Towards a Philosophical Account of Explanation in Mathematics

A PhD thesis by Josephine Salverda

University College London

I, Josephine Salverda, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

All proofs show that their conclusions are true; some also explain why they are true. But what makes a proof (or argument) explanatory, if it is? That is the central question of my thesis.

I begin by identifying four accounts of scientific explanation which look like they might be useful for the intra-mathematical case, assessing the prospects for extending each account to the mathematical case.

I examine whether we could get to a general result about mathematical explanation while drawing only on general assumptions about explanation. I argue that this methodology is flawed and that we need to pay serious attention to specific examples from mathematical practice, not just to general assumptions.

I examine two existing accounts of intra-mathematical explanation: first, Steiner's 1978 account. I propose a new and sympathetic reading that provides a better understanding of his account than can be found in the existing literature. Although Steiner's account seems to focus on ontic aspects of explanation, I show how (my extension of) Steiner's proposal can also account for what I take to be the primary epistemic function of an explanation, namely, to help us see why the fact to be explained is true.

Second, I examine Lange's 2017 account, which focuses on salient features. Of the features proposed by Lange, I suggest that symmetry is the best candidate for a feature of mathematical explanation, and I argue that we should see symmetry as an objective mathematical property that may have the propensity to appear salient to creatures like us in certain contexts.

I argue that it is philosophically fruitful to play close attention to candidate examples of mathematical explanation, and in Chapter 5 I present an in-depth case study of a proof in Galois theory and propose a positive account of its explanatory value.

Impact statement

The primary impact of my thesis is likely to be inside academia: helping to add clarity to the debate on intra-mathematical explanation, situated within my own research area of philosophy of mathematics and also relevant for adjacent research areas in philosophy of science.

The research output will also be largely academic: two papers based on work in this thesis have been published so far [Salverda 2017, 2018] and I hope to build on later chapters to produce future papers.

It is possible that a successful account of intra-mathematical explanation may be of pedagogical use in mathematics education, and this is something I would be keen to explore.

Acknowledgements

First and foremost, my thanks go to my two supervisors, Luke Fenton-Glynn and Marcus Giaquinto, for stimulating discussion and their wise and sharp-eyed comments; and to Marcus especially for taking the time to supervise me (and learn Galois theory!) in his retirement.

My studies were funded by the UK Arts and Humanities Council, the Royal Institute of Philosophy and the UCL philosophy department.

I am grateful to feedback received from audiences in London, Cambridge, Umea, Pescara and Brussels, as well as from two anonymous referees for the articles which have been published based on work in this thesis. My thanks also to the publishers for allowing me to reproduce that work here.

To my friends and PhD colleagues for providing much-needed company and support, in particular Lottie Aron, Neil Barton, Lauren Clark, Annika Coughlin, Miriam Craik-Horan, Kathrine Cuccuru, Mog Hampson, Katharina Herold, Ruth Macdonald, Anna Mayer, Holly Page, Lea Salje and Daniel Wood.

Finally, my thanks to my family, Reinier Salverda, Elisabeth Salverda, Johanna van Benthem, and my partner Christopher Cook, without whom this thesis would not exist.

Contents

0 Introduction

0.1 Mathematical motivation

This thesis focuses on mathematical explanation. As my primary motivation for taking this topic seriously, I will first provide some evidence that the phenomenon is taken seriously by mathematicians. Some philosophers deny this. For example, Mark Zelcer writes that ‘Great pains are taken by mathematicians to further the goals of mathematics. Among their goals are proof, perhaps precision, and the development of new methods, not explanation’ [Zelcer 2013: 179]. Michael Resnik and David Kushner remark that ‘Mathematicians rarely describe themselves as explaining’ [Resnik and Kushner 1987: 151], and Emily Grosholz claims that ‘mathematical truths are constructed or demonstrated, not explained’ [Grozholz 2000: 81].

However, judgments made by mathematicians tell a different story. For example, professor of mathematics John Baldwin considers Resnik and Kushner’s remark and writes that ‘As a mathematician, I can only explain such a remark as a lack of exposure to mathematicians’ [Baldwin 2016: 70].

In a joint article, Fields medallist Timothy Gowers writes that ‘for mathematicians, proofs are more than guarantees of truth: they are valued for their explanatory power, and a new proof of a theorem can provide crucial insights’ [Gowers and Nielsen 2009: 879]. In his PhD thesis, David Sandborg performs a database search of the text of *Mathematical Reviews* between 1980 and 1997. Eliminating cases where words like ‘explain’ and ‘explanation’ are used in other contexts (such as historical and sociological), he identifies 377 published occurrences of ‘mathematical explanations of mathematical facts’ [Sandborg 1997: Chapter 3 and Appendix].

Johannes Hafner and Paolo Mancosu provide further evidence that mathematicians often describe themselves as explaining, citing illustrative examples from a wide range of mathematical fields [Hafner and Mancosu 2005: 218-21].

Even where the words ‘explain’ or ‘explanation’ are not explicitly used, mathematicians are often sensitive to the distinction between proofs that merely show that their conclusion is true and those that explain why it is true. For example, in an interview Fields medallist Michael Atiyah recalls:

> 'I remember one theorem that I proved and yet I really could not see why it was true. It worried me for years ... I kept worrying about it, and five or six years later I understood why it had to be true. Then I got an entirely different proof ... Using quite different techniques, it was quite clear why it had to be true' [Minio 1984: 17].

Similarly, mathematician Edward Frenkel describes the role that explanation-seeking plays in his mathematical practice:

> 'My proof of this result was technically quite involved. I was able to explain how the Langlands dual group appeared, but even now, more than twenty years later, I still find it mysterious why it appears. I solved the problem, but it was ultimately unsatisfying to feel that something just appeared out of thin air. My research since then has been motivated in part by trying to find a more complete explanation.
>
> It often happens like this. One proves a theorem, others verify it, new advances in the field are made based on the new result, but the true understanding of its meaning might take years or decades.' [Frenkel 2013: 181]

Now, we might worry about taking judgments made by mathematicians as a reliable guide towards a philosophical account, despite the high pedigree of the mathematicians cited (including two Fields medallists). Note however that many of the remarks are not throwaway comments but arise from careful reflection on the commenter's own practice, such as the cited remarks by Michael Atiyah and Edward Frenkel. I suggest that even where mathematicians are initially sceptical that there is an interesting distinction between explanatory and non-explanatory proofs, this scepticism often dissolves on deeper reflection.

For example, when initially considering the topic of mathematical explanation, renowned mathematician David Gale writes: 'In mathematics the explanation is the proof. It's as simple as that' [Gale 1990: 4]. A year later, Gale considers a specific example of a proof about Somos sequences by Dean Hickerson, and writes:

> 'But what have we learned? As Hickerson puts it, "The thing I dislike about my proof is that it doesn't explain why the result is true. It depends primarily on the fact that when you compute a_{12} there's an unexpected cancellation. But why does this happen?" Indeed the proof, rather than illuminating the phenomenon, makes it, if anything, more mysterious. I report this with some embarassment [sic], since I have earlier asserted in this same journal that a proof in mathematics is in some sense equivalent to an explanation. We now see that this clearly need not be the case.' [Gale 1991: 41].

On careful reflection, Gale takes it that there is after all an interesting distinction between explanatory and non-explanatory proofs.[1]

This selection of remarks provides, I think, enough initial evidence for us to take the phenomenon of mathematical explanation seriously. For those readers still sceptical that there is such a thing as explanation in mathematics, I hope the examples discussed throughout the thesis will provide further motivation.

0.2 Motivation from philosophy of science

Now, supposing there is indeed such a thing as explanation in mathematics, we might still wonder why my project is of general philosophical interest. As philosophers, of course, we try to account for phenomena such as explanation in various fields of knowledge. But a secondary motivation to care about an account of explanation in mathematics in particular is the potential benefits such an account could bring to philosophy of science.

[1] To the detriment of his argument, Zelcer considers only Gale's first remark as evidence for the claim that there is no interesting explanatory / non-explanatory distinction in mathematics. Marc Lange is more careful and cites both papers [Lange 2014: 490-1, fn 10].

For example, Mark Steiner thinks there is a close connection 'between mathematical and physical explanations of physical phenomena ... In the former, as in the latter, physical and mathematical truths operate. But only in mathematical explanation [in science] is this the case: when we remove the physics, we remain with a mathematical explanation – of a mathematical truth!' [Steiner 1978b: 19]. On this view, an account of intra-mathematical explanation would do much to account for mathematical explanation in science. Similarly, in some of his papers Marc Lange makes assumptions about intra-mathematical explanation on the grounds that if these assumptions did not hold, 'mathematical explanation would be nothing at all like scientific explanation' [Lange 2009: 206]. This suggests Lange takes the two cases to be relevantly similar (and he presents specific ideas about what makes for a distinctively mathematical explanation in science in e.g. [Lange 2013]).

On the other hand, Alan Baker claims that 'the predominant view is that mathematical explanation is qualitatively different both from scientific explanation and from explanation in "ordinary non-scientific contexts"' [Baker 2012: 244]. I am open to a pluralist view of explanation, but it also seems possible that intra-mathematical and scientific explanation have features in common. An account of explanation in science could provide insights that carry over to the purely mathematical case; alternatively, the failure of certain features to carry over could also be illuminating.

Recent discussions on explanation in philosophy of science have included a focus on specific examples of mathematical explanation in science, i.e. cases where mathematical statements or facts are taken to play an explanatory role in the scientific explanation of a physical phenomenon. Such examples are popular in the debate between mathematical realists and nominalists. Mathematical realists such as Alan Baker claim that mathematical claims play a genuinely (and indispensable) explanatory role in scientific explanations, and argue (roughly) that since we should believe in the truth of our best scientific explanations, we should also believe in the truth of the mathematical claims indispensably involved in those explanations. Nominalists disagree, arguing for example that mathematical claims merely serve to model or index the physical phenomenon and do not play a genuinely explanatory role, and that we should only believe in the truth of claims that play a genuinely explanatory role. The debate over what it means for mathematics to play a genuinely explanatory role continues (see e.g. [Baker 2017]), and an account of intra-mathematical explanation might help the debate to progress. Indeed, two of the most prominent authors in this debate note that '... a philosophical account of mathematical explanation is something sorely needed for both philosophy of mathematics and philosophy of science' [Baker and Colyvan 2011: 333].

Throughout this thesis, I will reserve judgement about potential links between mathematical and scientific explanation; if there is an analogy between the two cases, this should become apparent from our reasoning about each case, rather than featuring as an assumption. At the very end of the thesis, however, I will return to this issue and present an example of intra-mathematical explanation that looks like it might transfer very neatly to a concrete example of mathematical explanation in science.

0.3 Thesis outline

In Chapter 1, I first draw a distinction between ontic and epistemic accounts of explanation, which will be useful in examining various existing accounts of explanation. I identify four accounts of scientific explanation which look like they might be useful for the intra-mathematical case, two ontic accounts (counterfactuals and grounding), and two epistemic accounts (why-questions and unification). I discuss each of these four accounts in some depth, assessing the prospects for extending each account to the mathematical case. I identify two accounts as the most promising to extend to the mathematical case: (1) counterfactuals and (2) why-questions. I will discuss the two possible extensions in Chapters 3 and 4.

In Chapter 2, I first examine whether we could get to a general result about mathematical explanation while drawing only on general assumptions about explanation. This is the methodology taken by Lange in his 2009 paper on induction, in which he argues that inductive proofs are never explanatory. I argue that this methodology is flawed and that we need to pay serious attention to specific examples from mathematical practice, not just to general assumptions. I present some convincing examples of explanatory inductive proof. These examples will also help me to test the two accounts to be considered in Chapters 3 and 4. I suggest that the presentation of a proof is an important and explanatorily relevant feature, a suggestion that will be further developed in Chapters 3 to 5.

In Chapter 3, I examine Mark Steiner's account of explanation, which I argue can be seen as an extension of counterfactual accounts of explanation in science. I propose a new and sympathetic reading that provides a better understanding of his account than can be found in the existing literature. Although Steiner's account seems to focus on ontic aspects of explanation, I show how (my extension of) Steiner's proposal can also account for what I take to be the primary epistemic function of an explanation, namely, to help us see why the fact to be explained is true. I emphasise the importance of presentation (in this case the presentation of the relevant characterizing property) and lay the ground for my own ideas about the connection between explanation and presentation, which will be explored in Chapter 5.

In Chapter 4, I examine Marc Lange's recent account of explanation, investigating a worry that it might fail as a why-question account in cases where an explanatory proof answers a why-question that could not have been formulated simply by understanding the meaning of the theorem-proposition. I consider a number of the examples Lange puts forward and investigate what seems to be the main potential problem facing his account, namely that it is too context-relative. I extend Lange's account to include a deeper analysis of the kinds of features that tend to count as salient features. On my reading, there are certain objective explanatory features or patterns (such as symmetry, unity and more) which have the propensity to appear salient to creatures like us in certain contexts. This kind of context-relativity is no problem and should indeed be seen as a virtue of a successful account of mathematical explanation.

In Chapter 5, I present a detailed case study of Galois theory, drawing together insights gained from earlier accounts to present a novel analysis of the case and arguing that attention to such cases is fruitful in our progress towards a philosophical account of mathematical explanation.

1 Chapter 1: Ontic and epistemic accounts of explanation

1.1 Ontic and epistemic aspects of explanation

Historically, philosophers have distinguished between ontic, epistemic and modal conceptions of explanation. For example, Salmon writes:

> 'In its classic form – the inferential version – the epistemic conception takes scientific explanations to be arguments. ... According to the modal conception, scientific explanations do their jobs by showing that what did happen had to happen. ... The ontic conception sees explanations as exhibitions of the ways in which what is to be explained fits into natural patterns or irregularities.' [Salmon 1984: 293]

However, the idea of a modal explanation is a little unclear in the mathematical case where all truths are necessary truths – that is, anything that 'happens', 'had to happen'. Salmon's categories have also been criticised in the scientific case, based on the charge that they cannot account for certain types of explanation in science like mechanistic explanation [Illari 2013]. So I will focus just on the distinction between ontic and epistemic conceptions of explanation.

On a strong reading, the distinction between these two conceptions lies in their differing views on the ontological status of explanations: whether they are representations (e.g. exhibitions or arguments) or things in the world (e.g. causes or dependence relations). But I favour a weaker reading of the distinction, as follows.

An ontic conception of explanation, I suggest, focuses on the link between the explanation and things in the world. For example, does the putative explanation correctly identify the relevant causes of the phenomenon to be explained? If so, the explanation can take the form of an argument or representation, but it derives its explanatory power from identifying causes.

An epistemic conception of explanation, I suggest, focuses on the link between explanation and understanding. For example, does the putative explanation help us to understand why the phenomenon to be explained occurred? If so, the explanation derives its explanatory power from increasing our understanding.

We can go further and suggest that the same putative explanation can have explanatory power along both dimensions, if it both identifies the relevant causes and also describes them in a way that helps us to understand why those causes led to the phenomenon occurring.

For example, Illari presents a picture on which a successful explanation must meet both ontic and epistemic constraints. Illari is interested in the case of mechanistic explanation in science, and proposes that a successful mechanistic explanation must:

- 'Describe the (causal) structure of the world: to be distinctively mechanistic, describe the entities and activities and the organization by which they produce the phenomenon or phenomena.
- Build a model of the activities, entities and their organization that scientists can understand, model, manipulate and communicate, so that it is suitable for the ongoing process of knowledge-gathering in the sciences.

> If this is right, then both epistemic and ontic constraints are essential for mechanistic explanation, but neither is prior. Without the first constraint, we are not explaining the production of a phenomenon by a mechanism; without the second we do not achieve the understanding essential to explanation.' [Illari 2013: 250]

In a more general context, Saatsi argues that focusing on different types of explanatory roles allows us to make fine-grained distinctions within any given account of explanation. Saatsi suggests that:

> '*Thick explanatory role* is played by a fact that bears an ontic relation of explanatory relevance to the explanandum in question.

Thin explanatory role is played by something that allows us to grasp, or (re)present, whatever plays a 'thick' explanatory role.' [Saatsi 2016: 1056]

I think a successful account of mathematical explanation should include both ontic and epistemic components, whether this is by proposing ontic and epistemic constraints on explanation as Illari does, or by allowing for thick and thin explanatory roles as Saatsi does.

To support the inclusion of an epistemic component, note that epistemic considerations were central to many of the motivating quotes I presented in the Introduction. For example, Gowers writes that 'a new proof of a theorem can provide crucial *insights*', Frenkel expresses his discomfort with a non-explanatory proof by writing that 'I still find it *mysterious*', and Atiyah describes his relief when 'five or six years later I *understood* why [the theorem] had to be true' [Gowers and Nielsen 2009: 879, Frenkel 2013: 181, Minio 1984: 17, emphasis added]. Without a link to understanding, a putative account of mathematical explanation will not fit with mathematical practice.

To support the inclusion of an ontic component, note that we want to rule out putative mathematical explanations which we might call 'merely' pedagogical, such as explaining a particular kind of mathematical notation or explaining the content of the theorem. Consider a simple example.

Theorem: For all $n, k \in \mathbb{N}_{>1}$, $k \begin{pmatrix} n \\ k \end{pmatrix} = n \begin{pmatrix} n-1 \\ k-1 \end{pmatrix}$

In order to understand the content of the theorem, we need to understand the notation used, in particular that $n! = n \times n-1 \times n-2 \times ... \times 2 \times 1$ and $\begin{pmatrix} n \\ k \end{pmatrix} = \frac{n!}{(n-k)!k!}$. Once we get to grips with the notation, we can see that $\begin{pmatrix} n-1 \\ k-1 \end{pmatrix} = \frac{(n-1)!}{(n-k)!(k-1)!}$. We now see that theorem tells us that $\frac{kn!}{k!(n-k)!} = \frac{n(n-1)!}{(n-k)!(k-1)!}$. Assuming we are familiar with the concepts of multiplication and division, we might feel that we now have a better understanding of the content of the theorem.

A better understanding still can be gained by introducing combinatorial concepts. In particular, $\binom{n}{k}$ is equal to the number of ways to choose k objects from n objects (without replacement, and where the order doesn't matter). This may not seem obvious at first. Why, we might ask, is the number of ways to choose k objects from n objects without replacement equal to $\frac{n!}{(n-k)!k!}$?

Consider the case where we are choosing two balls from a box containing five balls. There are five ways to choose the first ball, and four ways to choose the second ball. Imagine we have numbered the balls from 1 to 5. If we first choose ball 1, that gives us four possible combinations: $(1,2),(1,3),(1,4),(1,5)$. If we first choose ball 2, that gives us four combinations: $(2,1),(2,3),(2,4),(2,5)$, and so on. So there are $4 \times 5 = 20$ possible combinations. In general, to choose k objects from n objects, there will be $n \times (n-1) \times (n-2) \times ... \times (n-(k-1))$ possible combinations. Notice that $n \times (n-1) \times (n-2) \times ... \times (n-(k-1)) = \frac{n!}{(n-k)!}$.

However, we specified that the order didn't matter, so that $(1,2)$ counts as the same combination as $(2,1)$. That is, in choosing two balls from five balls we have in fact counted each relevant combination twice. The total number of relevant combinations is therefore $\frac{5 \times 4}{2} = 10$. And in general there are $k!$ ways to arrange a sequence with k elements, so if the order doesn't matter we get $\binom{n}{k} = \frac{n!}{(n-k)!k!}$.

We can now see that the theorem tells us that k times the number of ways of choosing k objects from n objects is equal to n times the number of ways of choosing $k-1$ objects from $n-1$ objects. I suggest that the material presented so far helps us to understand the *content* of the theorem .

However, this is not sufficient to count as a mathematical explanation in the sense I will be interested in throughout this thesis: I am interested in arguments that increase our understanding of why a certain mathematical theorem or fact is *true*. In mathematics, such arguments take the form of proofs, and so I will be focusing on proofs, or proof sketches. I do not mean to claim that only proofs can be explanations. But since nearly all of the philosophers writing on this topic focus on proofs, I am in good company.

In the next four sections, I consider some existing philosophical accounts of explanation and assess the prospects for extending these into an account of intra-mathematical explanation.

1.2 Grounding

In this section, I consider a suggestion made by Kit Fine (among others) that grounding relations between facts play an important explanatory role. This suggestion has gained in popularity in the recent philosophical literature, and it might seem a promising type of account to extend to the intra-mathematical case since the grounding relation is not restricted to causally active entities.

Fine 'recommend[s] that a statement of ground be cast in the following "canonical" form:

> Its being the case that S consists in nothing more than its being the case that T, U, ...

where S, T, U, ... are particular sentences. As particular examples of such statements, we have:

> Its being the case that the couple Jack and Jill is married consists in nothing more than its being the case that Jack is married to Jill.' [Fine 2001: 15]

Furthermore, Fine connects the grounding relation to explanation as follows:

> 'We take *ground* to be an explanatory relation: if the truth that P is grounded in other truths, then they *account* for its truth; P's being the case holds *in virtue* of the other truths' being the case. There are, of course, many other explanatory connections among

> truths. But the relation of ground is distinguished from them by being the tightest such connection.' [ibid.]

My main concern about this claim is that even if the grounding relation is an explanatory relation in some sense – grounding has for example been described as 'metaphysical explanation'[2] – it does not connect very well to the kinds of concerns we often have when we make an explanatory demand.

For example, a conjunction is often taken to be grounded in its conjuncts. Fine's suggestion is made in the context of the debate between realists and antirealists; on Fine's account, both sides in this debate are supposed to be able to agree that 'the fact that P & Q consists in the fact that P and the fact that Q' [ibid.].

But it does not seem plausible to me that the conjuncts always explain the conjunction. Consider for example the finite conjunctive statement: 'Everyone in the seminar room is a philosopher' where 'everyone' ranges over a finite set with 26 members: {Anna, Bob, Caroline, Dave, ... , Zelda}. We might ask 'Why is everyone in the seminar room a philosopher?'.

The answer 'Anna is in the seminar room and she is a philosopher, Bob is in the seminar room and he is a philosopher, ...' does not seem to provide an explanatory answer to this question, in contrast to the answer 'Because the seminar is being held at a philosophy conference'.

Similarly, in the mathematical case there are plenty of examples where tallying up instances does not seem to add to a proof's explanatory value. Lange suggests a nice case, considering two proofs by Dreyfus and Eisenberg of the following claim:

> 'Suppose you write down all of the whole numbers from 1 to 99,999. How many times would you write down the digit 7? The answer turns out to be 50,000 times.' [Lange 2017: 267, Lange ms: 1][3]

[2]For discussion here see e.g. [Thompson 2016].

[3]Phrased in this way, the example looks like it concerns a physical phenomenon. But we can simply rephrase it so that we are interested in the list of numbers, whether or not it is written down.

The first proof tallies all of the appearances of 7 in the list of numbers:

> ‘The digit 7 appears once between 1 and 10, once between 11 and 20, in fact once in every “regular” 10-plet of numbers; here “regular” means that there are no digits 7 in the tens or higher decimal places. Between 61 and 70, there are two digits 7; between 71 and 80, there are 10; collecting all of these, one concludes that the digit 7 appears 20 times between 1 and 100, and thus 20 times in every “regular” 100-plet. In the 100-plet from 601 to 700, there is an additional 7, i.e., the digit 7 occurs 21 times, and in the following 100-plet there are 99 additional 7’s, yielding all together 300 7’s between 1 and 1000. Proceeding in a similar way, one finds that the digit 7 appears 4,000 times between 1 and 10,000, and it appears 50,000 times between 1 and 100,000.’ [Dreyfus and Eisenberg 1986: 3]

The second proof appeals to a symmetry in the order of digits in the list:

> ‘Include 0 among the numbers under consideration – this will not change the number of times that the digit 7 appears. Suppose all of the whole numbers from 0 to 99,999 (100,000 of them) are written down with five digits each, e.g., 1306 is written as 01306. All possible five-digit combinations are now written down, once each. Because every digit will take every position equally often, every digit must occur the same number of times overall. Since there are 100,000 numbers with five digits each – that is, 500,000 digits – each of the 10 digits appears 50,000 times. That is, 100,000 x 5 / 10 = 50,000’. [Dreyfus and Eisenberg 1986: 3]

I agree with Lange that the second proof is more explanatory, even though the first proof cites the grounds of the result. I will discuss this particular example in more depth in Chapter 4.

