Unsolvability of the Quintic: A Case Study in Abstract Mathematical Explanation', *Philosophers' Imprint*, 15(3): 1–19.

[33] Raman-Sundström, M. and Öhman, L. Forthcoming. 'Mathematical Fit: A Case Study', *Philosophia Mathematica*.

[34] Resnik, M. and Kushner, D. 1987. 'Explanation, Independence, and Realism in Mathematics', *British Journal for the Philosophy of Science*, 38: 141–158.

[35] Rotman, J. J. 1965. *The Theory of Groups: An Introduction*, Boston: Allyn and Bacon.

[36] Steiner, M. 1978a. 'Mathematical Explanation', *Philosophical Studies*, 34: 135–151.

[37] Steiner, M. 1978b. 'Mathematics, Explanation, and Scientific Knowledge', *Noûs*, 12: 17–28.

[38] Strevens, M. 2004. 'The Causal and Unification Approaches to Explanation Unified—Causally', *Noûs*, 38(1): 154–176.

[39] Tappenden, J. 2005. 'Proof Style and Understanding in Mathematics I: Visualization, Unification and Axiom Choice', in P. Mancosu, K. F. Jørgensen and S. A. Pedersen (eds.), *Visualization, Explanation and Reasoning Styles in Mathematics*, Dordrecht: Springer, pp. 147–214.