Now, of course, Fine and I are essentially talking at cross-purposes here: Fine is interested in metaphysical questions, while I want to pursue an account of mathematical explanation that accounts for the epistemic role played by an explanatory proof in helping us to see why the theorem proved is true.

My point here is that it seems unlikely that a grounding account can help me in this project. As Lange suggests in recent work, it seems likely that grounding and explanation come apart in mathematics [Lange ms]. To motivate this further, consider the following issue.

In the mathematical domain, we might take the ultimate 'grounds' of a mathematical fact or theorem to be the mathematical axioms. However, the axioms are rarely cited directly in any actual presentation of a proof. Furthermore, choice of axioms is not always motivated only by what is explanatorily basic. Sometimes convenience for use in proofs of theorem is a motivating factor; and various preferences may also enter in, such as a preference for first order logic, for brevity, for ease of proving facts about the axiom system (such as mutual independence of axioms, consistency relative to other systems) and so on.

As an example of an axiom, consider the well-ordering principle about natural numbers, which describes a structural fact about this domain: namely that the ordering is strictly linear with no non-empty set of ever-decreasing natural numbers. Consider the following theorem:

> **Theorem:** $\sqrt{2}$ is irrational.
>
> **Proof using the well-ordering principle:**
>
> We start by assuming that $\sqrt{2}$ is rational, i.e. that $\sqrt{2} = a/b$ for positive integers a and b. Consider the set $S = \{k\sqrt{2} : k$ and $k\sqrt{2}$ are positive integers$\}$. Since $a = b\sqrt{2}$ is an integer, it is a member of set S. Since S is a non-empty set of positive integers, it must have a minimum element by the well-ordering principle. Assume that the smallest element is s, where $s = t\sqrt{2}$ for some positive integer t.
>
> Since t and $t\sqrt{2}$ are positive integers and $1 < \sqrt{2} < 2$, $2t - t\sqrt{2}$ is also a positive integer. But $2t - t\sqrt{2} = \sqrt{2}t(\sqrt{2} - 1) = s(\sqrt{2} - 1)$,

and since $0 < \sqrt{2} - 1 < 1$, we have $s(\sqrt{2} - 1) < s$. So we have found a smaller positive integer in the set S, and by contradiction $\sqrt{2}$ must be irrational.

This proof explicitly cites the well-ordering principle, but it is a proof by contradiction and such proofs are often thought to be non-explanatory. Lange quotes Pierre Nicole and Antoine Arnauld in the Port-Royal Logic of 1662 on this point:

> 'such Demonstrations constrain us indeed to give our Consent, but no way clear our Understandings, which ought to be the principal End of Sciences: For our Understanding is not satisfied if it does not know not only that a thing is, but why it is? which cannot be obtain'd by a Demonstration reducing to Impossibility' [Nicole and Arnauld 1717, quoted in Lange 2017: 234]

I actually think that a proof by contradiction can be explanatory, as does Lange [ibid.]. We will see some further examples in Chapter 2. But, even if the well-ordering proof above is explanatory, I suggest it is the specific way in which the axiom is cited that is relevant for the proof's explanatory value. We could prove the same result by citing the principle of induction instead:

Proof using strong induction:[4]

We need to show that $\sqrt{2} \neq \frac{a}{b}$ for all positive integers a and b. That is, $2 \neq \frac{a^2}{b^2}$ for all positive integers a and b.

We can see directly that $2 \neq \frac{1}{b^2}$ for all positive integers b (the base case, $a = 1$). Suppose that $2 \neq \frac{a^2}{b^2}$ for all positive integers b and all a such that $1 \leq a \leq k$ (the inductive hypothesis).

Suppose $2 = \frac{(k+1)^2}{b^2}$ for some positive integer b (i.e. the property fails to holds at $a = k + 1$). Then we have $2b^2 = (k + 1)^2$, so we have $b < k + 1$, and $(k + 1)^2$ is even. Hence $(k + 1)$ is even, i.e. $(k + 1) = 2c$ for some integer c, and we have $2b^2 = (2c)^2$,

[4] I give a more detailed exposition of inductive proof in Chapter 2.

or $\frac{b^2}{c^2} = 2$. But since $b < k+1$, this contradicts the inductive hypothesis. Hence $2 \neq \frac{(k+1)^2}{b^2}$ for any positive integer b.

Therefore, by the principle of induction, $\sqrt{2} \neq \frac{a}{b}$ for all positive integers a and b, so $\sqrt{2}$ is irrational.

I suggest the well-ordering proof is more explanatory because it draws a direct connection between the well-ordering principle and the structure of the set S, while the inductive proof cites the induction principle only to indicate that the induction schema is instantiated in the proof.

Of course, this suggestion needs further support. But I hope I have given some motivation for the thought that if axioms or grounds are explanatorily relevant, it is because of the way they feature in a given example. Baker puts the point succinctly:

> 'If a mathematical claim is provable then it is provable from the axioms of the theory in which it is embedded. So why not cite the axioms as providing an explanation of *any* provable claim? Here the axioms are analogous to the boundary conditions right after the Big Bang, and the same point about indirectness applies. Tracing inferential paths back to axioms is no more explanatory *per se* than tracing causal chains back to the Big Bang. Sometimes axioms are explanatory and sometimes they are not, but this depends on the nature of the proof and not on its bare existence.' [Baker 2009: 149]

I will now move on to consider a counterfactual account of explanation proposed by Alexander Reutlinger.

1.3 Counterfactual

Counterfactual accounts of causal explanation are inspired by 'what-if-things-had-been-different' questions: for example, a counterfactual account holds roughly that event A causes event B if it is true that if A had not happened, B would not have happened either; and if it is true that if A had happened in similar circumstances to the actual circumstances, then B would also have happened. We describe B as counterfactually depending on A, and proponents of this type of account often link counterfactual dependence to explanatory value. See for example Woodward [2003] for one of the most influential counterfactual accounts of causal explanation.

With their focus on causation, counterfactual accounts of explanation do not seem an obvious candidate to extend to the mathematical case. However, Reutlinger has recently proposed a version of a counterfactual account that he claims can handle both causal and non-causal explanation, and which I will consider here.

According to Reutlinger's counterfactual account of explanation, explanations have a two-part structure. First, the explanation includes a statement, E, about the phenomenon to be explained. Second, the explanation includes an explanans, which is formed of generalizations G_1 to G_m and auxiliary statements S_1 to S_n. Given this structure, the relationship between these parts is explanatory if and only if the following three conditions are all satisfied.

1. '*Veridicality condition*: $G_1, ..., G_m$, $S_1, ..., S_n$, and E are (approximately) true.

2. *Implication condition*: $G_1, ..., G_m$ and $S_1, ..., S_n$ logically entail statement E or logically entail a conditional probability $P(E|S_1...S_n)$.

3. *Dependency condition*: $G_1, ..., G_m$ support at least one counterfactual of the following form: had $S_1, ..., S_n$ been different than they actually are (in at least one way deemed possible in the light of the generalizations), then E or the conditional probability of E would have been different as well.' [Reutlinger 2016: 737].

A distinction between causal and noncausal explanations is then drawn within this framework, namely: 'noncausal explanations are explanatory by virtue of exhibiting noncausal counterfactual dependencies; causal explanations are explanatory by virtue of exhibiting causal counterfactual dependencies' [ibid.]. A positive account of causal and noncausal dependencies is given in Reutlinger [2013].

My focus here is on the details of the noncausal case on Reutlinger's account, since intra-mathematical explanation must presumably fall on this side of the distinction. Reutlinger considers a popular putative case of noncausal explanation that involves mathematics: the bridges of Königsberg, an interesting historical example.

In 1735, the city now called Kaliningrad was known as Königsberg. The city was divided into four parts, which were connected by seven bridges. No one had ever managed to take a walk that crossed each bridge exactly once, something which called for an explanation since apparently the city residents had spent many Sunday afternoons searching for such a route. Indeed, the mayor of Danzig (now Gdansk) wrote to the mathematician Leonhard Euler asking for an explanation. At first, Euler felt that the problem was trivial, responding to the mayor that 'this type of solution bears little relationship to mathematics, and I do not understand why you expect a mathematician to produce it, rather than anyone else, for the solution is based on reason alone, and its discovery does not depend on any mathematical principle' [Hopkins and Taylor 2004: 201].

Later, however, Euler wrote to fellow mathematician Giovanni Marinoni that 'This question is so banal, but seemed to me worthy of attention in that [neither] geometry, nor algebra, nor even the art of counting was sufficient to solve it' [Hopkins and Taylor 2004: 202].

Euler developed a new kind of 'Geometry of Position' to tackle the problem, which formed the beginnings of the rich mathematical field now known as graph theory. Using his new theory, Euler was able to explain the phenomenon in need of explanation: namely that no one had ever managed to take a walk that crossed each of the seven bridges exactly once.

Modern accounts of the theory are slightly different from Euler's original paper, but according to Reutlinger, Euler's explanation runs as follows. First, we can represent the layout of Königsberg as a graph, G, with four nodes (representing the four parts of the city) A, B, C and D, and seven edges (representing the bridges). See Figure 1 below.[5]

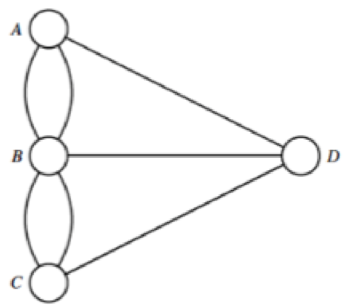


Figure 1: Graph G

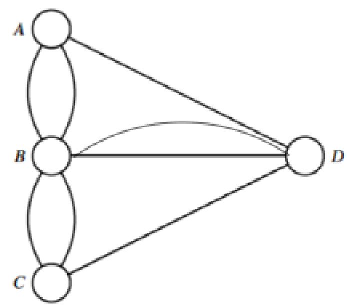


Figure 2: Graph G'

We then define an Euler path as a path through G that includes each edge in G exactly once. (Note that this does not mean we need to end up in the same place we started at: that extra condition would require the path to be an Eulerian cycle). The explanandum phenomenon can then be translated using this new conception to get statement E: 'No one had ever managed to take a walk along an Euler path around Königsberg'.

On Reutlinger's account, the explanation has two parts:

1. '*Euler's theorem*: there is an Euler path through a graph G iff G is an Eulerian graph. Euler proved that a graph G is Eularian [sic] iff (i) all the nodes in G are connected to an even number of edges, or (ii) exactly two nodes in G (one of which we take as our starting point) are connected to an odd number of edges.

2. *Contingent fact*: The actual bridges and parts of Königsberg are not isomorphic to an Eulerian graph, because conditions i and ii in the definition of an Eulerian graph are not satisfied: no part of town (corresponding to the nodes) is connected to an even number of bridges (corresponding to the edges), violating condition i; and more than two parts of town

[5]As presented e.g. in [Räz 2018: 344]. Reutlinger does not supply a diagram.

(corresponding to the nodes) are connected to an odd number of bridges (corresponding to the edges), violating condition ii. Königsberg could have been isomorphic to an Eulerian graph in 1736, but as a matter of contingent fact it was not.' [Reutlinger 2016: 740]

Reutlinger claims that the example meets his three conditions as follows:

1. 'The *veridicality condition* holds because (a) Euler's theorem, (b) the statement about the 'contingent fact' that each part of Königsberg is actually connected to an odd number of bridges, and (c) the explanandum statement are all true.

2. The *implication condition* is met, since Euler's theorem and the statement about the 'contingent fact' entail the explanandum statement.

3. The *dependency condition* is satisfied, because Euler's theorem supports the counterfactual [*c*]: "if all parts of Königsberg were connected to an even number of bridges, or if exactly two parts of two were connected to an odd number of bridges, then people would not have failed to cross all of the bridges exactly once."' [Reutlinger 2016: 740]

In this case, we see that the relevant generalization is supposed to be Euler's theorem, and the relevant auxiliary statement is the statement about the contingent fact that each part of Königsberg is connected to an odd number of bridges. The entailment in the implication condition is valid.

However, I suggest that contrary to Reutlinger's claim, the dependency condition is not quite satisfied. After all, it could be the case that even if there were an Euler path through Königsberg, no one had managed to find it. Here we can make use of the probability clause mentioned by Reutlinger, and claim that Euler's theorem supports the counterfactual c': "if all parts of Königsberg were connected to an even number of bridges, or if exactly two parts of town were connected to an odd number of bridges, then *it is likely that* people would not have failed to cross all of the bridges exactly once."[6]

[6]Though note that this looks like an unconditional rather than conditional probability.

Now, Reutlinger notes that the counterfactual dependency condition of his account does not rely on an interventionist understanding of counterfactuals [Reutlinger 2016: 738]. This means we do not need to consider possible interventions we might carry out to alter the bridges of Königsberg, but rather need only consider an alternative situation where the arrangement of bridges already happens to be different.

Reutlinger emphasises this point in order to sidestep the debate about Woodward's influential account, which explicitly involves interventionist counterfactuals. Moreover, the point also helps the prospects for extending Reutlinger's account to the intra-mathematical case: it seems clear that an interventionist account of counterfactuals would not be particularly helpful with regard to mathematical entities, which (if they exist) are abstract entities on which we presumably cannot intervene.

Now, if the relevant counterfactuals are not given an interventionist reading on Reutlinger's account, how should we understand them? At first glance this looks reasonably clear. We just need to imagine a Königsberg where, say, an extra bridge is present between the parts of town represented by nodes B and D in Figure 1. In this situation, the antecedent of counterfactual c' is fulfilled, since exactly two parts of Königsberg (those represented by nodes A and C) would then be connected to an odd number of bridges. We can easily find an Euler path on the new corresponding graph G' (presented in Figure 2), so it is likely that the residents of Königsberg would have found the corresponding route around town which crosses each bridge exactly once.[7] Hence the consequent of the counterfactual c' is fulfilled as well, and Reutlinger's dependency condition is satisfied. Presumably Reutlinger has in mind something like this.

My main concern about extending Reutlinger's account to cover cases of intra-mathematical explanation is that it is less clear how the relevant coun-

[7] Of course, the residents would not be reading the route off the graph, since the modern graph-theoretic representation was not available at the time. Nevertheless, presumably the residents had some systematic way of testing routes (perhaps by brute force – attempting all possibilities), or they would not have found the explanandum phenomenon interesting to begin with. Hence we can surmise that it is likely they would find the relevant route in the modified Königsberg.

terfactual works in the intra-mathematical case, because mathematical statements, if true, are necessarily and not contingently true. Additionally, perhaps the individuation conditions for mathematical entities are stricter than in many physical cases.

Suppose we translate the explanandum phenomenon into a mathematical claim. Rather than 'no one had ever managed to take a walk that crossed each of the seven bridges exactly once', the new explanandum is 'there is no Euler path through graph G'. The relevant generalisation would still be Euler's theorem, and the auxiliary statement would be that each node of the graph is connected to an odd number of bridges. The explanandum is entailed by the theorem and auxiliary statement, so the question is how we can understand a situation where the auxiliary statement is false.

As before, we can draw a different graph, G', with an extra edge between nodes B and D, just as we did when we imagined Königsberg with an extra bridge. It is true that in graph G', there is a version of the original explanandum statement: there is an Euler path through graph G'. But the original explanandum statement was about graph G, not graph G'. It is not immediately clear how to link these two statements together in order to give a good account of the intra-mathematical explanation.

The problem is that in the physical case, the individuation conditions may allow us to change the situation while talking about the same object. For example, we might be happy to consider Königsberg-with-an-extra-bridge as just another possible version of Königsberg. Euler's theorem helps us to understand why, if the layout of Königsberg were slightly different, the explanandum phenomenon would not occur. In turn this helps us to understand that the actual layout of Königsberg explains why the explanandum actually did occur.

In the intra-mathematical case, we can evaluate whether the explanandum phenomenon holds in a different graph, G'. But what makes us sure that G' is relevantly similar to the original graph, G? The graphs are not isomorphic, since one is Eulerian and one is non-Eulerian. So it's not clear that there is a mathematical sense in which they are similar.

The graphs both count as representations of different versions of the same physical entity (Königsberg). But in an intra-mathematical explanation we cannot use reference to a physical entity as part of the explanation. So to extend Reutlinger's counterfactual account we need some other way of linking the two graphs.

To describe the problem more clearly: in the physical case, the two versions of Königsberg, K and K', are related to graphs G and G'. Euler's theorem tells us something about the graphs G and G', which we can translate back to K and K'. K and K' are sufficiently similar for us to take both results to be telling us something about Königsberg. In particular, the result about K' helps us to understand something about the actual situation K.

In the intra-mathematical case, Euler's theorem tells us something about G and G'. These graphs are not the same and are not mathematically similar (they are non-isomorphic). How does a result about G' help us to understand a result about G?

I think the answer to this question is unclear on Reutlinger's account, so let us now step back from the details of his account, and think about the spirit behind the counterfactual account. The central idea is that explanations work by identifying what we might call a 'difference-maker', that is, a relevant property which G has and G' does not, and which makes the result go through for G and not for G'. Mark Steiner presents an account of mathematical explanation inspired by this central idea [Steiner 1978], and I will analyse Steiner's proposal in depth in Chapter 3.

For now, I move on to look at two epistemically-focused accounts of explanation.

1.4 Unification

Philip Kitcher presents one of the few accounts of explanation that is explicitly intended to cover the intra-mathematical as well as the scientific case. Kitcher quotes a remark by Feigl as his inspiration for taking explanation to be concerned primarily with *unification*:

> 'The aim of scientific explanation throughout the ages has been unification, i.e., the comprehending of a maximum of facts and regularities in terms of a minimum of theoretical concepts and assumptions.' [Feigl 1970: 12, quoted in Kitcher 1981: 508]

The idea here is that our search for explanation is the search for an 'explanatory store' of theoretical concepts that enables us to explain as many of the phenomena we seek to understand as possible. As Kitcher puts it, the explanatory store over a set of beliefs K is 'the set of derivations that makes the best tradeoff between minimizing the number of patterns of derivation employed and maximizing the number of conclusions generated' [Kitcher 1989: 432].

Kitcher's account is a global rather than local account, in the sense that a putative explanation is explanatory in virtue of belonging to the global explanatory store. The account can be formalized as presented by Hafner and Mancosu [2008]:

> 'Let us start with a set K of beliefs assumed consistent and deductively closed (informally one can think of this as a set of statements endorsed by an ideal scientific community at a specific moment in time) ... A *systematization* of K is any set of arguments which derive some sentences in K from other sentences in K. The *explanatory store* over K, $E(K)$, is the best systematization of K... Corresponding to different systematizations we have different degrees of unification; the highest degree is that given by $E(K)$.' [Hafner and Mancosu 2008: 153]

The members of the explanatory store are derivations or argument patterns. Again, Hafner and Mancosu present a formalization of this idea:

> 'Let us begin with the notion of *schematic sentence*. This is an expression obtained by replacing some or all of the non-logical expressions in a sentence by dummy letters. A set of *filling instructions* tells us how the dummy letters in a schematic sentence are to be replaced. A *schematic argument* is a sequence of schematic

sentences. A *classification* for a schematic argument is a set of sentences which tells us exactly what role each sentence in a schematic argument is playing, e.g. whether it is a premise, which sentences are inferred from which and according to what rules, etc. A *general argument pattern* $< s, f, c >$ is a triple consisting of a schematic argument s, a set f of filling instructions and a classification c for s.' [Hafner and Mancosu 2008: 155]

Let us now consider a mathematical example.

Theorem: For all n in $\mathbb{N}$, $1 + 2 + ... + n = n(n + 1)/2$

Proof:

For $n = 1 : 1 = (1 \times 2)/2$

If for $n = k$, $1 + 2 + ... + k = k(k + 1)/2$

then for $n = k + 1$, $(1 + 2 + ... + k) + (k + 1) = (\frac{k(k+1)}{2}) + (k + 1)$
$= \frac{k^2+3k+2}{2} = \frac{(k+1)(k+2)}{2}$.

We can give the argument pattern in the format required.

S1. Theorem: (For all n in $\mathbb{N}$) The sum from 1 to n is equal to $n(n + 1)/2$.

S2. For $n = 1$, the theorem is true.

S3. If the theorem is true for $n = k$, then the theorem is true for $n = k + 1$.

S4. Hence, the theorem holds for all n in $\mathbb{N}$.

SA: Schematic argument

S1. Theorem: (For all n in $\mathbb{N}$) $P(n)$

S2. $P(1)$

S3. If $P(k)$, then $P(k + 1)$

S4. Hence, for all n in $\mathbb{N}$, $P(n)$.

F: Filling instructions. Replace $P(i)$ by a sentence of the form 'Property P holds for $n = i$'.

C: Classification. S1 is a statement of theorem, S2 and S3 are premises, S4 follows from S2 and S3 by the principle of induction.

The triple forms an argument pattern $< SA, F, C >$ that can generate a lot of conclusions – in fact, infinitely many. Does this mean inductive proofs are likely to belong to the explanatory store over our set of beliefs about, say, the natural numbers $\mathbb{N}$?

Kitcher writes: 'Suppose I prove a theorem by induction ... It would seem hard to deny that this is a genuine proof. ... Further ... the proof explains the theorem' [Kitcher 1975: 265]. So it seems that Kitcher would agree that inductive proofs belong to the explanatory store (or strictly speaking, that the argument pattern of inductive proofs belongs to the explanatory store).

This is a worry for Kitcher's account, because I will argue in Chapter 2 that we should not assume that all inductive proofs have the same explanatory value. It seems very plausible that some inductive proofs are explanatory and some are not, and Kitcher's account does not allow for this possibility.

Furthermore, the principle of induction is logically equivalent to the well-ordering principle on $\mathbb{N}$, in the sense that any proof using the principle of induction can be transformed into a proof using the well-ordering principle, and vice versa. That is, argument patterns using the principle of induction and argument patterns using the well-ordering principles prove exactly the same number of conclusions. We don't want to add both argument patterns, since we are trying to minimise the total number of argument patterns; one of the principles, therefore, is presumably redundant and not involved in the explanatory store. But Kitcher's account does not give us a way to choose between the two.

Now, Kitcher could simply point to the fact that the principles are logically equivalent and say that we do not need to reject either one of them since they are essentially the 'same' principle in this sense. But this doesn't allow for the fact that we do sometimes perceive a difference in explanatoriness between two proofs of the same result, where one proof uses the principle of induction and one uses the well-ordering principle. We have already seen a putative example of this in Section 1.2 on grounding accounts of explanation.

Consider another persuasive example here.

> **Theorem:** There are infinitely many prime numbers.
>
> **Well-ordering proof** (due to Euclid)[8]
>
> Take any finite collection of prime numbers. According to the well-ordering principle, we can write them down in an ordered list: $p_1, p_2, ..., p_k$ for some natural number k. Now let $N = 1+p_1 \cdot p_2 \cdot ... \cdot p_k$. Either N is prime, or it has a prime factor. None of the primes $p_1, p_2, ..., p_k$ divide N, since the remainder will be 1 in each case. So if N has a prime factor, its prime factor p_N does not appear on the list $p_1, p_2, ..., p_k$. In either case, therefore, we have found a new prime number: either N or p_N. So for any finite collection of prime numbers, we can find a new prime number not already in the collection. Hence, there are infinitely many prime numbers.

We can prove the same result using the inductive argument pattern described above:

> **Inductive proof**
>
> We rephrase the statement of the theorem as follows: For any natural number n there is a prime greater than n.
>
> Base case: For $k = 1$, we have $2 > 1$ and 2 is prime.

[8]This proof is often given as a proof by contradiction, but this is not required and Euclid's own version was direct. See [Siegmund-Schultze 2014] for an interesting historical discussion.

Inductive step: Suppose for some $k > 1$ that there is a prime greater than k. We show that there is also a prime greater than $k + 1$, as follows.

Note that for any natural number $j > 2$, there is a prime number between j and $j!$ (this result is often proved as a lemma on the way to proving Bertrand's postulate, also known as the Bertrand-Chebyshev theorem[9]).

The number $N = j! - 1$ is greater than 1, and hence N has a prime factor, p. W must have $p > j$ because otherwise p would be a prime factor of 1, which is impossible (since 1 is not prime). Additionally, $p \leq N$, so we have $j < p \leq N = j! - 1 < j!$, that is, $j < p \leq j!$, as required.

Now, note that since $k > 1$, we have $k + 1 > 2$. So there is a prime number between $k + 1$ and $(k + 1)!$, i.e. a prime number greater than $k + 1$, as required.[10]

I suggest that the well-ordering proof is more explanatory. Again, this needs further support, but I hope the example provides some motivation for the thought that the explanatory value of an argument pattern does not reside solely in the number of conclusions it affords, and that the non-logical content of the argument is explanatorily relevant.

A similar point is made by Hafner and Mancosu when they consider cases where we seem to need an infinite number of argument patterns. They suggest that Kitcher's account cannot explain a 'controversy over the use of transcendental methods in real algebraic geometry [where] the point at issue concerns *qualitative* differences in the proof methods. ... In general there is more to explanation than unification in Kitcher's sense, a more fine-grained analysis of different types of unification seems to be needed.' [Hafner and Mancosu 2008:

[9]See e.g. http://mathworld.wolfram.com/BertrandsPostulate.html

[10]We don't actually make use of the fact that there is a prime greater than k. But we have still proved the truth of the conditional 'If the theorem is true for $n = k$, then the theorem is true for $n = k + 1$'.

170-1]

I agree, and so I will now move on to the final type of account to be considered in this chapter: a why-question account of explanation.

1.5 Why-questions

A natural thought about explanation is that when we search for an explanation, we are seeking to answer a particular why-question. For example, we might ask 'Why do plants need sunlight to grow?', or 'Why is theorem T true?'. Now, natural language why-questions can be ambiguous, and their intended content depends on emphasis and context. For example, when we ask 'Why did Adam eat the apple?' we might be wondering why Adam ate the *apple* (as opposed to some other fruit); why Adam *ate* the apple (as opposed to throwing it away); or why *Adam* ate the apple (as opposed to somebody else eating it). [van Fraassen 1980: 127]

This ambiguity motivates Bas van Fraassen's formal account of why-questions, according to which why-questions consist of:

1. A topic, P_k;
2. A contrast class, $X = P_1, P_2, \ldots, P_n$;
3. A relevance relation, R, which constrains admissible answers. [van Fraassen 1980: 143]

On this account, a why-question takes the form 'Why is P_k true, rather than some other possibility P_i?'. The formal account of contrast classes helps to specify which of these logical why-questions is being asked on a particular occasion. We then also need an account of what counts as an *answer* to a why-question. Van Fraassen proposes that:

> 'B is a direct answer to a question $Q = < P_k, X, R >$ exactly if there is some proposition A such that A bears relation R to $< P_k, X >$ and B is the proposition which is true exactly if (P_k; and for all $i \neq k$, not P_i; and A) is true.' [van Fraassen 1980: 144]

B takes the form 'P_k, in contrast to the rest of X, because A', where P_k is true, the other P_i's are false, and there is some relevant A which is true.

For example, take the following natural language why-question, Q: 'Why are the philosophers in Lecture Room 2?'. We can analyse the situation according to van Fraassen's account as follows.

- P_k: the philosophers are in Lecture Room 2
- X: {the philosophers are in Lecture Room 1, the philosophers are in Lecture Room 2}
- A: Lecture Room 1 is booked for an HR meeting
- B: 'The philosophers are in Lecture Room 2, rather than Lecture Room 1, because Lecture Room 1 is booked for an HR meeting.'

Or

- P_k: the philosophers are in Lecture Room 2
- X: {the philosophers are in Lecture Room 2, the HR delegates are in Lecture Room 2}
- A: there are too many HR delegates to fit in Lecture Room 2
- B: 'The philosophers are in Lecture Room 2, rather than the HR delegates being in Lecture Room 2, because there are too many HR delegates to fit in Lecture Room 2.'

The relevance relation, R, is not explicitly given here, and indeed van Fraassen sees the relevance relation as a mostly schematic part of his account, leaving us to fill in the details in each case. He describes explanatory relevance as 'the respect-in-which a reason is requested' [van Fraassen 1980: 142]. For example, when we ask 'Why does the blood circulate through the body?', we might be seeking a mechanical answer in terms of the heart's pumping action, or instead we might want a functional answer in terms of circulating oxygen to the tissues.

In our example, the relevance relation might constrain admissible answers to those linked to the university's room booking system, say. So the following answer will not count as acceptable, because there is no link between Kiwi fruit and the venue of the philosophy lecture:

- P_k: the philosophers are in Lecture Room 2
- X: {the philosophers are in Lecture Room 1, the philosophers are in Lecture Room 2}
- A: Kiwi fruit are also known as Chinese gooseberries
- B: 'The philosophers are in Lecture Room 2, rather than Lecture Room 1, because Kiwi fruit are also known as Chinese gooseberries.'

In the faulty answer, A is true but it doesn't stand in the required relevance relation to $< P_k, X >$.

Now, let us try to apply this formal account to a mathematical example. Take any of the theorem-proof pairs we have already considered. For example:

> **Theorem** (call it T): For all n in $\mathbb{N}$, $1 + 2 + ... + n = n(n + 1)/2$
>
> **Proof:**
>
> For $n = 1 : 1 = (1 \times 2)/2$
>
> If for $n = k$, $1 + 2 + ... + k = k(k + 1)/2$
>
> then for $n = k + 1$, $(1 + 2 + ... + k) + (k + 1) = (\frac{k(k+1)}{2}) + (k + 1)$
> $= \frac{k^2+3k+2}{2} = \frac{(k+1)(k+2)}{2}$.

A mathematician might ask 'Why is the theorem true?'; suppose that she is interested in why Theorem 2 is true, rather than false. On van Fraassen's account we might present the situation as follows:

- P_k: Theorem T is true
- X: {Theorem T is true, Theorem T is false}
- A: The proof of T is valid and sound

- B: 'T is true, rather than false, because the proof of T is valid and sound.'

Here P_k is true, the other P_i's are false, and A is true. It looks like A is relevant in the mathematical context in which the why-question is asked: after all, A describes mathematical conditions that are sufficient for the truth of P_k.

Note that the details of the proof are irrelevant here: any proof of Theorem T which is valid and sound could be inserted instead. This is a worry if our aim is to draw a distinction between explanatory and non-explanatory proofs. As far as van Fraassen is concerned, of course, we could simply deny that this should be our aim in the context described; the mathematician simply asks 'Why is Theorem T true, rather than false?', and in this context perhaps any proof would indeed suffice.

In this thesis, however, I am interested in a potential distinction between explanatory and non-explanatory proofs, so let us now examine whether van Fraassen's account can help us in this project.

Van Fraassen develops the notion of *telling* answers as a way to distinguish between the explanatory value of multiple potential answers to a given why-question.[11] In a telling answer, A must:

1. Be probable in light of our background knowledge;

2. Probabilistically favour the topic [P_k] over other members of the contrast-class relative to our background knowledge;

3. Do better than other potential answers in these two respects. [van Fraassen 1980: 146]

Unfortunately, this won't help us in our mathematical example, because P_k is not just contingently true (with probability < 1), but necessarily true (with

[11] Van Fraassen expresses less confidence in this part of the account, and it has met with a number of objections in the case of scientific explanation (for example, cases where the explanatory answer does not probabilistically favour the topic B). I will put these worries aside and focus only on the intra-mathematical case here.

probability 1), while the other element of X is necessarily false (with probability 0). Again, therefore, any other proof of Theorem T could play the same role.

Sandborg points out that this problem will arise in any case where the members of the contrast class are mutually exclusive. For example, suppose we ask 'why does $1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + ...$ converge to $\frac{\pi}{4}$, rather than some other real number?' [Sandborg 1998: 613]. Any proof of this fact will count as a telling answer, and so, again, on van Fraassen's account 'a proof must either be explanatory or not; there is no middle ground' [ibid.].

Since the notion of telling answers doesn't help us in the mathematical case, we might try instead to lean on the relevance relation as a way to distinguish explanatory from non-explanatory proofs. Indeed, as Sandborg puts it, '*only* the relevance relation can distinguish explanatory from non-explanatory answers to exclusive-contrast why-questions, since it is the only evaluative component of van Fraassen's theory not left trivialized' [Sandborg 1998: 613, emphasis added].

Now, a number theorist might seek a number theoretic answer to explain why a certain theorem is true, for example. The relevance relation might therefore specify that only a number theoretic proof would be relevant in this context, while in another context where a mathematician desires a geometric explanation, only geometric proofs would count as relevant. This suggestion might allow us to 'cut back' some of the troubling overgeneration that counts all proofs as explanatory. However, it will not solve the problem entirely, because presumably a number theorist will still want to distinguish between number theoretic proofs with differing explanatory values. And it will be difficult to cut the overgeneration back any further, for the simple reason that we can't always be sure what kind of proof might be relevant for a given purpose when we set out to prove a conjecture.

Consider the famous case of Fermat's Last Theorem, which says that no three positive integers a, b, and c satisfy the equation $a^n + b^n = c^n$ for any integer value of n greater than two. Understanding the content of Fermat's last theorem requires only an understanding of integers and exponents, which are not especially complex. We might reasonably expect that a proof of the theorem would draw on similarly accessible concepts, and a search for this kind of proof motivated mathematicians and others for quite some time. Eventually, however, Andrew Wiles provided a proof that involves extremely complex concepts such as the theory of elliptic functions [Singh 1997]. It's not clear how we could have allowed for this when formulating a relevance relation at the outset.

Similarly, in Chapter 5 we will see a case where a whole new theory – group theory – is brought in to explain a result about quintic polynomials that already had a proof. When the original proof was provided as an answer to the why-question, we couldn't have formulated a relevance relation that mentioned group theory, because group theory didn't exist (or hadn't been discovered) yet.

Sandborg presents another convincing example involving the convergence of sequences with specific growth rates, where 'what is relevant to the explanandum is not known prior to the explanation itself' [Sandborg 1998: 615].

So, van Fraassen's account faces a dilemma here. Either the account of the relevance relation is formulated after we have already decided which proofs are explanatorily relevant (in which case the relation is redundant, or at least ad hoc); or the relevance relation is broad enough to be set at the outset (in which case it is in danger of being too broad and overgenerating examples: all proofs, or all proofs of a certain type, will end up counting as explanatory). As with the unification account, the why-question account fails to allow for fine-grained qualitative distinctions to be drawn between explanatory and non-explanatory proofs.

As with the counterfactual account, however, let us take a step back and consider the spirit behind the why-question account. The central idea is that explanation is context-relative: an argument or proof counts as explanatory only in a given context. Marc Lange presents an account of mathematical explanation inspired by this central idea [Lange 2017], and I will analyse Lange's proposal in depth in Chapter 4.

1.6 Concluding remarks

In this chapter, I examined four existing accounts of scientific explanation: grounding, counterfactual, unification, and why-question accounts. I presented a few simple examples from mathematics to see whether these accounts might transfer over to intra-mathematical explanation. I argued that each of the accounts faces some problems.

The grounding account is problematic because it focuses on ontic features and does not fit well with epistemic aspects of explanation, such as whether a proof helps us to understand why the theorem is true. The counterfactual account does not transfer neatly to the intra-mathematical case, because it is hard to make sense of the antecedent of the relevant counterfactual in cases where a statement is necessarily true. The unification account is problematic because it does not allow for fine-grained differences in explanatory value between proof types which are equivalent, like the principle of induction and the well-ordering principle. The why-question account involves a relevance relation which would need to do a lot of work in the intra-mathematical case, and it is difficult to restrict the relevance relation in a non ad hoc way to prevent the account overgenerating and counting all proofs as explanatory.

I have mentioned inductive proof a number of times already, and in the next chapter, I will focus on inductive proof as a case study.

2 Chapter 2: Mathematical induction

In the previous chapter, I examined some of the existing accounts of scientific explanation, isolating two accounts that looked like they might provide a promising extension to the mathematical case. Before looking at possible extensions, however, I first examine whether it might be possible to come to conclusions about mathematical explanation based solely on general intuitions about the kind of features any successful account of explanation must have.

There seems to be a fairly common view in the philosophical literature on the following point: that inductive proofs are not explanatory. Among others, Mark Steiner, Marc Lange, Johannes Hafner and Paolo Mancosu have made this claim (see [Hafner and Mancosu 2005], [Lange 2009] and [Steiner 1978]).

Indeed, inductive proofs are often seen as a suitable test case. For example, Hafner and Mancosu pose a dilemma for Steiner's account of explanation on the grounds that the account either overgenerates by counting inductive proofs as explanatory, or undergenerates by ruling out a promising example from mathematical practice [Hafner and Mancosu 2005: 237]. Often the claim that inductive proofs are not explanatory is made quickly as a side remark or footnote, and not given much evidential support beyond a strong feeling or intuition.

Marc Lange is a notable exception to this trend. In a 2009 paper he gives an argument in support of the claim that inductive proofs can never be explanatory, which is intended to rely only on very general assumptions about explanation. It is interesting to note that in more recent work Lange somewhat modifies his strong claim to the effect that inductive proofs could fall 'somewhere *between* an explanation and a proof utterly lacking in explanatory power. (Explanatory power is a matter of degree)' [Lange 2014: 511, fn 21, emphasis in the original].

I also am in favour of a more moderate view on which some inductive proofs can indeed count as (somewhat) explanatory. I think this fits more closely with mathematical practice. In this chapter, therefore, I will show what goes wrong in Lange's original argument for the claim that there are no explanatory inductive proofs; put forward some plausible examples of explanatory inductive proofs; and finally, suggest what might (mistakenly) motivate the seemingly strong and shared philosophical intuition that inductive proofs are not explanatory.

The chapter will proceed as follows. In Section 2.1, I provide my reading of Lange's argument for the claim that proof by induction is not explanatory, outlining the approach that motivates Lange's argument. In Section 2.2, I briefly examine each of the four assumptions on which Lange's argument relies, providing a number of simple examples that cast doubt on the assumptions. I also discuss a problematic 'missing link' in Lange's argument.

In Section 2.3, I present some putative examples of explanatory inductive proofs. These simple cases show that attention to concrete examples is important. Lange avoids relying on intuitions about specific proofs for methodological reasons; I press home the methodological point that ignoring concrete cases leaves the debate open to conflict with mathematical practice. Attention to examples is important in order to develop an account of explanation in mathematics that practitioners in the field can recognise.

2.1 Lange's argument

2.1.1 What is proof by induction?

Many proofs involving mathematical induction are given on the natural numbers, $\mathbb{N}$, and Lange's paper focuses on such proofs.[12]

Inductive proofs on $\mathbb{N}$ come in one of a number of different forms, all grounded using the same basic fact about the natural numbers: that they form a well-ordered and non-empty set. Each of the possible variants involves an inductive inference, which is represented by a slightly different logical principle each time. Following Lange, I will list a few of the possible variants of proof by induction, giving the form of inductive inference on which each variant relies.

[12] I will touch on induction in other domains in Chapter 4.

First, many inductive proofs start from $n = 1$, with the following structure: first, show that the property in question, P, holds for $n = 1$, i.e. $P(1)$. This is called the *base case* of the proof. Then, show that if $P(k)$, then $P(k+1)$ – known as the *inductive step*. Mathematical induction is an inference from the base case and the inductive step to the claim that the property in question holds for all natural numbers, i.e. that $\forall n P(n)$. In this case the inductive inference is represented by the following principle: $P(1) \wedge [\forall k P(k) \rightarrow P(k+1)] \vdash \forall n P(n)$. Lange provides the following familiar example, where $P(n) \longleftrightarrow 1 + 2 + ... + n = \dfrac{n(n+1)}{2}$:

> Show that for any natural number n, the sum of the first n natural numbers is equal to $n(n+1)/2$. For $n = 1$, the sum is 1, and $n(n+1)/2 = 1(2)/2 = 1$. If the summation formula is correct for $n = k$, then the sum of the first $(k+1)$ natural numbers is $[k(k+1)/2] + (k+1) = (k+1)[(k/2) + 1] = (k+1)(k+2)/2$, so the summation formula is correct for $n = k+1$. [Lange 2009: 204]

Second, a proof by induction might start from the base case $n = 5$, say, in which case two inductive steps are needed: one to show that if $P(k)$, then $P(k+1)$, and one to show that if $P(k)$, then $P(k-1)$ [Lange 2009: 207]. Following Baker, call a proof that starts from the $P(1)$ base case as above a $P1P$ proof, and a proof that starts from the $P(5)$ base case a $P5P$ proof [Baker 2010: 683]. A $P5P$ proof involves the following inductive inference: $P(5) \wedge [\forall k P(k) \rightarrow P(k+1)] \wedge [\forall k > 1\, P(k) \rightarrow P(k-1)] \vdash \forall n P(n)$. Note that there is nothing special about the case $n = 5$ here; in principle we can choose any $n \in \mathbb{N}$ for the base case.

Third, note that the variants of mathematical induction discussed so far are known as 'weak' induction, and there is a variant of mathematical induction called 'strong' induction. Strong induction involves only the following inductive step: Show that for every natural number k, if $P(n)$ holds for all natural numbers $n < k$, then $P(k)$ holds. The inductive inference in this case is represented as follows: $\forall k[\forall n < k\, P(n) \rightarrow P(k)] \vdash \forall n P(n)$. Strong and weak induction are logically equivalent in $\mathbb{N}$ in the sense that any proof by weak induction can be transformed into a proof by strong induction, and vice versa. I will say more about strong induction in Section 2.2.1.

These are not all of the possible forms of induction on $\mathbb{N}$; for example, another form is known as Cauchy induction after Augustin-Louis Cauchy, and is used to prove results like the arithmetic mean-geometric mean inequality (see for example [Dubeau 1991]). Nevertheless, the forms cited above are sufficient for the purposes of this chapter. We are now in a position to consider Lange's argument for the claim that proof by induction cannot be explanatory.

2.1.2 Lange's argument

Consider the fixed but arbitrary claim that some property, $P(n)$, holds for all n. Suppose there is some explanatory proof by induction of this general result, and suppose additionally (without loss of generality[13]) that this proof is a $P1P$ proof. Then on my reading of Lange's argument, we can proceed as follows:

1. There is an explanatory $P1P$ proof of the claim that $P(n)$ holds for all n.

2. (Reformulation) For a given $P1P$ proof by induction, there is a PkP proof by induction of the same result for some $k > 1$. [Recall that a $P1P$ proof involves base case $n = 1$, and a PkP proof involves base case $n = k$.]

3. (All or nothing) If a $P1P$ proof by induction is explanatory, then the reformulated PkP proof is also explanatory.

4. (Explanatory condition) If a proof is explanatory, then each of its premises partially explains the conclusion.

[13] Strictly speaking, there is some loss of generality here because (All or nothing) is not formulated as a biconditional. Lange initially states the assumption as a one-way conditional, but goes on to say of the $P1P$ and $P5P$ cases that 'There is nothing to distinguish them, except for where they start' [Lange 2009, 209]; so if there are any worries about cases where a PkP proof is explanatory while the $P1P$ proof is not, we can simply recast (All or nothing) as a biconditional.

5. (Asymmetry) For mathematical statements A and B, if A partially explains B, then B does not partially explain A.

6. By 1, 2 and 3, there is an explanatory PkP proof by induction that $P(n)$ holds for all n.

7. By 1 and 4, $P(1)$ partially explains the fact that $P(n)$ holds for all n.

8. Therefore $P(1)$ partially explains $P(k)$.

9. By 4 and 6, $P(k)$ partially explains the fact that $P(n)$ holds for all n.

10. Therefore $P(k)$ partially explains $P(1)$.

11. By 8 and 10, $P(1)$ partially explains $P(k)$ and $P(k)$ partially explains $P(1)$, which contradicts premise 5.

12. Hence, on pain of contradiction, there is no such explanatory proof by induction. See [Lange 2009: 207-9].

Before moving on to consider problems with this argument in Section 2.2, I will first say a bit about the motivation behind the argument.

2.1.3 Lange's approach

It seems in principle possible for it to be the case that some proofs by induction are explanatory while others are not, and that there are further borderline cases, as with many philosophical categories. Lange acknowledges this possibility, writing that 'it could be that ... some mathematical inductions are explanatory, others are not, and there is no broad truth about what they 'usually' or 'generally' are' [Lange 2009: 205, fn 3].

Nevertheless, Lange writes that his argument 'does not show merely that some proofs by mathematical induction are not explanatory. It shows that none are' [Lange 2009: 209]. So, when Lange claims that 'proofs by mathematical induction are generally not explanatory' [ibid.: 205], it seems that we should read 'generally' as meaning 'universally'. This fits with the fact that Lange's argument proceeds by reductio ad absurdum, showing that a contradiction follows from the claim that some given proof by induction is explanatory, given the assumptions made in premises 2–5.

So, Lange chooses to argue for a blanket ban on proof by induction being explanatory, rather than considering cases of proof by induction on an individual basis. Presumably Lange has some motivation for doing so; I think this motivation arises from Lange's attempt to avoid relying on intuitions about specific cases or types of proof. As Lange points out, such intuitions often conflict: 'Philosophers disagree sharply about which proofs of a given theorem explain why that theorem holds and which merely prove that it holds' [Lange 2009: 203].

Consider, for example, a proof by induction of the claim that the sum of numbers from 1 to n is $n(n+1)/2$, as given in Section 2.1.1. Gila Hanna writes that the inductive proof 'does not ... show *why* the sum of the first n integers is $n(n+1)/2$. ... Proofs by mathematical induction are non-explanatory in general' [Hanna 1990: 10]. But, according to Brown, 'a proof by induction is probably more insightful and explanatory than the picture proofs' of the same result [Brown 1997: 177].

Disagreements of this kind mean that a philosophical account of mathematical explanation is difficult to test against canonical examples: given the sharp conflict of intuitions, there are no generally accepted examples of explanatory proofs against which the proposed account can be tested [Lange 2009: 203].

Instead of focusing on examples, therefore, Lange aims to 'end this fruitless exchange of intuitions' altogether, with his 'neat argument' that avoids 'making any controversial presuppositions about what mathematical explanation would be' [Lange 2009: 203-5]. There seem to be two main components to Lange's approach: 1. Avoid relying on intuitions about specific cases of mathematical proof; and 2. Avoid relying on controversial assumptions about explanation in mathematics.

At first glance, these two conditions place a reasonable restriction on an account of explanation in mathematics. In particular, although our intuitions can be a useful guide, it seems right to say that we should avoid relying solely on intuition in order to build up an account of explanation in mathematics, not least because it is unclear whose intuitions should count.

However, there is a difference between avoiding reliance on our intuitions about examples and avoiding examples altogether. General assumptions can be just as problematic as intuitions about particular cases. Why think that our intuitions about general claims are more likely to be reliable than our intuitions about specific cases? It seems plausible that our intuitions about general cases arise from a (possibly unconscious) generalisation from examples. Such generalisations are in danger of becoming overgeneralisations, if they do not take into account a sufficiently wide range of cases.

In the rest of this chapter, I will show that none of Lange's assumptions are uncontroversial, and moreover that attention to examples can help to illuminate where each assumption goes wrong.

Let us now examine Lange's four assumptions in more detail.

2.2 Four assumptions and a missing link

2.2.1 Reformulation

Lange's first and seemingly unproblematic assumption holds that any proof by mathematical induction can be reformulated to start from a different base case. So, for example, any $P1P$ proof can be converted into a $P5P$ proof, though not, as Lange concedes, without some labour [Lange 2009: 207]. In short, Lange accepts:

> (Reformulation) For a given $P1P$ proof by induction, there is a PkP proof by induction of the same result.

Lange focuses on the case $k = 5$, and shows that a $P5P$ proof can indeed be found for the claim that the sum of numbers from 1 to n is equal to $n(n+1)/2$.

> For $n = 5$, the sum is $1 + 2 + 3 + 4 + 5 = 15$, and $n(n + 1)/2 = 5(6)/2 = 15$. If the summation formula is correct for $n = k$, then (I

showed earlier) it is correct for $n = k+1$. If the summation formula is correct for $n = k$ (where $k > 1$), then the sum of the first $k - 1$ natural numbers is $[k(k+1)/2] - k = k[(k+1)/2 - 1 = k(k-1)/2$, so the summation formula is correct for $n = k - 1$ [Lange 2009: 209].

However, reformulating a proof by induction is not so simple in every case. An example will help us to see that there are cases where it is less clear that (Reformulation) applies. Consider the following proof that proceeds by strong induction.

Proof A

Claim: For all n, if n is composite, then n is a product of primes, where n is composite iff $n = b.c$ for $b, c > 1$.

Inductive step: Suppose the claim is true for all $k < n$. Suppose $n = b.c$, where $b, c > 1$. Then $b, c < n$. If b is not composite, it is prime (since $b > 1$, and only 1 is neither prime nor composite); if b is composite, it is a product of primes by the inductive hypothesis. If c is not composite, it is prime (since $c > 1$); if c is composite, it is a product of primes by the inductive hypothesis. In either case, $n = b.c$ is a product of primes.

Proof A is not a $P1P$ proof, and indeed is not a PkP proof of any kind, since there is no base case $n = k$ that is handled separately from the rest of the domain. Hence (Reformulation) does not apply to Proof A, and therefore the assumption that Proof A is explanatory does not lead to a contradiction. This is not to say that Proof A must count as explanatory, but rather that Lange's argument does not cover all proofs involving induction, as he claims. Instead, it seems that Lange's conclusion holds only for proof by weak induction, since such proofs must involve a base case.

This seems an odd result. Recall that strong and weak induction are equivalent in $\mathbb{N}$, in the sense that any proof by weak induction can be transformed into a proof by strong induction of the same result, and vice versa. This means there is some proof, A*, that establishes the same result as Proof A but using weak induction.

To find Proof A*, note that Proof A as given above is a proof of the claim that $\forall nP(n)$, where $C(k) \longleftrightarrow k\ is\ composite$, $\Pi(k) \longleftrightarrow k\ is\ a\ product\ of\ primes$, and $P(k) \longleftrightarrow [C(k) \rightarrow \Pi(k)]$. In order to transform Proof A into a proof by weak induction of the same result, let $P'(k) \longleftrightarrow P(1) \wedge P(2) \wedge ... \wedge P(k-1)$. We can use weak induction to show that $\forall nP'(n)$, and since $\forall nP'(n) \longleftrightarrow \forall nP(n)$, we will thereby have shown that $\forall nP(n)$:

Proof A*

Claim: For all n, if n is composite, then n is a product of primes.

Base case: $n = 1$. $P'(1)$ holds iff $P(0)$ holds, and $P(0)$ holds iff $C(0) \rightarrow \Pi(0)$, which is vacuously true, since 0 is not composite.

Inductive step: Suppose $P'(k)$. holds, i.e. $P(1) \wedge P(2) \wedge ... \wedge P(k-1)$. Suppose $k = b.c$, where $b, c > 1$. Then $b, c < k$. If b is not composite, it is prime (since $b > 1$); if b is composite, it is a product of primes, since $b < k$ and $\forall n < kP(n)$ by the inductive hypothesis. If c is not composite, it is prime (since $c > 1$); if c is composite, it is a product of primes, since $c < k$ and $\forall n < kP(n)$ by the inductive hypothesis. In either case, $k = b.c$ is a product of primes, so $P(k)$. Then by the inductive hypothesis $P(1) \wedge P(2) \wedge ... \wedge P(k-1) \wedge P(k)$, and therefore $P'(k+1)$.

Now, we could could add two further premises to Lange's argument, as follows:

(Reformulation)* For any proof involving strong induction, there is a reformulated proof by weak induction of the same result.

(All or nothing)* If a proof that uses strong induction is explanatory, the reformulated proof by weak induction of the same result is equally explanatory.

Suppose Proof A is explanatory. (Reformulation)* is true, and we have seen there is a proof, Proof A*, using weak induction and which proves the same result as Proof A. According to (All or nothing)*, Proof A* is also explanatory. Since Proof A* uses weak induction, the original assumption, (Reformulation), applies to Proof A*. Lange's argument from Section 2.1.2 can then be run to get a contradiction, and we can conclude that Proof A cannot be explanatory. Using this method, Lange's argument could cover both weak and strong induction.

However, it seems to me that (All or nothing)* is quite implausible, since the reformulated Proof A* is often quite contrived. For example, I see no immediately compelling reason to think that Proofs A and A* above must be equally explanatory, and to me Proof A* seems far less explanatory than Proof A.

Of course, my intuition about Proofs A and A* is not enough to undermine Lange's argument. But in this section we have seen that the first assumption, (Reformulation), faces a challenge: the premise does not apply to all cases of proof by induction, since proofs involving strong induction need not treat a base case separately. Either Lange's argument does not have the generality he claims, or two further premises covering strong induction must be added to the argument to achieve the general result Lange desires.

Dougherty has recently raised another concern about (Reformulation), arguing that the ease of transformation between $P1P$ and $P5P$ proofs suggests they are in fact not different proofs at all, but different presentations of the same proof [Dougherty 2017].

So, Lange's first assumption is clearly not uncontroversial. Let us now move on to consider his second assumption.

2.2.2 All or nothing

According to (All or nothing), if a $P1P$ proof by induction is explanatory, then the reformulated PkP proof is also explanatory.

This assumption looks implausible in a number of cases. For example, consider a case of induction on the complexity of formulas in propositional logic. Suppose we represent well-formed formulas of propositional logic using parse trees, and define the 'height' of each wff as the length of the longest path of its parse tree (so that the objects of the induction remain in $\mathbb{N}$). We can then prove the following theorem using standard (strong) induction on the height of wffs:[14]

> **Theorem:** Every well-formed formula of propositional logic has an equal number of right and left parentheses.

I will not give the full proof here for reasons of space, but note that in the standard base case we have $n = 1$ and so we just need to show that the result holds for atoms, which is easy as atoms have no parentheses. If we were to reformulate the proof to start from base case $n \leq 5$,[15] we would need to show that the result holds for all formulas of height up to and including 5. Moreover, there is nothing special about the case $n = 5$: we could reformulate the proof to start from base case $n = 9992414$, say, in which case we would need to show that the result holds for all formulas of height (up to and including) 9992414. This would take hours of calculation if the idea is to show the base case by inspection rather than doing a separate inductive step (and if we were to add an extra inductive step, then this extra complexity would presumably give us one way to 'choose between' the $P1P$ and $P9992414P$ proofs).

According to (All or nothing), both the $P5P$ and $P9992414P$ proofs are explanatory if the $P1P$ proof is explanatory. But it looks highly implausible that the base case in the $P992414P$ proof will count as explanatory, if the base case is simply proved by inspection.

Other respondents to Lange's paper have raised similar worries. For example, Cariani also discusses cases of induction on the complexity of formulas and further questions whether the downwards inductive step in a PkP proof can 'support or can be supported by explanatory arguments' [Cariani ms: 6, fn 6].

[14] I consider structural induction, arguably a more natural way to prove this theorem, in Chapter 4.

[15] The base case in the (natural) reformulated proof will be $n \leq 5$ rather than $n = 5$ because the original proof uses strong induction.

Baker claims that it is possible to distinguish between the explanatory value of $P1P$ and $P5P$ proofs even in the classic cases considered by Lange [Baker 2010]. In the rest of this section, I will examine Baker's argument against Lange's paper, as I think it is interesting to note that both authors fall prey to similar methodological problems.

Baker's formulation of Lange's argument is slightly different from mine. Baker objects to step 4 in his own presentation of the argument:

> (4) 'There is no difference between the explanatoriness of a $P(1)$-based inductive proof of $\forall nP(n)$ and a $P(5)$-based inductive proof of $\forall nP(n)$. So $P(5)$ also partly explains $\forall nP(n)$' [Baker 2010: 682].

In my presentation, this step reads as follows:

> (All or nothing) If a $P1P$ proof by induction is explanatory, then the equivalent $P5P$ proof is equally explanatory.

Since Baker's objection focuses on the first sentence of his (4), and this is roughly equivalent to my condition, Baker's objection will have force against my presentation of Lange's argument, if it goes through. So let us now examine Baker's objection.

As Baker points out, there clearly are some differences between the $P1P$ and $P5P$ proofs, for example where they start and what kind of inductive step they involve (upwards or downwards), and Lange acknowledges these differences. But as Baker writes, 'The real issue is whether either of these differences has any implications for the relative explanatory power of these proofs. In other words, do these differences make a difference?' [Baker 2010: 683].

Baker suggests a number of differences that might be thought to make a difference to the proofs' relative explanatory power. I will focus on two differences which he claims do 'make a difference'.

First, note that a $P5P$ proof is more disjunctive than a $P1P$ proof: it deals with the cases $\{1, 2, 3, 4\}$ separately from the cases where $n > 5$. Baker claims 'that there is suggestive evidence from various sources that the degree of disjunctiveness of a proof does impact negatively on its perceived explanatoriness' [ibid.]. The evidence Baker has in mind comes from three different sources:

(i) *Support from mathematical practice*: Mathematicians seem to be unsatisfied with some proofs that proceed by going through several thousand cases, for example Appel and Haken's computer-based proof of the Four Colour Theorem. According to Baker, 'because the core of the proof involved going through several thousand cases (using a computer program), ... it therefore did not provide a satisfying explanation for why the result is true' [Baker 2010: 684]. However, it is not clear that mathematicians object primarily to the thousands of cases involved or to the use of computers. Even supposing mathematicians do object to the thousands of cases involved in the Appel-Haken proof, it is not clear whether this objection arises due to a felt lack of explanatoriness or due to a worry about the proof being in some human sense unsurveyable. There are too many confounding factors here to claim that it is disjunctiveness that leads to lack of explanatoriness in this case. Hence, this evidence alone is not particularly compelling.

(ii) *Support from intuition*: According to Baker, some very disjunctive proofs are clearly non-explanatory: 'For example, going through the first 98 even numbers greater than 2 and verifying that each can be expressed as the sum of two primes clearly counts as a perfectly acceptable proof of the proposition 'All even numbers less than 200 satisfy Goldbach's Conjecture.' Equally clearly, however, it does nothing to explain why this proposition is true' [ibid.]. Although there is no known non-disjunctive proof in this case with which to compare explanatoriness, Baker's intuition here seems right. Nevertheless, this intuition does not establish that explanatoriness is in general negatively correlated with disjunctiveness.

(iii) *Support from the philosophical literature*: Baker points to general models of explanation as unification, for example the models developed by Kitcher and Friedman, which support the view that very disjunctive proofs are less explanatory than more unified, non-disjunctive proofs [ibid.: 685]. I have already presented some worries about extending the unification account to the intra-mathematical case in Chapter 1. So I take this evidence to be fairly weak.

In summary, Baker has pointed to three ways of supporting his claim that more disjunctive proofs are generally less explanatory. The evidence Baker provides for his claim is fairly weak, since he relies on only one case from mathematical practice, one case of intuition and one potentially problematic model of explanation as unification.

Nonetheless, it may be that more compelling evidence can be found, and Baker's claim does seem to have some intuitive appeal. It may be instructive to examine some more cases from mathematical practice to find out whether 'the degree of disjunctiveness of a proof does impact negatively on its perceived explanatoriness' [Baker 2010: 684]. In particular, it could be interesting to examine cases in geometry, where the domain might naturally divide up into three cases: one with a point inside a circle, one with a point outside the circle, and one with a point on the circle, for example. Perhaps some degree of disjunctiveness is unproblematic in this context.

For now, though, Baker's claim faces a more pressing problem. Note that as Baker points out, both the $P1P$ and the $P5P$ proofs are disjunctive, 'since the base case is always treated differently from all other cases in the domain' [ibid.: 685]. The P1P proof is less disjunctive, since it divides the domain into only the base case and the rest of the domain, whereas the $P5P$ proof divides the domain into the base case, cases $\{1, 2, 3, 4\}$, and the rest of the domain; nevertheless, in order to undermine Lange's argument, Baker's aim is to conclude that the $P1P$ proof is explanatory and that the $P5P$ proof is not (rather than merely less) explanatory. As Baker notes, 'our goal is to draw an absolute conclusion ... from a comparative difference' [ibid.: 686].

In order to overcome this problem, Baker introduces a second difference between $P1P$ and $P5P$ proofs, which I will now discuss. According to Baker, '$P1P$ is minimal, among inductive proofs of $\forall n P(n)$, and $P5P$ is not minimal', where

> *Definition*: 'A proof, X, of a theorem, P, is minimal, relative to some larger family of proofs, F, if every part of X is present in every other proof (in F) of P'. [Baker 2010: 686]

This definition is somewhat unclear. Suppose some proof, X, is minimal. If another proof, Y, contains all of the same parts as X together with some additional parts, then it seems that Ycontains some redundant parts, which are not required in order to establish the result in question. Mathematicians are not in the habit of including redundant parts in proofs, so it seems that Baker's idea of a family of proofs surrounding a minimal proof may not map onto an interesting class of proofs on a literal reading of 'part' here.

Instead, it seems that Baker means that a minimal proof contains only the *kinds* of parts that are also present in all other proofs in the same family. The idea is that all proofs by induction contain a part of type 'base case' and a part of type 'inductive step'. As Baker puts it:

> '$P1P$ has only two basic parts: a base case and an upward induction step. This is significant because *every* proof by induction includes each of these kinds of part. In other words, every inductive proof requires at least one base case, and each inductive proof requires at least one upward induction step. $P5P$, by contrast, contains a part – the downward induction step – which is not common to all proofs by induction. Thus $P1P$ is minimal, while $P5P$ is not. [...] [N]ot only is the $P1P$ proof minimal, in the sense defined above, but ... it is *uniquely* minimal among inductive proofs. The *only* way for an inductive proof of $\forall n P(n)$ to have just two parts is for it to have a base case of $n = 1$ and an upward induction step of size 1. Every other variation on this basic inductive form will have some sort of base case and some sort of upward inductive step, but all of this [sic] alternatives will have other, extra parts.' [Baker 2010: 686-7]

Some of the claims Baker makes here are problematic. As Proof A in Section

2.2.1 showed us, some inductive proofs involving strong induction do not contain a base case at all.[16] Such proofs contain only an inductive step and hence only one 'kind' or part. According to Baker's definition, therefore, such proofs are minimal among all inductive proofs of the same result. So $P1P$ proofs are not always uniquely minimal, if they are minimal at all.

The problem here is that Baker does not allow for the fact that some proofs using strong induction do not involve a base case. Baker's focus on general claims and lack of attention to concrete examples means that he has missed some simple counterexamples to his claims.

A second problem is that it is not clear how minimality is meant to map onto explanatoriness. Why think that minimal proofs are explanatory? Baker writes:

> 'One idea is to view [minimality] as analogous to an oft-cited condition on scientific explanations, namely that such explanations cite only factors that are relevant to the given explanandum. For example, appealing to the fact that a given sample of salt was placed in hexed water fails to count as a genuine explanation for why the salt dissolved. Why not? Because the fact that the water had been hexed is irrelevant to the behaviour of the salt. One point about an irrelevant factor is that it is superfluous: the given line of (putative) explanation can be reformulated without it. So too with certain features of non-minimal proofs. The downward induction step in $P5P$, for example, is not essential to an inductive proof of $\forall n P(n)$. This seems like a good reason to conclude that $P5P$ is not explanatory.' [Baker 2010: 688].

Now, as we will see in Section 2.2.4, Baker himself claims that mathematical

[16] Baker seems to deny this, writing that 'it is still possible to defend the minimality of the $P1P$ proof. To see why, note that even though the $P(1)$ base case need not be listed independently, it must still be proved separately within the context of the induction step. This is because for $n = 1$, the antecedent of the initial conditional of the induction step is vacuously true: $P(m)$ holds for all $m < 1$ because there are no such cases. Hence for this conditional to be established, one needs to check that $P(1)$ holds' [ibid.: 688]. Proof A, however, shows that this is not the case.

explanation is thought to be qualitatively different from scientific explanation, so it is not clear that using an analogy to cases of scientific explanation here helps his case. I suggest that an account of explanation in mathematics should not proceed by analogy with explanation in science; if there is an analogy, this should become apparent rather than featuring as an assumption.

Moreover, in this case the analogy itself seems unhelpful. In the scientific case, minimality seeks to reduce an explanation to those factors that play a genuine role in bringing about the outcome. As Baker puts it, 'minimality echoes the spirit of causal models of explanation' [Baker 2010: 689]. By contrast, no such reduction can be carried out in the mathematical case, since it is generally assumed that mathematics is a non-causal domain. As I will discuss in Section 2.2.4, it is difficult to understand in what sense a mathematical statement, A, can be responsible for another mathematical statement, B, except perhaps in the sense of A's forming part of a deduction of B.

Suppose relevant factors in a mathematical explanation are those which are responsible for bringing about the conclusion in this deductive sense. Then it seems that minimal proofs in mathematics are simply those which arrive at their conclusion by the shortest deduction possible. But then it seems implausible that minimal proofs are always the most explanatory. As Baker himself concedes, 'it seems plausible to think that length per se – in other words, sheer number of symbols – has no particular correlation with (lack of) explanatoriness' [Baker 2010: 683]. We can imagine a case where a shortcut in a deduction is found by using some principle which is itself quite difficult to understand. In this case there seems little reason to think that the shorter deduction must be the most explanatory.

In summary, Baker needs to clarify the notions of *relevance, responsibility* and *part* in mathematics. Although I think Baker is right that (All or nothing) is implausible, Baker has not yet succeeded in undermining Lange's argument on this account. Both Baker and Lange make very general claims about proof by induction which are contradicted by simple counterexamples. This points to a methodological problem with the approach taken by both Baker and

Lange: relying on general assumptions can be just as problematic as relying on intuitions about specific cases, and we must take care not to do the former in trying to avoid the latter.

In the next section, I consider Lange's third assumption.

2.2.3 Explanatory condition

Lange writes that he will 'presuppose that a mathematical explanation of a given mathematical truth F may consist of a proof (i.e. a deduction) of F from various other mathematical truths $G_1, ..., G_n$. ... In such an explanation, the G_i collectively explain why F obtains; each of G_1, G_2 and so forth helps to explain F (e.g. F is explained partly by G_1)' [Lange 2009: 206]. I have formulated this assumption as follows:

> (Explanatory condition) If a proof is explanatory, then each of its premises partially explains the conclusion.

Lange provides no argument for this claim, except to draw an analogy with cases of mathematical explanations of physical phenomena, where, for example, 'Kepler's laws of planetary motion are explained by being deduced from Newton's laws of motion and gravity ... hence, Newton's law of gravity helps to explain why Kepler's laws hold' [Lange 2009: 206]. More needs to be said about this condition, since the notion of 'helping to explain', or partial explanation, is left unclear. I will say a bit more about partial explanation in the next section.

My main worry here is that Lange's paper starts off by talking about proofs as potential explanatory items, while the moves made in his argument focus on parts of proofs (specific mathematical claims) doing explanatory work. It is not clear that these two concepts of explanation are the same. We might think of a proof as an explanation in an epistemic sense – for example as an argument that convinces an audience – while we might see the explanatory relation between two facts as involving some kind of ontic dependence relation (as in the recent debate on grounding).

A similar kind of worry is pointed out by Baldwin, who asks:

> 'What are the relations between? We have argued that the explanatory object is a proof and such is the title of [Lange's] paper. But he concludes (where P is some property that may or may not hold of a natural number), 'It cannot be that $P(1)$ helps to explain why $P(5)$ holds and that $P(5)$ helps to explain why $P(1)$ holds, on pain of mathematical explanations running in a circle.' This leap exacerbates the reduction of the argument to considering only 'premises and consequence' by conflating the entire explanation with any component of it' [Baldwin 2016: 78-9].

So as with the previous assumptions, we see that (Explanatory condition) is not uncontroversial. Let us therefore move on to the final assumption, (Asymmetry).

2.2.4 Asymmetry

Lange writes that he 'presuppose[s] only that mathematical explanations cannot run in a circle ... that when one mathematical truth helps to explain another, the former is partly responsible for the latter in such a way that the latter cannot be partly responsible for the former. Relations of explanatory priority are asymmetric' [Lange 2009: 206]. Thus Lange endorses:

> (Asymmetry) For mathematical statements A and B, if A partially explains B, then B does not partially explain A.

Lange's defence of (Asymmetry) seems to rest on the idea that if (Asymmetry) were false, then 'mathematical explanation would be nothing at all like scientific explanation' [ibid.]. This does not seem a good reason in itself to accept (Asymmetry), especially given Lange's intention to avoid 'making any controversial presuppositions about what mathematical explanations could be' [Lange 2009: 203]. The assumption that mathematical explanation is like scientific explanation is certainly not uncontroversial. For example, Baker claims that 'the predominant view is that mathematical explanation is qualitatively different both from scientific explanation and from explanation in 'ordinary non-scientific contexts" [Baker 2012: 244]. The principle of (Asymmetry) thus needs substantial further support.

For now, note that further worries can be raised about (Asymmetry). Baker, for example, writes that 'I have some doubt about the inevitability of this condition, especially since Lange's argument is formulated in terms of partial explanation' [Baker 2010: 682]. I share Baker's worry about (Asymmetry), especially since Lange's notion of partial explanation is left unclear.

If 'A partially explains B' means simply that A contributes to an explanation of B, then (Asymmetry) is unconvincing. Consider the following example: Mary likes John, partly because he is kind, and partly because he likes her. John likes Mary, partly because she is witty, and partly because she likes him. It seems plausible that A: 'John likes Mary' partly explains B: 'Mary likes John', and vice versa. Here, then, is an initial reason to doubt (Asymmetry) in general.

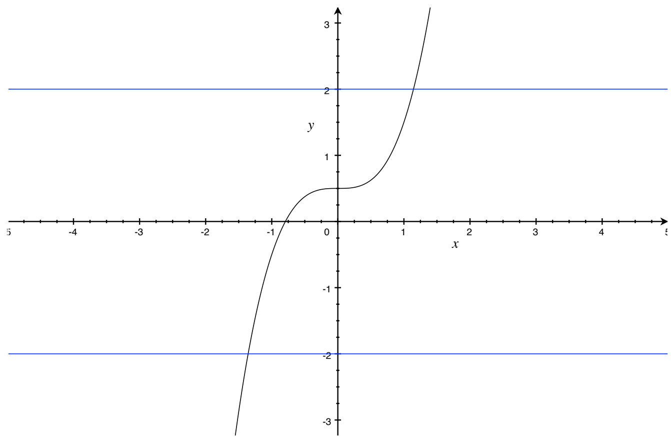
Now, the mathematical case must involve non-causal explanation. Nevertheless, it seems plausible that a mathematical counterexample to (Asymmetry) also exists, if partial explanation amounts simply to contribution to an explanation. For example, let A be the intermediate value theorem:

> (A) Let $f : [a, b] \to \mathbb{R}$ be a continuous function, and let u be a real number such that $f(a) < u < f(b)$. Then for some $c\epsilon[a, b], f(c) = u$.

And let B be the intermediate zero theorem:

> (B) Let $f : [a, b] \to \mathbb{R}$ be a continuous function, and suppose $f(a) < 0 < f(b)$ or $f(a) > 0 > f(b)$. Then for some $c\epsilon[a, b], f(c) = 0$.

B is a special case of A. Therefore, it seems that A can contribute to an explanation of B, for we can explain why B holds as follows: B is true because B is a special case of A, where $u = 0$, and A is true. On the other hand, it also seems that B can contribute to an explanation of A, as follows. We can 'see' why B holds when $f(a) < 0 < f(b)$ by looking at a diagram:



It is clear that a continuous curve passing from below the x-axis to above the x-axis must at some point, c, pass through the x-axis, where $y = f(c) = 0$. But this fact can be used to explain why A holds. For a similar diagram can be given where instead of the x-axis we consider the line $y = u$. The diagram above helps us to understand why a continuous curve passing from below the line $y = u$ to above the line $y = u$ must at some point, c, pass through the line $y = u$. At this point, c, $y = f(c) = u$ as required.[17]

This sketch of an argument does not amount to a proof of A from B, but it does seem that the special case, B, represented in a diagram, helps to explain why A holds. Therefore, it is possible for A to contribute to an explanation of B and for B to contribute to an explanation of A. So Lange's notion of partial explanation cannot involve mere contribution to an explanation; some stronger conception of partial explanation must be intended.

Lange writes that 'when one mathematical truth helps to explain another, the former is partly responsible for the latter in such a way that the latter cannot then be partly responsible for the former' [Lange 2009: 206]. Unfortunately, this does not clarify the matter, since the idea of A 'being responsible

[17]For worries about the reliability of diagrammatic reasoning in this example, see e.g. [Giaquinto 2011].

for' B is unclear in a mathematical context, where an appeal to causation is ruled out. We might attempt to formulate the notion of 'being responsible for' as the instantiation of some kind of grounding relation. However in Chapter 1 we saw that on Lange's own view, grounding relations are unlikely to be linked to explanation in mathematics.

An alternative thought might be that A is partly responsible for B if A forms part of a deduction of B. However, this cannot be Lange's intention since (Asymmetry) would then be false: It is possible for A to be part of a deduction of B and for B to be part of a deduction of A. Indeed, this relation will occur whenever two mathematical propositions are shown to be equivalent. For example, in classical propositional logic, $A = \neg p \to q$ can be used to derive $B = p \vee q$, and vice versa.

In the absence of a more detailed account of partial explanation, (Asymmetry) is problematic as well as controversial. Note that a few simple examples have sufficed in order to cast doubt on each of Lange's four assumptions.

2.3 A missing link and bridging principles

So far, I have put forward some simple examples that cast doubt on the assumptions on which Lange's argument relies. Lange may provide further defence for these assumptions, but there are further problems with Lange's argument that proof by induction is not explanatory. In particular, there is a problematic 'missing link' in Lange's argument, which I will discuss in this section. The missing link is an inferential gap from steps 7 to 8 in Lange's argument, that is, from the claim that $P(1)$ partially explains the fact that $P(n)$ holds for all n, to the claim that $P(1)$ partially explains $P(k)$.[18] Some further principle is needed to justify this inference.

One such principle could be: 'If A partially explains B, and B entails C, then A partially explains C'. However, this principle is incorrect in general, as we see from the following statements.

[18]The same gap is present in the step from 9 to 10, i.e. from the claim that $P(k)$ partially explains the fact that $P(n)$ holds for all n, to the claim that $P(k)$ partially explains P(1).

A. Harry thought that Sally would arrive at 5, and Harry wanted to arrive when Sally did.

B. Both Harry and Sally arrived at 5.

C. Sally arrived at 5.

Assuming that beliefs and desires can play a role in explanation, it seems that A partially explains B, in the sense that A explains one of B's conjuncts. Furthermore, B entails C; and yet A does not partially explain C. So Lange needs to provide a better alternative principle to close the gap in his argument.

Another possible bridging principle would be that 'Any fact which partially explains a universally quantified truth also partially explains every other instance of that truth'. Cariani for example formulates this as a principle of Minimal Closure: 'If $p_1, ..., p_n$ explain $\ulcorner \forall x \Phi x \urcorner$, then $p_1, ..., p_n$ explain $\ulcorner \Phi t \urcorner$ for any (referring) singular term t in the language (provided, of course, that $\ulcorner \Phi t \urcorner$ is not one of the p_i's' [Cariani ms: 3].

However, in a later paper Lange explicitly denies relying on this principle, writing 'I do not base [the missing link] on the premise that if a fact helps to explain a given universal generalization, then it must help to explain every instance of that generalization' [Lange 2010: 326, fn 15].

Unfortunately Lange does not himself provide an explicit principle to support the inference from $P(1)$ partially explaining $\forall n P(n)$ to $P(1)$ partially explaining $P(k)$. Instead, Lange's support for the inference comes from an analogy:

> Compare a scientific example: Coulomb's law (giving the electrostatic force between two stationary point charges) explains why the magnitude E of the electric field of a long straight wire (of negligible thickness) with uniform linear charge density λ is $2\lambda/r$ at a distance r from the wire. In explaining why for any λ and r, $E = 2\lambda/r$, Coulomb's law explains in particular why $E = 4\ dyn/statcoulomb$ if $\lambda = 10\ statcoulombs/cm$ and $r = 5\ cm$. By the same token, if $P(1)$ explains why for any n, $P(n)$, then $P(1)$ explains in particular

why $P(5)$. [Lange 2009: 210]

As before, we might question Lange's assumption that an analogy to scientific explanation is appropriate here. But even if mathematical explanation is relevantly like scientific explanation, the problem with the analogy here is that it does not involve partial explanation: instead, Lange suggests that Coulomb's law explains a general relation, $E = 2\lambda/r$, and hence explains any particular instance of that relation. For the analogy to provide a convincing parallel, we need a case where some instance of a given law helps to explain a general relation, and hence helps to explain any other instance of that relation. Lange needs to provide an explanation of the fact that $E = 4d/s = 2(10s/cm)/5cm$, say, by another instance of the equation $E = 2\lambda/r$. Without this further case, the situation in Lange's scientific analogy does not seem to match up to the situation in the inductive proof.

How, then, might the missing link be filled? In a recent paper, Hoeltje *et al.* aim to find a principle on which Lange might implicitly be relying in order to bridge the inferential gap [Hoeltje *et al.* 2013]. Hoeltje *et al.* put forward the following putative principles of explanation (among others):

> **Case-By-Rule** A universally quantified truth explains its instances.
>
> **Rule-By-Case** A universally quantified truth is explained by its instances. [Hoeltje *et al.* 2013.: 512-13].

According to Hoeltje *et al.*, Lange must appeal to Case-By-Rule in order to bridge the gap in his argument. But, they argue, Case-By-Rule is unconvincing, because Rule-By-Case has greater appeal and conflicts with Case-By-Rule, if we accept Lange's (Asymmetry) assumption [ibid: 517].[19]

In order to find out whether either Case-By-Rule or Rule-By-Case is plausible, we first need to clarify the intended scope of these principles. Hoeltje *et*

[19] As the authors point out, it is unlikely Lange actually intends to appeal to Case-By-Rule, because together with (Asymmetry) the principle would prove the stronger claim that no instance of a universally quantified truth, T, can explain T. This stronger claim allows for a direct route to Lange's desired conclusion, relying only on (Explanatory condition) and circumventing any worries about the relative explanatoriness of $P1P$ and PkP proofs. We may assume that Lange would have taken this route if he believed it to be a viable one.

al. argue that Case-By-Rule and Rule-By-Case conflict, i.e. it cannot be the case that both 'A universally quantified truth explains its instances' and 'A universally quantified truth is explained by its instances'. Now, (Asymmetry) tells us that a given instance of a universally quantified truth cannot both explain and be explained by that truth. But (Asymmetry) leaves open the possibility that a universally quantified truth explains some of its instances and is explained by some other of its instances, with some instances not involved in an explanatory relation at all. This seems to be a situation in which both (limited scope versions of) Case-By-Rule and Rule-By-Case hold. Therefore, it seems that the intended reading of these principles is something like 'A universally quantified truth explains *each of* its instances'.

Furthermore, it seems that Hoeltje *et al.* intend Case-By-Rule and Rule-By-Case to hold as universal principles of explanation, since they argue that it is plausible that there is a general explanatory principle for the universal quantifier [ibid.: 515]. That is, the intended reading of these principles is something like '*Every* universally quantified truth explains each of its instances'. This reading is backed up by the conclusion drawn by Hoeltje *et al.* about Case-By-Rule: 'Not only is the latter principle false because it has *some* false instance, *Asymmetry* and *Rule-By-Case* imply that it *only* has false instances' [ibid.: 515, emphasis in the original]. That is, Case-By-Rule would be false if it had any false instance, and so we can take it that Case-By-Rule is intended to apply universally.

For these reasons, I will focus on the versions of Case-By-Rule and Rule-By-Case given below:

> **Case-By-Rule*** Every universally quantified truth explains each of its instances.
>
> **Rule-By-Case*** Every universally quantified truth is partially explained by each of its instances.

Rule-By-Case must involve partial explanation, because it is highly implausible to think that each instance of a universally quantified truth fully explains the truth. Rather, the idea seems to be that a universally quantified truth is explained by its instances collectively, and hence partially explained by each instance. We are still lacking an account of partial explanation in mathematics, but I will not say more about that here.

Now, why think that either of these principles holds? It seems to me that the two principles are in fact highly implausible, both as general principles of explanation and as principles of mathematical explanation.

First, take Case-By-Rule* and consider the question: 'Why yesterday did the sun not set where it rose?'. The fact that the sun did not set where it rose yesterday is an instance of the universally quantified truth that the sun never sets where it rises. But this truth surely seems lacking as an explanatory answer to the question. So, Case-By-Rule* seems incorrect as a universal principle of explanation.

In the mathematical case in particular, a further counterexample can be found. Consider the following universally quantified truth: 'All prime numbers greater than 2 are odd'. The statement quantifies over the domain of prime numbers which are greater than 2, and states that all members of this domain have the property of being odd. But this universally quantified truth surely does not explain the fact that 3 is odd, say, although 3 is an instance of the universal generalisation. Hence there is at least one universally quantified truth in mathematics that does not explain its instances, and Case-By-Rule* is unconvincing as a general principle of mathematical explanation.

On the other hand, Rule-By-Case* does not look very convincing either. Rule-By-Case* holds that a universally quantified truth is partially explained by its instances. Consider the question (familiar from Chapter 1): 'Why is everyone in the seminar room a philosopher?'. The answer 'Anna is in the seminar room and she is a philosopher, Bob is in the seminar room and he is a philosopher, ...' does not seem to provide an explanatory answer to this question, in contrast to the answer 'Because the seminar is being held at a philosophy conference'. So here is a universally quantified truth that is not explained by all of its instances collectively, and is not partially explained by its instances individually. Hence, Rule-By-Case* seems incorrect as a universal principle of explanation.

In the mathematical case in particular, a further counterexample can be found. Consider the following universally quantified truth: 'For all f such that f is a function defined on the natural numbers, $\mathbb{N}$, and $m, n \in \mathbb{N}$, $f(m) \neq f(n) \rightarrow m \neq n$'. Consider an instance of this universal generalisation, namely the constant function f defined as follows: $f(n) = 1$ for all $n \in \mathbb{N}$. The relevant property here holds only vacuously, since the antecedent of the conditional is never fulfilled. Although this instance might confirm the universally quantified truth (to the extent that there is confirmation in mathematics), surely the instance does not explain the universally quantified truth. Hence there are universally quantified truths in mathematics that are not explained by each of their instances, and Rule-By-Case* is unconvincing as a general principle of mathematical explanation.

In summary, there are simple counterexamples to both Case-By-Rule* and Rule-By-Case*, if they are taken as universal principles of explanation. I think it is plausible, therefore, that there is no universal explanatory principle linking a universally quantified truth with its instances; rather, I suggest that in some cases the instances help to explain the truth, and in some cases the truth helps to explain the instances. In the same vein, I think it is likely that there are some explanatory and some non-explanatory inductive proofs. (Other respondents to Lange's paper agree; see e.g. [Cariani ms]).

In the next section, I will back up this claim by presenting some putative examples of explanatory inductive proofs.

2.4 Explanatory inductive proofs

I will present the examples in Sections 2.4.1 and 2.4.2, saving discussion for Sections 2.4.3 and 2.4.4.

2.4.1 Two pictorial proofs

Case 1: The sum of numbers from 1 to n

The statement to be proved is the following claim: for all n in $\mathbb{N}$, $1+2+...+n = \frac{n(n+1)}{2}$. This can be proved by induction in (at least) two different ways. I provide two proofs below, and suggest that Proof C is more explanatory than Proof B.

Proof B

Base case

$For\, n = 1 : 1 = 1 \times \frac{(1+1)}{2}$.

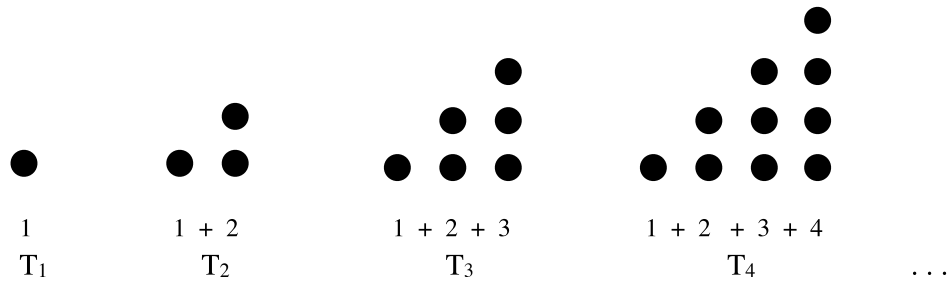
Inductive step

If for $n = k$, $1 + 2 + ... + k = \frac{k(k+1)}{2}$, then for $n = k + 1$,

$$\begin{aligned}(1+2+...+k)+k+1 &= (\frac{k(k+1)}{2})+k+1 \\ &= \frac{k^2+3k+2}{2} \\ &= \frac{(k+1)(k+2)}{2}\end{aligned}$$

Proof C

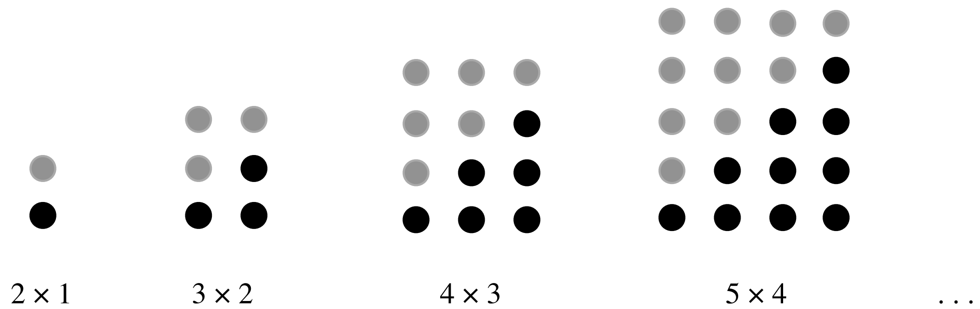
First, note that the n^{th} triangular number is equal to the sum from 1 to n by definition, ie. $T_n = \sum_{k=1}^{n} k$. The diagram below shows the first four triangular numbers.



Now, in order to prove our desired result that $1+2+...+n = \frac{n(n+1)}{2}$, we simply need to show that the value of the n^{th} triangular number is equal to $\frac{n(n+1)}{2}$.

Base case

For small values of n, we can see that the value of the triangular number T_n can be found by halving the rectangular array of dots with sides of length n dots and $n+1$ dots respectively. This is because two copies of T_n fit together to make a rectangular array of dots, as shown in the diagram below:

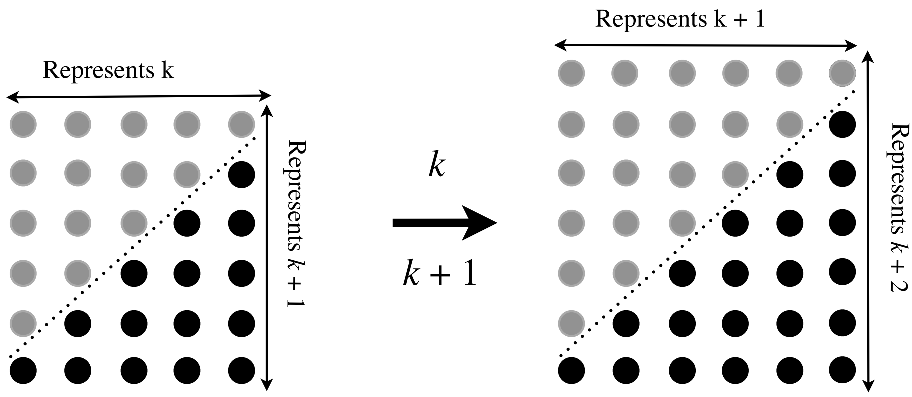


The first triangular number, T_1, is equal to half of the number of dots in the rectangular array with sides of length 1 dot and 2 dots: $1 = \dfrac{1 \times (1+1)}{2}$. And since the sum of numbers from 1 to 1 is equal to T_1, we have proved our base case.

Inductive step

Suppose that the value of the triangular number T_k is equal to half of the number of dots in the rectangular array of dots with sides of length k and

$k+1$. We now need to show that T_{k+1} is equal to half of the number of dots in the rectangular array of dots with sides of length $k+1$ and $k+2$. But this is easy to see from the following diagram:



The first rectangular array represents two copies of T_k. To get from the first to the second rectangular array, we add on a row of $k+1$ dots and a column of $k+1$ dots. This corresponds to adding on a new row of $k+1$ dots to each of the triangular components in the first array, so the second rectangular array represents two copies of T_{k+1}. So if $2T_k = k(k+1)$, then $2T_{k+1} = (k+1)(k+2)$.

Therefore, the value of the n^{th} triangular number is equal to $\frac{n(n+1)}{2}$ for all n. Since we saw that the sum of the first n numbers is equal to the n^{th} triangular number, it follows that $1+2+...+n = \frac{n(n+1)}{2}$.

Case 2: Sums of odd numbers

Claim: For any natural number n, n^2 is equal to the sum of the first n odd numbers. Again, I provide two proofs, suggesting that Proof E is more explanatory than Proof D.

Proof D

Base case

For $n = 1 : n^2 = 1$, which is the sum of the first odd number, 1.

Inductive step

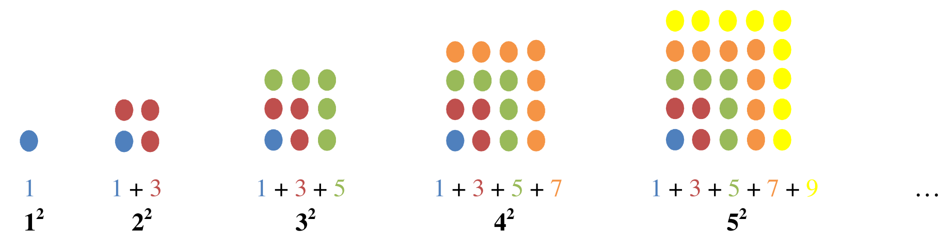
Note that the k^{th} odd number is $2k-1$, and the $(k+1)^{th}$ odd number is $2k+1$. Now, suppose that for $n = k, k^2 = 1 + 3 + 5 + ... + (2k - 1)$.

Then $(k + 1)^2 = k^2 + 2k + 1 = 1 + 3 + 5 + ... + (2k - 1) + (2k + 1)$, which is the sum of the first $k + 1$ odd numbers, as required.

Proof E

Base case

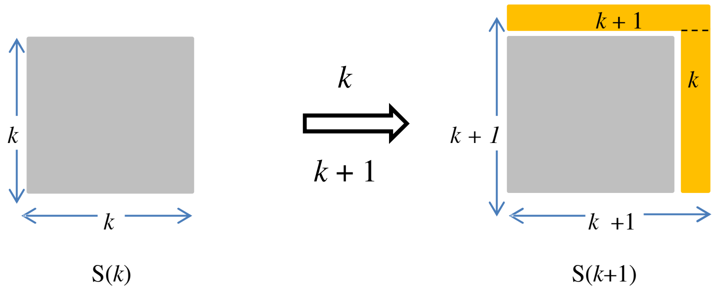
For $n = 1 : n^2 = 1$, which is the sum of the first odd number, 1. We can see that the sum of the first n odd numbers forms a square for small values of n by inspection of the following diagrams:



Here n^2 is represented by a square composed of dots, with sides of length n dots. Each time the new 'layer' of dots represents the latest odd number to be added on. The pictures show that when we add on the next odd number in this way we can form a new square, for odd numbers up to 9. In each case the new square has sides which are one dot longer than the sides of the previous square. We see that 1^2 is represented by a square composed of just one dot – the first odd number – since $1^2 = 1$.

Inductive step

The picture given above shows that the sum of the first n odd numbers is equal to n^2 for small values of n. We now need to prove the general result by proving the inductive step. Let $S(k) = k^2$. Suppose that the desired result holds for $n = k$. So $S(k)$ is equal to the sum of the first k odd numbers. We need to show that $S(k+1)$ is equal to the sum of the first $k+1$ odd numbers. We can represent $S(k)$ as a square with sides of length k, since this square has area k^2.



In the way we added on dots in the base case, we can now add on two rectangles to the square $S(k)$ – one with sides of length 1 and $k+1$, and one with sides of length 1 and k. In this way we can form a square with sides of length $k+1$. Since this square has area $(k+1)^2$, it represents $S(k+1)$. Note that the rectangles added on have an area of $2k+1$ in total. So $S(k+1) = S(k)+2k+1$. But since the $(k+1)^{th}$ odd number is $2k+1$, and $S(k)$ is the sum of the first k odd numbers, we see that $S(k+1)$ is the sum of the first $k+1$ odd numbers, as required.

2.4.2 Intuitions

I find Proof C explanatory, and I find it more explanatory than Proof B; I find Proof E explanatory, and I find it more explanatory than Proof D. How might we account for this intuition? Various respondents to Lange's paper have proposed that explanatory inductive proofs should be explanatory 'all the way through', as we might put it. For example, Hoeltje *et al.* suggest that explanatory inductive proofs might exist, 'if they involve explanatory subproofs of *both* inductive basis and step' [Hoeltje *et al.* 2013: 520, emphasis in the original]. Similarly, Cariani proposes a 'Transmission Requirement: a proof by mathematical induction is explanatory only if the arguments for all of its components are themselves explanatory', adding that 'Sometimes one or more of these arguments will be completely trivial; that's enough to pass this requirement' [Cariani ms: 8].

Do my preferred proofs meet this condition? It looks like the pictorial proofs (C and E) do give us some insight into why the result holds for small values of n. For example, proof C displays the connection between the sum $1 + 2 + ... + n$, the triangular number T_n and the term $\frac{n(n+1)}{2}$ in an easily accessible way, so that we can see directly why the result holds in the base case. The diagrams in Proof C also help to display the construction from one case to the next in an accessible way: we visualise a generic instance and the steps are operations on visualised instances, helping us to see directly why the result holds for $k + 1$ if it holds for k.

By contrast, Proof B links the relevant properties through symbol manipulation, using a chain of deductive inferences. Somewhere in the process of symbol manipulation, as we follow Proof B step by step, it is easy to lose track of what the terms mean. The steps are applications of rules of symbol manipulation, rather than operations on visualised instances. This might be one reason to think that Proof B is less explanatory than Proof C.

Note that this suggestion is not uncontroversial. For example, Baldwin denies that keeping track of the meaning of terms is necessary for a proof to have explanatory value:

> 'The glory of algebra is that one does not need (and possibly cannot) keep track of the meaning of each term in a derivation; nevertheless, the variables have the same interpretation at the end of the derivation as the beginning. The thought seems to be that losing track of the explicit reference of each term means the argument is non-explanatory but mere calculation. We have just seen the fallacy of this assertion ...' [Baldwin 2016: 75-6].

This is no problem for me, however: if Baldwin is right and Proof B could or should count as explanatory after all, so much the better for my claim that some inductive proofs are explanatory.

The reader might (I hope) agree with my claims about the two cases presented, and my intuitions do have some support in the literature. For example, Cariani also presents a version of Proof E as a putative example of an explanatory inductive proof [Cariani ms: 10]. However, as I noted earlier, Lange emphasises that he wishes to end what he calls the 'fruitless exchange of intuitions' surrounding proof by induction. So more needs to be said in support of my suggestion that some inductive proofs are explanatory.

On the other hand, in order for Lange to reject my intuitions, he needs to justify this rejection, perhaps by arguing that the explanatory work is done by some factor external to the proof. For example, Lange might argue that the diagrams present in proof C are doing all of the explanatory work, and are not themselves an essential part of proof C, since they could be removed. In that case proof C itself is not explanatory after all.

In response, I would question the standards of proof which hold that diagrams such as those in proof C are not a genuine component of proof C. Although there is a proof that does not involve diagrams – namely proof B – proof C, including its diagrams, seems to be a perfectly reasonable proof, according to the usual notions of proof used by mathematicians. Proof C appeals to another area of mathematics, namely geometry. For example, proof C appeals to the fact that a product of two natural numbers can be represented as a rectangular array of dots. But it seems quite reasonable for facts from one area of mathematics to be used to explain facts in another area of mathematics.

The point here is that the diagrams in proof C play a genuine role in the proof: they specify the way in which the extra row and column of dots are added on to the original rectangular array, for example. This could be specified in another way, but simply eliminating the diagrams from proof C would leave a gap in the proof.

The role of diagrams in proofs C and E brings out an interesting issue. So far, we have been assuming that degree of explanatoriness is a property of proofs. Instead, we could take degree of explanatoriness to be a property of proofs under a given presentation. So, perhaps we should say that a proof by induction of the claim $1 + 2 + ... + n = \frac{n(n+1)}{2}$ under presentation C is more explanatory than a proof by induction of the same claim under presentation B. Some work needs to be done in order to establish what counts as the same presentation of a given proof: which cognitive differences are relevant? The inclusion of diagrams is presumably a cognitively relevant difference, but the printing format (say typed or handwritten) surely is not.

As Dougherty puts it:

> '... some rearrangements of a proof's presentation are clearly changes in presentation alone, not in the proof itself. Changing the order of the premises, or changing the names of variables, or translating the proof into French does not change the mathematical content. In these cases it's easy to see that such manipulations change nothing about the mathematics, because they make no reference to the mathematical facts involved or the relationships between them.' [Dougherty 2017: 5476]

For my purposes here, I merely want to allow for the possibility that a proof under one presentation is more explanatory than under another presentation.[20] It seems to me that any account of mathematical explanation should at least allow for this possibility, in order to be faithful to mathematical practice.

Now, one might be tempted to press the dialectical point here and simply claim that intuitions about the three cases I have presented are more secure than Lange's four assumptions, which we saw earlier were controversial.

However, I think the point is more subtle. The claim is not that we should rely exclusively on intuitions about examples, but that we should use examples to inform the questions we consider and the constraints we place on a successful account of mathematical explanation. In particular, it is important to pay attention to a sufficiently wide range of cases, so that we don't inadvertently jump to overgeneralisations about all inductive proofs from a small number of cases taken from the same domain. A lack of attention to cases from a wide range of mathematical domains probably helps to explain the common philosophical intuition that inductive proofs are not explanatory.

I further support this claim in Chapter 4, where I consider Lange's more recent work on mathematical explanation, in which he somewhat modifies his strong claim that proof by induction can never be explanatory.

[20] I will expand on this suggestion in the next chapter.

2.5 Concluding remarks

In this chapter, I have used simple examples from mathematics to suggest that there is reason to doubt Lange's argument for the claim that inductive proofs are never explanatory. In Section 2.1, I outlined Lange's argument and the background approach that motivates it. In Section 2.2, I examined the four assumptions on which Lange's argument relies. Attention to specific examples helped us to see, for example, that (Reformulation) is problematic and that (Asymmetry) is false under various interpretations. I also discussed a missing link in Lange's argument pressed by Hoeltje *et al.*, suggesting that it is unlikely the link can be bridged by a general principle, and that it is plausible that there is no general rule linking induction and explanatory value. That is, it seems likely that some inductive proofs are explanatory and some are not. In Section 2.3, I put forward two putative examples of explanatory inductive proofs.

The general methodological points I have tried to make are that our assumptions about mathematical explanation should stand up to scrutiny of particular examples, and that general intuitions can be just as problematic as intuitions about specific cases. The debate surrounding Lange's argument has focused on very general claims about proof and explanation in mathematics, without detailed discussion of examples. This approach has been taken because, as we have seen, Lange wishes to avoid appeal to our intuitions about specific proofs and proof types.

Nevertheless, there is a difference between avoiding reliance on our intuitions about examples and avoiding examples altogether. We have seen that general assumptions can be just as problematic as intuitions about particular cases. Why think that our intuitions about general claims are more likely to be reliable than our intuitions about specific cases? It seems plausible that our intuitions about general cases arise from a (possibly unconscious) generalisation from examples. Such generalisations are in danger of becoming overgeneralisations: I suggest that (some) philosophers have been too hasty to tar all inductive proofs with the same brush, perhaps due to an overfamiliarity with standard arithmetical cases.

My conclusions: (1) We have reason to doubt the claim that no proof by mathematical induction is explanatory; (2) intuitions about particular cases (of candidate mathematical explanations) should not be dispensed with in favour of exclusive reliance on intuitions about general principles (of mathematical explanation).

3 Chapter 3: Characterizing properties

In this chapter, I return to the type of account described in Section 1.3, which is based on the central idea that explanations work by identifying a relevant 'difference-maker'.

Steiner's account [1978] is one of the earliest contemporary accounts of mathematical explanation, and it appeals to characterizing properties of entities referred to in proofs. Unfortunately Steiner's remarks are often quite vague, sometimes described as 'very puzzling indeed' [Hafner and Mancosu 2005: 233], and this lack of clarity has led to a lack of understanding and a tendency to reject Steiner's account in the philosophical literature.

I argue that Steiner's account repays deeper analysis by providing a sympathetic reading that makes sense of his puzzling remarks and draws out some important questions.

I focus on a simple mathematical example involving sums of number sequences and identify three key conditions that the proof must meet to count as explanatory for Steiner. I propose a suitable characterizing property and show that on my suggestion, the proof indeed fits Steiner's account. Subsequently, I present a few potential problems relating to Steiner's focus on the generalizability of proofs, and show how my reading of generalizability helps to avoid these worries.

Finally, I show how (my extension of) Steiner's proposal can account for what I take to be the primary epistemic function of an explanation, namely, to help us see why the fact to be explained is true.

3.1 Steiner's account of mathematical explanation

3.1.1 Three conditions on explanation in mathematics

According to Mark Steiner:

> '... an explanatory proof makes reference to a characterizing property of an entity or structure mentioned in the theorem, such that

> from the proof it is evident that the result depends on the property. It must be evident, that is, that if we substitute in the proof a different object of the same domain, the theorem collapses; more, we should be able to see as we vary the object how the theorem changes in response. In effect, then, explanation is not simply a relation between a proof and a theorem; rather, a relation between an array of proofs and an array of theorems, where the proofs are obtained from one another by the 'deformation' prescribed above. (But we can say that each of the proofs in the array 'explains' its individual theorem.)' [Steiner 1978: 143]

So, a proof must fulfil three conditions in order to be explanatory:

1. The proof makes reference to a characterizing property of an entity or structure mentioned in the theorem.

2. It is evident from the proof that the result depends on the property, that is if we substitute in a different object of the same domain, the theorem collapses.

3. We should be able to see as we vary the object how the theorem changes in response: the proof is generalizable.

Steiner's suggestion is usually broken down into two conditions where points 1 and 2 are combined (see e.g. [Hafner and Mancosu 2005], [Resnik and Kushner 1987]), but the first condition here is important to Steiner. As we will see in the next section, he stresses the claim that the characterizing property must be a property of something mentioned in the theorem. Therefore, I include condition 1 as a separate criterion.

Steiner does not explicitly say that these conditions are meant to be necessary and sufficient criteria for a proof to be explanatory. Nevertheless, he puts forward examples which he takes it meet the schema and are therefore explanatory, suggesting that the conditions are sufficient on his view. Additionally, he considers a possible counterexample proposed by Feferman: a proof that is thought to be explanatory but does not seem to meet Steiner's criteria. In response, Steiner finds an appropriate characterizing property in order to fit the proof to his schema [Steiner 1978: 148]. So, it also seems that the three conditions are necessary for a proof to be explanatory, on Steiner's view.

At first glance, there is an immediate tension in Steiner's account between conditions 2 and 3. To show that a proof is explanatory, we are required to show both that the theorem 'collapses' for a different object of the same domain and that it generalizes for some other object (presumably also in the domain; Steiner does not specify otherwise). So the second condition can't be understood as saying that the theorem 'collapses' for *any* other object in the domain; rather, the theorem should collapse for *some* other object in the domain, and in particular presumably for some object that doesn't have the characterizing property. I will say more about this in Section 3.2.1.

3.1.2 The sum of positive integers from 1 to n

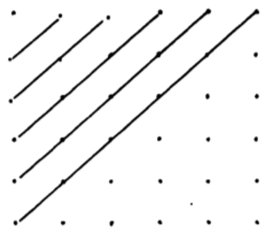
Steiner considers three proofs (strictly speaking, proof sketches) of the fact (by now familiar) that the sum of positive integers from 1 to n is equal to $\frac{n(n+1)}{2}$. [Steiner 1978: 136-7]

1. Inductive proof (strictly speaking, this is only the inductive step):
$S(n+1) = S(n) + (n+1) = n(n+1)/2 + 2(n+1)/2 =$
$(n+1)(n+2)/2.$

2. 'Symmetry' proof:

$$\begin{array}{ccccccccccc} 1 & + & 2 & + & 3 & + & ... & + & n & = & S \\ n & + & (n-1) & + & (n-2) & + & ... & + & 1 & = & S' = S \\ \hline (n+1) & + & (n+1) & + & (n+1) & + & ... & + & (n+1) & = & n(n+1) \end{array}$$

3. 'Geometrical' proof:



‘By dividing a square of dots, n to a side, along its diagonal, we get an isosceles right triangle containing

$S(n) = 1 + 2 + 3 + ... + n$

dots. The square of n^2 dots is composed of two such triangles – though if we put the triangles together we count the diagonal (containing n dots) twice. Thus we have

$S(n) + S(n) = n^2 + n$, q.e.d.’ [Steiner 1978: 137]

According to Steiner, Proof 1 is not very explanatory, Proof 2 is ‘more illuminating’ and Proof 3 is ‘perhaps an even more explanatory proof’ [ibid.: 145].

In this paper, I will focus on Proof 2 above and will give a detailed analysis of how the proof might fit Steiner’s schema. Neither Steiner nor his respondents do so. Steiner makes only the following brief remark:

> ‘Both explanatory proofs that the sum of the first n integers equals $n(n+1)/2$ proceed from characterizing properties: the one by characterizing the symmetry properties of the sum $1 + 2 + \ldots + n$; the other its geometrical properties. By varying the symmetry or the geometry we obtain new results, conforming to our scheme.’ (ibid.)

Resnik and Kushner [1987] focus on a different example: the irrationality of $\sqrt{2}$, which I will discuss in Section 3.4.1. And Hafner and Mancosu write:

> ‘Steiner’s remarks imply that he apparently takes the symmetry properties as well as the geometrical properties of the sum $1 + 2 + \ldots + n$ as something – entities or structures? – *mentioned* in [the theorem]. This is very puzzling indeed and just highlights the need

for precise definitions here. In the absence of such definitions ... we don't even have a clear enough grasp of Steiner's theory in order to apply and assess it in general'. [Hafner and Mancosu 2005: 233]

Given the lack of analysis in the literature, I think it is useful to look at this example in more depth. I hope my investigation will help us to understand Steiner's three conditions and bring out some important questions for Steiner's account.

3.2 Identifying a characterizing property

3.2.1 Commutativity and associativity of addition

Recall Steiner's example, Proof 2 from the last section:

$$\begin{array}{ccccccccccc} 1 & + & 2 & + & 3 & + & ... & + & n & = & S \\ n & + & (n-1) & + & (n-2) & + & ... & + & 1 & = & S' = S \\ \hline (n+1) & + & (n+1) & + & (n+1) & + & ... & + & (n+1) & = & n(n+1) \end{array}$$

As we have seen, it is important that the characterizing property applies to something mentioned in the theorem, for Steiner. Steiner's formulation of the theorem to be proved is as follows:

(SUM) $S(n) = 1+2+3+\ldots+n = n(n+1)/2$. [Steiner 1978: 136]

The theorem mentions the numbers 1, 2, 3, some arbitrary number n, the sum $S(n)$, the operations of addition, multiplication, subtraction and division, and the relation of equality. Is Steiner's point that these entities (taken broadly) have symmetry and geometrical properties? Hafner and Mancosu suggest that Steiner takes the symmetry properties themselves to be mentioned in the theorem, but I don't think this is required: recall Steiner's stipulation that 'an explanatory proof makes reference to a characterizing property of an entity or structure mentioned in the theorem' [ibid.: 147]. Although the statement is ambiguous, I read this as follows: the *proof* must mention the characterizing property; and the characterizing property must be a property of something mentioned in the *theorem*.

Now, Proof 2 presents the sum of integers from 1 to n in two different ways. First, the sum is given in that order: $1 + 2 + 3 + \ldots + n$. Then, the sum is given in reverse: $n + (n-1) + (n-2) + \ldots + 1$. One central insight of the proof is that, since addition is commutative and associative, these sums are the same. So we can add both sums together to get $2S(n)$, forming a new sum of n elements. Each element of this new sum is equal to $n + 1$, using the commutativity of addition again. So $2S(n)$ is equal to n lots of $n + 1$, leading to the desired result.

An initial suggestion, then, might be to propose the following characterizing property, P: *the commutativity and associativity of addition for positive integers.* This can be seen as a 'symmetry propert[y] of the sum' in the sense that the two different representations of the sum $S(n)$ are symmetrical around their midpoint, and the property 'behind' this symmetry is property P.

Proof 2 implicitly appeals to property P (and in this sense perhaps makes reference to it), and the theorem mentions addition, some integers, and the sum of integers. So it seems that P holds of some entity or structure mentioned in the theorem, as required, hence fitting Steiner's first condition.

However, I will now show that P cannot be the property Steiner has in mind, because it doesn't fit Steiner's second condition. Recall that condition 2 stipulates that the result must depend on the property in the sense that the theorem collapses if we substitute in a different object.

It is true that the theorem depends on P in the following sense: if addition for integers were not commutative and associative, we wouldn't be able to get to the desired result using this proof method. But counterfactuals like this are difficult to understand in the mathematical case where results hold necessarily. According to Steiner, the appropriate counterfactual to consider instead runs as follows: 'If we substitute in a different object, the theorem would collapse'.

It is important to consider here the domain from which the substituted object must come. The domain of the theorem is clearly (sums of) elements of $\mathbb{N}$, and Steiner explicitly specifies that the 'different object' should come from the same domain. So we want to substitute in a sum of some sequence of natural numbers.

Note that no matter which sum we choose, the cause of the theorem's collapse will not be a failure of the new object to instantiate property P. Any sum of elements in $\mathbb{N}$ is commutative and associative, so the theorem couldn't fail because of a breakdown of associativity or commutativity. But intuitively we want to choose a new object which is 'different' in the sense that it lacks the characterizing property of the original object.[21] If not, it would be hard to understand Steiner's counterfactual condition as a way to flesh out the idea of a result depending on the characterizing property.

This is a problem for our initial choice of characterizing property, and points to a feature any successful characterizing property must have: there must be at least one object in the domain for which the characterizing property does not hold. Otherwise condition 2 would be impossible to fulfil (or is fulfilled only vacuously). I doubt that Steiner has vacuous fulfilment in mind, so on a charitable reading of Steiner's account, I must have identified the wrong characterizing property.

I will suggest a new one in the next section; nevertheless, I hope the discussion here has helped to clarify the content of Steiner's second condition. In particular, I suggest that the best reading of condition 2 runs as follows: 'If we substitute an object from the same domain which lacks the characterizing property, then the theorem collapses'. I will say more about how to understand the 'collapse' of the theorem in the next section.

3.2.2 Arithmetic sequences

Take the following property, Q: *being an arithmetic sequence in* $\mathbb{N}$. (An arithmetic sequence is one with a constant difference between consecutive terms). Let us see whether this new property helps to fit Proof 2 to Steiner's account. For ease of reference, I repeat the theorem and proof (or proof sketch):

[21] I will speak interchangeably of entities and objects, staying neutral on matters of mathematical ontology.

Theorem: (SUM) $S(n) = 1 + 2 + 3 + \ldots + n = n(n+1)/2$**.**

Proof 2:

$$\begin{array}{ccccccccccc} 1 & + & 2 & + & 3 & + & ... & + & n & = & S \\ n & + & (n-1) & + & (n-2) & + & ... & + & 1 & = & S' = S \\ \hline (n+1) & + & (n+1) & + & (n+1) & + & ... & + & (n+1) & = & n(n+1) \end{array}$$

Steiner's three conditions are satisfied as follows.

1. The proof makes reference to a characterizing property of an entity or structure mentioned in the theorem.

The proof and theorem both mention an entity S or $S(n)$, the sum of the first n natural numbers. The first n natural numbers form a sequence in $\mathbb{N}$ and this sequence is also mentioned, in the sense that its terms are explicitly listed. The sequence has property Q.

Now, the proof does not explicitly make reference to property Q, but it does implicitly appeal to Q: if the sequence did not have a constant difference between consecutive terms, then the terms of the sequence and its mirror image would not 'match up' to sum to $(n + 1)$ in each case. I suggest that property Q is represented in the diagram appealed to in the proof, so Steiner's first condition is fulfilled.

We might worry whether property Q really fits with the fact that Steiner calls the desired property a 'symmetry property' of the sum $1 + 2 + \ldots + n$ [Steiner 1978: 145]. Property Q is in fact a property of the sequence $1, 2, .., n$, rather than of the sum. But since Steiner chooses to write $1 + 2 + \ldots + n$ here rather than $S(n)$, I think we can reasonably take him to be referring to the summand – in this case the terms of the arithmetic sequence – rather than the sum itself. The sum is simply a number in $\mathbb{N}$, and it is not clear what symmetry property might hold of the sum.

There is a certain symmetry, on the other hand, in the way terms of the sequence are regularly spaced; it's this fact which means that the terms in the sequence and its mirror image match up in each case. Steiner's remark is vague enough that I think this level of symmetry seems like a reasonable fit.

2. It is evident from the proof that the result depends on the property, that is if we substitute in a different object of the same domain, the theorem collapses.

Consider the following sequence: $1, 4, 6, 25, 49, 101$. The theorem collapses for this sequence because the terms of the sequence and its mirror image do not 'match up' in each case:

$$\begin{array}{rccccccccccc} S*: & 1 & + & 4 & + & 6 & + & 25 & + & 49 & + & 101 \\ S*: & 101 & + & 49 & + & 25 & + & 6 & + & 4 & + & 1 \\ \hline 2S*: & 102 & + & 53 & + & 31 & + & 31 & + & 53 & + & 102 \end{array}$$

We can see from the diagram that the method used to calculate S in Steiner's example does not work for $S*$, since $2S*$ is not a sequence with constant terms, unlike $2S$. It's not that there is nothing to say about the sum $S*$; we can still calculate its value. However, we can only do so by adding each of the terms. We don't get an equation like the one in Steiner's original theorem to calculate the value of the sum.[22]

So, we have: (i) taken another object from the same domain (taken to be sequences of elements of $\mathbb{N}$), where (ii) the object lacks the characterizing property and (iii) the theorem collapses. I hope that points (i) and (ii) are clear, but (iii) needs some further discussion. Steiner talks loosely about the theorem collapsing, but it seems from the discussion of the non-arithmetic sequence above that it is really the proof method that collapses. I think this is right; focusing on the proof method or argument, rather than the theorem, is a more fruitful way to interpret Steiner's second condition.

To back up my claim, consider the example of geometric sequences. Geometric sequences lack the characterizing property Q, and so in the spirit of Steiner's account, the theorem should collapse for geometric sequences. On my reading of 'the theorem collapses', this is true: the proof method above

[22]Don't be deceived by the kind of symmetry still present in 2S*, as it doesn't get you an equation for calculating the sum that doesn't just involve adding up the terms. We just get

$$S*(n) = \begin{cases} \frac{1}{2}(2(a_1 + a_n) + 2(a_2 + a_{n-2}) + ... + 2(a_{n/2} + a_{(n+2)/2})) & if\ n\ is\ even \\ \frac{1}{2}(2(a_1 + a_n) + 2(a_2 + a_{n-2}) + ... + 2(a_{(n-1)/2} + a_{(n+3)/2}) + 2a_{n+1/2}) & if\ n\ is\ odd \end{cases}$$

In both cases, this is simply a reformulation of $S*(n) = \sum_{k=1}^{n} a_k$.

does not work for calculating the sum of a geometric series. Consider a simple example, such as the case where each term of the sequence is double the previous term for $n > 1$:

$$
\begin{array}{rccccccccccc}
S'': & 1 & + & 2 & + & 4 & + & 8 & + & 16 & + & 32 \\
S'': & 32 & + & 16 & + & 8 & + & 4 & + & 2 & + & 1 \\
\hline
2S'': & 33 & + & 18 & + & 12 & + & 12 & + & 18 & + & 33
\end{array}
$$

Just as before, $2S''$ is not a sequence with constant terms and the proof method fails to help us find a value for the sum S''. However, unlike in the previous example, it's not the case that no version of the *theorem* holds for the new sequence. There is a formula for finding the value of sums of geometric sequences: in general $S_n = \dfrac{a_1(1-r^n)}{1-r}$, $r \neq 1$, where n is the number of terms, a_1 is the first term and r is the common ratio. So it's not clear that the theorem collapses, on a reading where this does not refer to the proof method used.

Hence I will take the following approach. Wherever Steiner writes 'the theorem collapses', I will take this to be shorthand for the following: Given a certain proof using a characterizing property R of an entity referred to in the theorem, the *theorem collapses* just in case the same argument applied to objects lacking the characterizing property R is not a proof of the modified proposition that is now the conclusion.

With this in mind, let's move on to condition 3 and see how the original proof is generalizable.

3. We should be able to see as we vary the object how the theorem changes in response: the proof is generalizable.

In keeping with my reading of the theorem collapsing, I will read condition 3 as follows. Given a certain proof using a characterizing property R of an entity referred to in the theorem, the *proof generalizes* just in case the same argument applied to other objects with characterizing property R is a proof of the modified proposition that is now the conclusion. I think we can charitably assume that Steiner's talk of the theorem changing as we vary the object is simply a loose shorthand for the account of generalizability just sketched.

Here are two examples of the theorem from Proof 2 generalized to cover other sums of integers.

Theorem A: The sum of the first n odd positive integers is equal to n^2. Or $S(n) = \sum_{k=1}^{n} 2k - 1 = 1 + 3 + 5 + ... + (2n-1) = n^2$.

Proof A:

$$\begin{array}{lccccccccccccc} S(n): & 1 & + & 3 & + & 5 & + & 7 & + & \dots & + & 2n-1 \\ S(n): & 2n-1 & + & 2n-3 & + & 2n-5 & + & 2n-7 & + & \dots & + & 1 \\ \hline 2S(n): & 2n & + & 2n & + & 2n & + & 2n & + & \dots & + & 2n \end{array}$$

Since there are n terms in the sequence, we have $2S(n) = n.2n$ and hence $S(n) = n^2$.

Theorem B: The sum of the first n terms of the following sequence: $1, 4, 7, 10, 13, \dots$ is equal to $\frac{1}{2}(3n^2 - n)$. Or $S(n) = \sum_{k=1}^{n} 3k - 2 = \frac{1}{2}(3n^2 - n)$.

Proof B:

$$\begin{array}{lccccccccccccc} S(n): & 1 & + & 4 & + & 7 & + & 10 & + & \dots & + & 3n-2 \\ S(n): & 3n-2 & + & 3n-5 & + & 3n-8 & + & 3n-11 & + & \dots & + & 1 \\ \hline 2S(n): & 3n-1 & + & 3n-1 & + & 3n-1 & + & 3n-1 & + & \dots & + & 3n-1 \end{array}$$

Since there are n terms in the sequence, we have $2S(n) = n.(3n-1) = 3n^2 - n$ and hence $S(n) = \frac{1}{2}(3n^2 - n)$.

Strictly speaking, these are proof sketches. We might ask for further clarification of the fact that the n^{th} term in Theorem B is $3n - 2$, for example.[23] But Steiner seems happy with proof sketches, given the way he presents Proof 2 above.

[23] In general the n^{th} term of an arithmetic sequence is $a_1 + (n-1)d$, where a_1 is the first term in the sequence and d is the common difference between two successive terms in the sequence.

Now, in both Theorems A and B we use exactly the same proof method from Proof 2 to get to a result about the value of the sum of the first n terms of some other sequence in $\mathbb{N}$. In Steiner's terms, explanation can thus be seen as a relation between the array of Theorems SUM, A and B and Proofs 2, A and B. Indeed, the array can be further expanded to cover all arithmetic sequences in $\mathbb{N}$, since for any sequence of numbers with a constant difference between consecutive terms, the same proof method can be used to get to a result about the sum of the first n terms of that sequence.

Theorem C: The sum of the first n terms of an arithmetic sequence $\{a+(k-1)d\}_{k=1}^{n}$ is equal to $\frac{1}{2}n(2a+(n-1)d)$.

Proof C:

$$\begin{array}{rccccccccc}
S(n): & a & + & a+d & + & \dots & + & a+(n-2)d & + & a+(n-1)d \\
S(n): & a+(n-1)d & + & a+(n-2)d & + & \dots & + & a+d & + & a \\
\hline
2S(n): & 2a+(n-1)d & + & 2a+(n-1)d & + & \dots & + & 2a+(n-1)d & + & 2a+(n-1)d
\end{array}$$

Since there are n terms in the sequence, we have $2S(n) = n(2a+(n-1)d)$, so $S(n) = \frac{1}{2}n(2a+(n-1)d)$.

That is, Proof 2 in fact generalizes to cover all sums of (finite segments of) arithmetic sequences.

So, it seems that Proof 2 meets Steiner's third condition on the basis of characterizing property Q. In the next section, I will look at Steiner's account of characterizing properties in more depth.

3.3 Characterizing properties in more detail

3.3.1 Unique, partial and multiple characterization

In setting out his account, we have seen that Steiner writes 'I shall speak of 'characterizing properties', by which I mean a property unique to a given entity or structure within a family or domain of such entities or structures.' He goes on to say that '... a given entity can be part of a number of differing domains or families. Even in a single domain, entities may be characterized multiply' [Steiner 1978: 143].

Towards the end of the paper, he allows that '... an arbitrary equation with rational coefficients has not a unique Galois group, in the sense that no other equation has it ... The concept of 'characterization' will have to be weakened to allow for partial characterization. The Galois group of E characterizes it in that the properties of the Galois group tell us much about E' [ibid.: 149-50].

So, it seems that Steiner's account allows for unique, multiple and partial characterization. How should we understand these notions? Steiner gives an example: 'Thus, one way of epitomizing the number 18 is that it is the successor of 17. But often it is more illuminating to regard 18 in light of its prime power expansion, 2×3^2' [ibid.: 143].

That is, 18 is uniquely characterized as having prime power expansion 2×3^2. It is also uniquely characterized as being the successor of 17. Hence we see that 18 is multiply characterized in $\mathbb{N}$: there are (at least) two ways of picking 18 out uniquely from other objects in the domain.

Now, let us consider Steiner's remark about partial characterization. I suggest that 18 is partially characterized by being an abundant number, for example, where n is *abundant* if the sum of the divisors of n is at least $2n$. This partially characterizing property 'tells us much' about 18 in the sense that it picks out 18 as a member of the set of abundant numbers, a proper subset of the set of positive integers. Some interesting mathematical results rely on picking out this set, so it seems that the partial characterization is mathematically relevant. For example, it is easy to show that prime numbers are not abundant and that any positive multiple of an abundant number is also an abundant number, and mathematicians including Erdős [1934] have proved various results about the density of abundant numbers in $\mathbb{N}$.

Note that 18 is also partially characterized by being an even number, and by being equivalent to 0 mod 18. So it is clear that multiple partial characterization is also possible.

Let us now examine the characterizing property I suggested in the last section, Q: *being an arithmetic sequence in* $\mathbb{N}$. Property Q is clearly partially rather than uniquely characterizing. It does not pick out just one entity in the set of sequences in $\mathbb{N}$, but rather it picks out all objects with property Q, in the same way that 'being an abundant number' picks out 18 among many other numbers (12, 20, 24, 30, ...).

Although Steiner initially focuses on uniquely characterizing properties, I don't think this is problematic. For one, Q is 'better' at characterizing entities than the commutativity and associativity of addition for integers, as it does not apply to everything in the domain of sequences in $\mathbb{N}$. Property Q allows us to distinguish between number sequences, in a way which is relevant to the proof at hand. And we will need to allow for partially rather than uniquely characterizing properties to make sense of the suggestion that the proof should generalize to other entities with the same property, as stipulated in my reading of Steiner's third condition.

In the next section, I explore my reading in more depth.

3.3.2 Varying the property

Steiner's third condition is stated in his words as 'We should be able to see as we vary the object how the theorem changes in response: the proof is generalizable' [Steiner 1978: 143]. The meaning of 'vary the object' is vague. On my reading, we should take the condition to read as follows: 'Given a certain proof using a characterizing property R of an entity referred to in the theorem, the *proof generalizes* just in case the same argument applied to other objects with characterizing property R is a proof of the modified proposition that is now the conclusion'. The idea is that 'varying the object' means we take a given characterizing property and look for other objects with the same property.

However, other interpretations of Steiner's account in the literature suggest that we should instead consider different (but closely related) properties in order to see whether the theorem generalizes.

For example, Weber and Verhoeven [2002] discuss a case where theorems about right triangles and obtuse angles are proved by holding the proof-idea constant but using a different characterizing property (cosine of right angles versus cosine of obtuse angles). Here the idea of 'varying the object' seems to be to look for closely related objects, where the objects are closely related in virtue of having closely related characterizing properties. This reading seems to fit with one of Steiner's remarks that 'generalizability through *varying* a characterizing property is what makes a proof explanatory' [Steiner 1978: 145, emphasis added].

In order to stay faithful to this remark, I could simply modify my proposal in Section 3.2.2 to suggest an alternative characterizing property, $Q_{1,1}$: 'being (an initial segment of) the arithmetic sequence with initial term 1 and constant difference 1'. To generalize the proof, then, we would simply consider objects with closely related characterizing properties, such as $Q_{1,2}$ 'being (an initial segment of) the arithmetic sequence with initial term 1 and constant difference 2' (the odd numbers), and so on.

However, I want to resist this move because I think my original reading is faithful to the spirit of Steiner's account and helps us to understand some of the remarks he makes about essences. Steiner writes:

> 'My view exploits the idea that to explain the behavior of an entity, one deduces the behavior from the essence or nature of the entity. Now the controversial concept of an essential property of x (a property x enjoys in all possible worlds) is of no use in mathematics, given the usual assumption that all truths of mathematics are necessary. Instead of 'essence', I shall speak of 'characterizing properties" [Steiner 1978: 143].

I will not take a stance here on the controversial topic of essential properties, but I suggest we can see an explanatory proof as one which explains why all objects with a certain nature – the characterizing property – have a certain 'behaviour' pattern (fulfilling the relevant theorem), while all objects lacking this nature do not.

To clarify, my aim is not to propose a reading that is faithful to Steiner's account at all costs; rather, to propose a constructive and charitable reading that captures what seems to me the guiding idea behind his account, while making the account as interesting and persuasive as possible. In the next section I defend my reading of generalizability by showing that it helps to overcome a number of potential worries about Steiner's account.

3.4 Generalizability

It is clear that generalizability is a crucial component of explanation for Steiner; a proof does not count as explanatory if it does not generalize. But suppose a mathematician has proved a result about a certain case. Putting aside cases of simple error, it might take some time for even a successful research mathematician to prove that the result generalizes. In such a case, it would seem odd to forbid the mathematician from classifying the proof as explanatory until she has proved the generalization. Why think a proof that doesn't (yet) generalize can't (yet) be classified as explanatory?

In response, we could argue that as long as what leads us to classify a proof as explanatory is the characterizing feature – in our example, the constant difference between consecutive terms – then it doesn't matter if we don't actually make the generalizing step. This fits with what I take to be the spirit behind Steiner's conception of explanation: it is the characterizing property (the 'essence' of a mathematical entity) which makes something explanatory, not whether we happen to have exploited that characterizing property to its full potential.

But a deeper point still remains. What if there is no further generalization (not just that we haven't discovered one)? It seems very plausible to me that a proof with no further generalization could nevertheless be explanatory. To deny this, I think, needs further argument.

In Section 3.4.2, I show that my reading of generalizability addresses these concerns. First, however, I want to point out another problem with taking generalizability as we usually understand it to be the cornerstone of explanation.

The problem is that generalizability admits of degree: some proofs generalize more widely than others, as I illustrate in Section 3.4.1. And we have seen that explanatoriness and generalizability are closely related for Steiner. Yet Steiner writes that his 'proposal is an attempt at explicating mathematical explanation, not *relative* explanatory value' [Steiner 1978: 143]; so it seems that Steiner's conception of explanation does not admit of degree. In the next section I discuss this apparent mismatch by focusing on a particular example.

3.4.1 Degrees of generalizability and the square root of two

We have already seen two proofs of the fact that $\sqrt{2}$ is irrational. Here we will consider two further proofs.

Theorem: $\sqrt{2}$ is irrational.

Proof: We proceed by contradiction. Suppose $\sqrt{2}$ were rational. Then $\sqrt{2} = \frac{a}{b}$ for $a, b \in \mathbb{N}$ and $b \neq 0$. Then $2 = (\frac{a}{b})^2$ so $2b^2 = a^2$. As Steiner presents the proof:

> '... by using the Fundamental Theorem of Arithmetic – that each number has a unique prime power expansion (e.g. 756 is uniquely $2^2 \times 3^3 \times 7^1$) – we can argue for the irrationality of the square root of two swiftly and decisively. For in the prime power expansion of a^2 the prime 2 will necessarily appear with an even exponent (double the exponent it has in the expansion of a), while in $2b^2$ its exponent must needs be odd. So a^2 never equals $2b^2$, q.e.d..' [Steiner 1978: 137-8]

How does this prime factorisation proof meet Steiner's three criteria on explanation?

1. The proof makes reference to a characterizing property of an entity or structure mentioned in the theorem: as Steiner suggests, 'the prime power expansion of a number is a characterizing property' since by the

Fundamental Theorem of Arithmetic, each number has a unique prime power expansion [ibid.: 138 and 144]. In this case the proof makes reference to the unique prime expansion of the number 2, which is an entity mentioned in the theorem.

2. It is evident from the proof that the result depends on the property, that is if we substitute in a different object of the same domain, the theorem collapses. Here, if we substitute in the number 4, we don't get a result about the irrationality of the square root of 4 because 'the prime power expansion of 4, unlike that of 2, contains 2 raised to an even power, allowing $a^2 = [4]b^2$'. [ibid.: 144] [24]

3. We should be able to see as we vary the object how the theorem changes in response: the theorem is generalizable to cover more numbers. For example, we can substitute in 5 or any other prime, p, to get directly to a result about the irrationality of $\sqrt{p}$. Indeed, we can generalize further to the claim that 'the square root of n is either an integer or irrational ... [and] almost the same reasoning gives us the same for the p^{th} root of n'. [ibid.]

As we see, this example fits Steiner's three criteria. Drawing on earlier discussion, we can easily identify the characterizing property[25]; we can easily find counterexamples and generalizations of the theorem; and both the theorem and proof collapse when we substitute in another entity like 4, since the square root of 4 is rational.

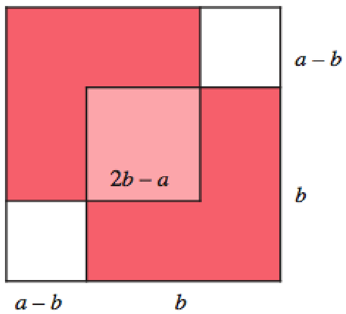
Now, there are many different proofs of the irrationality of the square root of two, and some of these generalize less widely than the one just discussed. Take the following visual proof, presented in [Miller and Montague 2012].

[24] Steiner has 'allowing $a^2 = 2b^2$' here, which must simply be a typographical error.

[25] Here I put aside concerns that the correct characterizing property is a bit more subtle than suggested by Steiner. The proof relies not simply on 2 having a unique prime expansion, but on the fact that the unique prime expansion of a prime number is that prime itself; or more generally that whenever n is not a perfect square, one of the exponents in the unique prime expansion of n is not even.

Tennenbaum's proof

> 'We now describe Tennenbaum's wonderful geometric proof of the irrationality of $\sqrt{2}$. Suppose that $\sqrt{2} = a/b$ for some positive integers a and b; then $a^2 = 2b^2$. We may assume that a is the smallest positive integer for which this is possible. We interpret this geometrically by constructing a square of side a and, within it, two squares of side b ... Since the combined areas of the squares of side b equals the area of the square of side a, the shaded, doubly-counted square must have the same area as the two white squares. We have therefore found a smaller pair of integers u and v with $u^2 = 2v^2$, which is a contradiction. Thus $\sqrt{2}$ is irrational.' [Miller and Montague 2012: 110]



Geometric proof of the irrationality of $\sqrt{2}$

This proof seems a good candidate for an explanatory proof to me (at least as good a candidate as the prime factorisation proof presented by Steiner). Although the proof is an indirect proof by contradiction, I suggest that it does not merely show that it would be impossible for $\sqrt{2}$ to be rational, but illustrates *why* this would be impossible: we would be able to find a concrete counterexample in every case.

However, Tennenbaum's proof does not automatically generalize to cover all primes. Indeed, although Miller and Montague generalize the proof to cover the case $n = 3$, they write that 'For the irrationality of root 5 we have to modify our approach' and they show that further generalizations to cover triangular numbers only work up to $n = 10$ [Miller and Montague 2012: 111-13]. So the proof generalizes less widely than Steiner's example.

I suggest that we can see the new visual proof as posing a dilemma for Steiner. On the one hand, Tennenbaum's proof seems like a good candidate for being an explanatory proof. If this is right, then it will be a point against Steiner's account if his schema for explanatoriness cannot readily accommodate the proof. On the other hand, if Steiner's account can accommodate the proof (given some reasonable characterizing property), then we have two explanatory proofs of the same result that generalize to a different degree. This is a problem if we think, as Steiner seems to, that generalizability tracks explanatoriness and that explanation does not admit of degree.

Now, an easy way out of this dilemma for Steiner would simply be to allow that explanation does after all admit of degree. For example, Steiner's third condition could be modified along the following lines: 'The further a proof generalizes, the greater the degree of explanatory value'.

This would allow Steiner's account to accommodate both proofs of the irrationality of $\sqrt{2}$. For example, we could simply hold that both proofs meet a minimum explanatory threshold: that they generalize to cover three further cases, say. The unique factorisation proof is nevertheless more explanatory than Tennenbaum's proof, since it generalizes to cover many more cases.

Although this may seem like a promising approach, I want to suggest one good reason not to take it. Apart from problems of where to draw the somewhat arbitrary threshold, the problem is that modelling explanatory value directly as a function of generalizability could lead to problems of incommensurability. Different proofs of the same result may generalize not only to different degrees, but also to cover different kinds of cases. This means it may be impossible to directly compare the generalizability of two proofs, in order to determine which is more explanatory.

For example, in one paper Stan Wagon presents fourteen different proofs of a result about tiling a rectangle, comparing and classifying the proofs according to their possible generalizations. Some of these proofs generalize to cover the cylinder, while others generalize to the torus. As Wagon points out, 'no one of the proofs is best in terms of its ability to generalize' [Wagon 1987: 601].

In the next section, I argue that my reading of Steiner's generalizability condition deals neatly with the problems discussed so far.

3.4.2 Advantages of my reading

Recall the reading of generalizability under which I interpreted Steiner's third condition: 'Given a certain proof using a characterizing property R of an entity referred to in the theorem, the *proof generalizes* just in case the same argument applied to other objects with characterizing property R is a proof of the modified proposition that is now the conclusion'.

Note that in fact generalizability does not admit of degree, on this reading. Suppose the characterizing property is R. The theorem generalizes just in case the same argument (resulting in a suitably modified proposition) applies to all other objects with property R. In some cases, there will be many objects with property R. And in some cases, there will be few such objects – perhaps only one. Condition 3 is met if the same argument can be applied to *all* such objects: whether there are many or only one. It is an all-or-nothing condition, and cannot be partially met.

How does this help us to resolve the dilemma posed in Section 3.4.1? Well, it is not the case that the prime factorisation proof that $\sqrt{2}$ is irrational is automatically more generalizable than Tennenbaum's pictorial proof, simply because the former covers more numbers (namely all the primes). Rather, each proof is generalizable just in case the proof's argument can be applied to all cases of objects with the relevant characterising property. In the prime factorisation case, it's obvious that the proof is generalizable in this sense (it's easy to see that substituting in another prime will work). In the pictorial case, this is less obvious and needs further work. Perhaps the pictorial proof is not generalizable in this way. But whether it is generalizable in this sense or not, is not a matter of degree.

So, Steiner can get around the dilemma I presented by claiming that neither explanation nor generalizability admit of degree on the best reading of his three conditions.

My reading of generalizability also has a number of further advantages. First, generalizability as presented here is independent of the interests of persons. So explanatoriness is not relative to persons' abilities and interests, on my reading of Steiner's account; to this extent it is an objective property and does not depend on whether we happen to have discovered the generalization.[26] Second, the generalizability requirement is not as restrictive as might first appear: (i) It does not exclude proofs that cover only one object and cannot be extended to apply to more objects, if the characterizing property applies only to a single object; (ii) In the same vein, the most general version of a proof (like Proof C in Section 3.2.2) will still count as generalizable, as long as it covers all objects with property R.

We might worry whether my strong reading of generalizability matches the usual way we use the term in mathematics. But we can forestall this objection by taking generalizability on my reading to be short for 'generalizability with respect to the relevant characterizing property' and allowing for 'trivial generalizability' in case (i) just mentioned, where the characterizing property applies to just one thing.

One serious disadvantage, however, is that the account so far gives no indication of how explanations fulfil the primary epistemic function of explanations, namely, to help us see why the fact to be explained is true. I will address this problem in the next section, in which I also return to examine Proof 3 from Section 3.1.2 in light of the discussion so far.

3.5 New directions

3.5.1 Ontic and epistemic aspects of explanation

Steiner's account seems at first glance to have both ontic and epistemic components. For example, Steiner stipulates that in an explanatory proof it should

[26]For different readings of objectivity, see for example [Burge 2010: 46-54].

be '*evident* that the result depends on the property', and that '*we should be able to see* as we vary the object how the theorem changes in response' [Steiner 1978: 143, emphasis added], which seems to point to an epistemic aspect of his account.

One important question is what it means for the result depending on the property to be 'evident' on Steiner's account. Recall the discussion of Proof 2 in Section 3.2.2. Why, we might ask, was the importance of property Q not immediately apparent when first analysing Proof 2? It's not clear how we might come to identify property Q except by thinking about how the proof might collapse or generalize to cover other sums of number sequences. This is how I came to identify the property. To a more practised mathematician or mathematics teacher, the required property might become apparent at first glance, perhaps based on familiarity with such results. If this is right, then identifying an appropriate characterizing property seems to build in an epistemic aspect to Steiner's account, where a proof is judged to be explanatory based on the cognitive capacities or background of the reader.

On the other hand, Steiner's account clearly focuses on properties and patterns of dependence, and indeed he rejects another proposed criterion of explanation connected with our ability to visualize on the basis that 'this criterion is too subjective to excite' [Steiner 1978: 143]. So it seems that Steiner's primary aim is to capture an ontic or at least objective account of explanatory proof.

We could maintain an emphasis on the ontic aspect by arguing that a less experienced mathematician may *incorrectly* classify proofs as explanatory (or non-explanatory) based on incorrectly identifying the characterizing property. After all, Steiner does not specify that 'evident' means anything like 'easy to grasp'. Instead, we could simply propose that 'evident' be read as 'evident to a mathematician'; where, of course, further work would be needed to say who counts as a mathematician. In this way, a Steinerian account can maintain that a proof is explanatory in some sense independently of whether the average person actually classifies it as such.

In what follows, I want to go beyond Steiner's account and propose a way to combine both epistemic and ontic aspects, because of the problem raised in the last section: how does a proof's meeting Steiner's three conditions help us to see why the theorem proved is true?

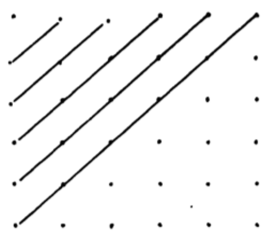
My suggestion is that a proof counts as explanatory in an ontic or objective sense if it in fact meets Steiner's three conditions, involving some suitable characterizing property. We can be justified in calling the proof explanatory if we latch onto the relevant characterizing property (even if, as I suggested earlier, we don't latch on to it in full generality or we don't actually make the generalizing step). The proof also counts as explanatory in an epistemic sense if the property is presented in a way that enables us (or a person with suitably advanced mathematical skills) to latch on to the relevant property.

In my proposed extension of Steiner's account, the primary epistemic function of an explanation is fulfilled to the extent that the proof presents the characterizing property in an accessible way: the more accessible, the more readily we see why the theorem proved is true.

I illustrate this proposal in the next section by going back to examine Proof 3 from Section 3.1.2.

3.5.2 The importance of presentation

I suggest that Proof 3 involves the same characterizing property as in Proof 2, presented in a different way. For ease of reference, I repeat Proof 3:



'By dividing a square of dots, n to a side, along its diagonal, we get an isosceles right triangle containing

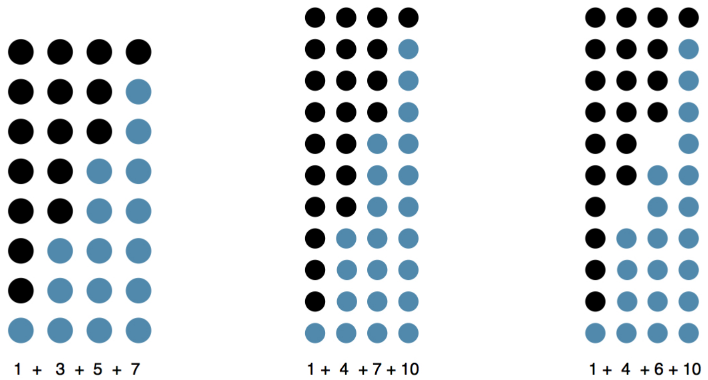
$S(n) = 1 + 2 + 3 + ... + n$

dots. The square of n^2 dots is composed of two such triangles – though if we put the triangles together we count the diagonal (containing n dots) twice. Thus we have

$S(n) + S(n) = n^2 + n$, q.e.d.' [Steiner 1978: 137]

Like Proof 2, Proof 3 also essentially involves counting each element of the sum twice. Here the second sum is upside down, rather than backwards as in Proof 2. But this is just a different geometrical representation of the same idea. The 'geometric' proof, as Steiner calls it, also only works because there is a constant difference between terms in the sum (which are represented by dots).

We could apply the same argument to other instances of arithmetic sequences as in the first two images below.



Whenever the sequence has a constant difference between terms, two dot copies of the sequence will form a rectangle. It's easy enough to see that no rectangle will be formed if there is not a constant difference between terms, as in the third array of dots.

Note that we can't generalize the proof to cover an arbitrary arithmetic sequence as we did with Proof 2, because we can't represent an arbitrary arithmetic sequence using dots. This might tempt us to say that Proof 3 fails to meet Steiner's third condition; but recall the reading of generalizability defended earlier. It's not part of the condition that the argument has to apply to the abstract general description of the case. Rather, the important thing is that the argument can be applied to any individual case of an object with the same characterizing property. This is true (allowing for a broad understanding of applying the argument: in some cases it may be hard to physically draw all of the dots!).

Let me clarify this. The relevant question is: Does the set S of all objects with property Q contain the 'arbitrary arithmetic sequence' $a, a+d, a+2d, \ldots$? If so, the argument in Proof 3 can't be used to cover this arbitrary case as the argument in Proof 2 can (as shown in Proof C, Section 3.2.2). This is because we can't represent $a + d$, for example, using a concrete number of dots in the way we can represent an arbitrary number using a symbol. However, I suggest that the arbitrary sequence is not really a sequence in S; rather, it stands for any sequence in this set, and the argument in Proof 3 does cover each one of these sequences. So Proof 3 meets Steiner's third condition, I suggest.

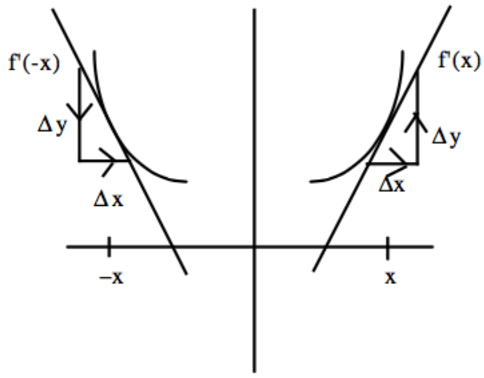
My proposal, then: Proofs 2 and 3 make use of the same characterizing property – being a sequence of numbers with a constant difference between terms. They both meet Steiner's three conditions. What differs between the proofs is the degree to which they make the relevant characterizing property accessible, and thereby the degree to which they achieve the primary epistemic function of an explanation: enabling or helping us to grasp why the proof's conclusion is true.

I find the characterizing property easier to spot and presented more clearly in Proof 2, but this may depend on the reader's cognitive background and preferences. In general, identifying a suitable characterizing property could be an epistemically challenging task, if the property is presented unclearly or in a way not accessible to someone with a particular set of cognitive skills. Such a proof can nevertheless be declared objectively explanatory, according to Steiner's account. It's just that the average reader can't immediately access the property – and hence the explanation.

In the next section, I present a new example which illustrates the same idea.

3.5.3 The derivatives of odd and even functions

Consider the following case:

> **Theorem:** The derivative of an even differentiable function is odd. That is, for any differentiable function f such that $f(x) = f(-x)$, the derivative f' is odd, that is $f'(-x) = -f'(x)$.
>
> **Proof D:** Using limits
>
> An even function is a function where $f(x) = f(-x)$. Using the definition of derivative,
>
> $$f'(-x) = lim_{h\to 0}\frac{f(-x+h)-f(-x)}{h} = lim_{h\to 0}\frac{f(x-h)-f(x)}{h} = -lim_{h\to 0}\frac{f(x-h)-f(x)}{-h} = -f'(x) \ .$$
>
> Then $f'(x)$ is a function such that $f'(-x) = -f'(x)$, as required.
>
> **Proof E:** Using diagrams
>
> 
>
>
> An even function is a function where $f(x) = f(-x)$. This means that even functions are symmetrical about the y-axis: the left-hand side is the mirror of the right-hand side. Therefore, the slope at any point x is the opposite of the slope at $(-x)$. This is just to say that $f'(-x) = -f'(x)$, as required. [Raman 2002: 20]

Clearly, E is more of a proof sketch, but we have seen that Steiner seems quite happy with proof sketches. Both proofs fit Steiner's scheme as follows:

1. The proof makes reference to a characterizing property of an entity or structure mentioned in the theorem

In both proofs, the characterizing property is that of *being an even function*, and the property holds of an entity mentioned in the theorem, namely the function f.

In Proof D, the characterizing property is represented as follows: (P_D) $f(x) = f(-x)$. The property is stated in this form in the theorem. Proof D makes use of (P_D) together with the limit definition for the derivative of f.

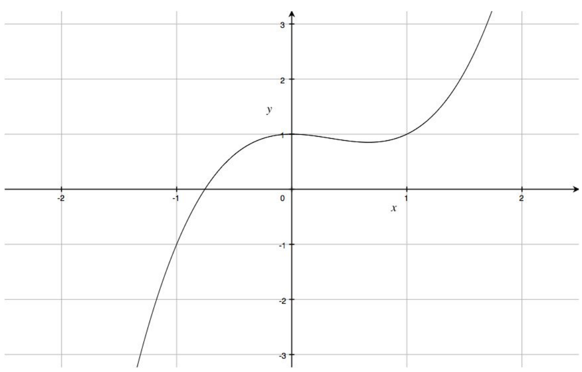
In Proof E, the characterizing property is represented as follows: (P_E) f is symmetric over the y-axis. Proof E makes use of (P_E) together with a graphical representation of the derivative of f.

2. It is evident from the proof that the result depends on the property, that is if we substitute in a different object of the same domain, the proof or theorem collapses.

In both proofs, we can substitute in a specific entity that lacks the characterizing property – that is, a differentiable function which is not even – and we can see that the derivative of the function is not odd. Take for example the function $g(x) = x^3 - x^2 + 1$. Note that $g(-x) = -x^3 - x^2 + 1$.

We have $g'(x) = lim_{h \to 0} \dfrac{g(x-h) - g(x)}{h}$ while $g'(-x) = lim_{h \to 0} \dfrac{g(-x+h) - g(-x)}{h}$. These limits are not equal in general.

Looking at the graphical representation of function g, we can see that the gradient of g does not have rotational symmetry around the origin, as would be required for g' to be an odd function.



3. We should be able to see as we vary the object how the theorem changes in response: the theorem is generalisable.

Both proofs are already maximally general, in the sense that they cover the case of an arbitrary even differentiable function, $f(x)$. But we can substitute in any specific even differentiable function – that is, any object of the domain which satisfies the property – and the proof will go through.

Now, let us consider how this example supports my suggestion in the previous section. The characterizing property here is the 'evenness' of the original function. In Proof D this is represented as follows: $f(x) = f(-x)$. The proof proceeds by symbol manipulation of the limit definitions for derivatives. In Proof E the evenness of the function is represented in the symmetry of the graphical representation: the graph of $f(x)$ is symmetrical about the y-axis. That is, the same characterizing property is presented differently in each case. We might even say the two cases are simply different presentations of the same proof: for example, Raman quotes a mathematics professor saying that 'If I were going to use that picture, I would take it and turn it into a proof. Although if you do that, it comes down to pretty much [Proof D]' [Raman 2002: 28].

In the form that Steiner presents his account, he cannot make sense of any perceived difference in explanatory value between the two proofs (or proof presentations). Both meet his three conditions on explanation. But on my proposed extension of Steiner's account, the perceived difference in explanatory value lies in the differing degrees of epistemic access which each proof presentation affords us to the characterizing property.

I find the characterizing property easier to spot and presented more clearly in Proof E, but a working mathematician may not need the diagram to access the property. For example, Raman cites the same mathematics professor saying that 'If I had had trouble writing up the analytic solution, then I would have drawn myself a picture.' [ibid.]

To summarise, I hope my proposal has provided reason to think that a successful development of Steiner's account should include both an ontic component – whether or not the proof contains a suitable characterizing property that meets conditions 1–3 – and an epistemic component – whether the characterizing property is presented in a way accessible to a reader with a certain cognitive background.

In Chapter 5, I will build on this suggestion to provide a new start to my own account of explanation in mathematics, adding in insights gained from examining Lange's account in Chapter 4.

Before moving on to Chapter 4, I briefly return to examine Proof 1, which involved mathematical induction, in light of my arguments from Chapter 2.

3.6 Proof by induction

At the beginning of Chapter 2, I noted that Hafner and Mancosu pose a dilemma for Steiner's account of explanation on the grounds that his account either overgenerates by counting inductive proofs as explanatory, or undergenerates by ruling out a promising example from mathematical practice [Hafner and Mancosu 2005: 237]. In Chapter 2, I argued that proofs involving mathematical induction can sometimes be explanatory. If this is correct, then it should not be a problem for Steiner's account of mathematical explanation if his proposal counts an inductive proof as explanatory, *pace* Hafner and Mancosu.

However, Steiner himself seems reluctant to count inductive proofs as explanatory. Recall Proof 1, presented at the start of this chapter, which involves mathematical induction.

> **Theorem:** The sum of positive integers from 1 to n is equal to $\frac{n(n+1)}{2}$.
>
> **Proof 1** (strictly speaking, this is only the inductive step):
>
> $S(n+1) = S(n) + (n+1) = n(n+1)/2 + 2(n+1)/2 =$
>
> $(n+1)(n+2)/2.$

According to Steiner, this proof is 'not very explanatory', and the proof makes no reference to a characterizing property. We might think that the set of natural numbers is characterized by the principle of induction, but Steiner denies that this principle can count as the characterizing property: 'The proof by induction does not characterize anything mentioned in the theorem. Induction, it is true, characterizes the set of all natural numbers; but this set is not mentioned in the theorem' [Steiner 1978: 145]. This response from Steiner is odd, since most formal proofs do mention the set $\mathbb{N}$ in the theorem, starting '$\forall n \in \mathbb{N}$... '. On a face value reading, Steiner's account then seems to be extremely dependent on the choice of set-up.

Hafner and Mancosu make the same point, writing that 'it seems quite odd that Steiner's theory qua theory of the explanatoriness of proofs should turn out to be so overly sensitive to what appears to be a rather minor detail in the exact wording of a theorem which doesn't affect its proof'. [Hafner and Mancosu 2005: 237].

Although Steiner could get round this problem by allowing that inductive proofs can be explanatory, he again appears reluctant to do so, arguing that even if we could find a suitable characterizing property,

> '... a characterizing property is not enough to make an explanatory proof. One must be able to generate new, related proofs by varying the property and reasoning again. Inductive proofs usually do not allow deformation, since before one reasons one must have already conjectured the theorem. Changing the equations for S will not immediately reevaluate $S(n)$ – it must be conjectured anew.' [Steiner 1978: 151, endnote 11].

Steiner's suggestion here seems to fit with my suggestion in Chapter 2 that in cases of non-explanatory inductive proof, we often fail to see where the equation 'comes from'. That is, the non-explanatory inductive proof gives us a method to verify that the result is true, but seems to leave it mysterious how the result could be discovered. As we saw earlier, deformations of Proof 2 do not face this problem: when we insert a new arithmetic sequence into the diagram, we can immediately see what form the sum $S(n)$ will take.

Now, recall that my reading of Steiner's generalizability condition took the condition to be independent of the interests and abilities of persons. To this extent I took generalizability to be an objective property which does not depend on whether we happen to have discovered the generalization.

It is interesting therefore to note that Steiner's response to the 'problem' of inductive proof raises an epistemic point. His concern is not that the inductive proof does not generalize to cover new results, but rather that the proof does not give the reader enough information to extend the proof – 'one must already have conjectured the theorem' [ibid.]. Steiner's response to the 'problem' of inductive proof thus suggests that I was right to look for an epistemic aspect to his account.

I will explore these issues in more depth in Chapter 5, where I build on the insights gained here to make a start towards my own account of mathematical explanation.

3.7 Concluding remarks

I have attempted to give a maximally charitable reading of Steiner's account, one that makes sense of the puzzling comments he makes about his sum-of-integers example. I began by analysing one of Steiner's three proofs (Proof 2), and I identified a characterizing property in order to show that the proof does indeed meet Steiner's three conditions and hence can be called explanatory on his account. I raised a few potential worries about Steiner's account, and showed how my reading of his third generalizability condition helps Steiner to avoid these problems.

Then, I examined Proof 3 and suggested that a proof may display its characterizing property more or less clearly, allowing for a new epistemic component to Steiner's account that lays the ground for my work in Chapter 5. Finally, I returned to look at Proof 1 and suggested that Steiner's response to the 'problem of induction' supports my extension of Steiner's account to include an epistemic component.

In the next chapter, I examine a more recent account of mathematical explanation proposed by Marc Lange.

4 Chapter 4: Salient features

In this chapter, I will explore Marc Lange's account of mathematical explanation, presented in [Lange 2014] and [Lange 2017]. According to Lange,

> 'What it *means* to ask for a proof that explains is to ask for a proof that exploits a certain kind of feature in the setup – the same kind of feature that is salient in the result. The distinction between proofs that explain why some theorem holds and proofs that merely establish that it holds exists only when some feature of the result being proved is salient. That feature's salience makes certain proofs explanatory. A proof is accurately characterized as an explanation (or not) only in a context where some feature of the result being proved is salient.' [Lange 2014: 507, emphasis in the original]

I will capture this view as follows: A proof is explanatory only if (1) the result, R, exhibits some salient feature, S_R; (2) the proof, P_R, exploits a salient feature, S_{PR}; and (3) the two salient features, S_R and S_{PR}, are similar.

The reason for these three conditions is that the two features salient in the result and exploited in the proof are not always precisely the same feature – for example, in some of Lange's examples they are different types of symmetry. I therefore take Lange's salient feautures to be broad types that decompose into more specific types, together with some kind of similarity relation.

Lange identifies three salient features in particular: unity, symmetry and simplicity. I will explore each feature in detail in Sections 4.1 to 4.3 by analysing some of Lange's proposed examples for each case, as well as presenting some problem cases for Lange.

Now, I said in Chapter 1 that I would be exploring Lange's account as an example of an epistemic why-question account of mathematical explanation. I think this fits with Lange's view – for example, Lange describes a salient feature as 'prompting a why question answerable by a proof deriving the result from a similar feature of the given' [Lange 2014: 507]. He rejects the suggestion that a salient feature is 'some kind of structure that lies behind the phenomena to be explained' because 'talk of "structure" tends to sound as though it is talking about a feature of the case that, by virtue of underlying the phenomena whatever the context, is explanatory whatever the context' [Lange 2017: 448], and he writes that 'if some extraterrestrials differ from us in which features of a given theorem they find salient, then it follows from my account that those extraterrestrials will also differ from us in which proofs they ought to regard as explanatory. I embrace this conclusion.' [Lange 2014: 525]

However, in the same way I extended Steiner's account in Chapter 3 to include both ontic and epistemic aspects, I will follow a similar approach here – I suggest that the salient features Lange proposes are best seen as objective features, in the sense of being independent of context, while the *salience* of the feature is relative to our cognitive capacities and hence context-dependent.

4.1 Unity and symmetry

4.1.1 Triples of integers

Consider the following result: the product of any three consecutive nonzero natural numbers is divisible by 6. Lange gives two proofs.

> **Proof by induction**
>
> 'The product of 1, 2, and 3 is 6, which is divisible by 6.
>
> Suppose that the product of $(n–1)$, n, and $(n+1)$ is divisible by 6. Let's show that the product of n, $(n+1)$, and $(n+2)$ is divisible by 6. By algebra, that product equals $n^3+3n^2+2n = (n^3–n)+3n(n+1)$. Now $(n^3–n) = (n–1)n(n+1)$, so by hypothesis, it is divisible by 6. And $n(n+1)$ is even, so $3n(n+1)$ is divisible by 3 and by 2, and therefore by 6. Hence, the original product is the sum of two terms, each divisible by 6. Hence, that product is divisible by 6.' [Lange 2014: 510]

Unified proof

> Of any three consecutive nonzero natural numbers, at least one is even (that is, divisible by 2) and exactly one is divisible by 3. Therefore, their product is divisible by $3 \times 2 = 6$. [ibid.]

Here the proof by induction divides the products of three consecutive numbers into two classes, handled in the base case and inductive step. Yet, according to Lange,

> 'Insofar as we found the theorem remarkable for identifying a property common to every triple of consecutive nonzero natural numbers, our point in asking for an explanation was to ask for a proof that treats all of the triples alike. This feature of the theorem is made especially salient by a proof that does treat all of the triples alike ... [the unified proof] proceeds entirely from a property possessed by every triple' [Lange 2014: 510].

Here the striking fact or salient feature is supposed to be that the result holds for all triples, which means an explanatory proof must be a unifying proof that exploits a property common to all triples.

To clarify, the result displays unity not by stating that each triple has some property of *unity* – it's not clear what that would be – but by stating that each triple has a certain property, P, *in common*, where in this case $P(a, b, c) \leftrightarrow 6|abc$. It seems to me that I was initially struck by the fact that the product was divisible by 6, rather than some other number – but Lange might suggest I was really struck (or should have been struck) by the fact that the product of *every* triple is divisible by 6.

We can then describe the unified proof as exploiting unity, not in virtue of showing that each triple displays some property of unity, but in virtue of identifying another property Q that each triple has in common. The relevant property Q in this case is that (at least) one of the members of each triple is even, and one of the members is divisible by three – or more generally, that any sequence of k consecutive positive integers is divisible by k.

The unified proof shows that every triple has property Q and that $Q \to P$, leading to the required result that every triple has property P.

As I suggested at the start, I think it's plausible to see Lange's feature of unity as an objective one here – it's a property P that holds of the domain independent of context, namely that each element has some property Q in common. However, our attention being drawn to this feature is dependent on context – the proof by cases draws our attention to the unified nature of the domain, creating a context where unity becomes a salient feature.

Now, one potential worry is that there are surely many proofs of universal generalisations that have this structure and do not seem to count as explanations. So it is important to note that Lange's three conditions are taken only to be necessary – and not sufficient – conditions on explanatory proof.[27]

But this leads to a further worry – Lange's conditions might be necessary for and yet not account for the explanatory value of a given proof. After all, one of the necessary conditions for being an explanatory proof is being a proof; yet being a proof does not itself account for explanatory value, as we saw in the motivating quotes given in the Introduction.

The next example, also one of Lange's, shows that even when a proof exploits unity in the way described above, it is the specific details of the properties involved – P and Q – that seem to account for the proof's explanatory value, rather than P and Q's holding for every entity in the domain.

4.1.2 The sum from 1 to n

Consider the following (by now very familiar) theorem about sums of positive integers: The sum of positive integers from 1 to n is equal to $n(n+1)/2$. As before, Lange considers two proofs.

Proof by cases

[27] As a reminder, the conditions were captured as follows: A proof is explanatory only if (1) the result, R, exhibits some salient feature, S_R; (2) the proof, P_R, exploits a salient feature, S_{PR}; and (3) the two salient features, S_R and S_{PR}, are similar. In the case of unity, the two salient features are not merely similar but the same (having a property in common with all other entities in the domain).

'There are two cases.

When n is even, we can pair the first and last numbers in the sequence, the second and second-to-last, and so forth. The members of each pair sum to $n + 1$. No number is left unpaired, since n is even. The number of pairs is $n/2$ (which is an integer, since n is even). Hence, $S = (n + 1)n/2$.

When n is odd, we can pair the numbers as before, except that the middle number in the sequence is left unpaired. Again, the members of each pair sum to $n + 1$. But now there are $(n–1)/2$ pairs, since the middle number $(n + 1)/2$ is unpaired. The total sum is the sum of the paired numbers plus the middle number: $S = (n + 1)(n–1)/2 + (n + 1)/2$. This simplifies to $(n + 1)n/2$—remarkably, the same as the expression we just derived for even n.' [Lange 2014: 511-12]

'Standard' Gaussian proof

$S = 1 + 2 + \ldots + (n–1) + n$

$S = n + (n–1) + \ldots + 2 + 1$

If we pair the first terms, the second terms, and so forth, in each sum, then each pair adds to $(n + 1)$, and there are n pairs. So $2S = n(n + 1)$, and hence $S = n(n + 1)/2$. [ibid.]

According to Lange, the proof by cases draws our attention to the fact that the result holds for both even and odd n. Although this fact may not initially have seemed remarkable, 'this feature of the result strikes us forcibly' and hence it becomes a salient feature for which we seek an explanation. The standard proof shows that this feature is 'no coincidence ... [and] traces the result to a property common to the two cases: that the terms are balanced around $(1 + n)/2$'. [Lange 2014: 513].

In this case, the result displays unity by stating that property P holds for all sequences $1, 2, .., n$, where $P(1, 2, ...n) \leftrightarrow 1 + 2 + 3 + ... + n = n(n + 1)/2$.

The result exploits unity by identifying a property Q common to all sequences $1, 2, .., n$, namely $Q(1, 2, ..., n) \leftrightarrow$ (the nth terms in two copies of S pair to $(n+1)$), and showing that for all n, $Q(1, 2, ..., n) \rightarrow P(1, 2, ..., n)$.

But, I suggest, the explanatory value comes in the details of property Q: specifically, in the symmetry involved in Q. As I suggested in the previous chapter, the commutativity and associativity of addition ensure that the sum and its 'mirror image' sum to the same result in each case[28], while the constant difference between terms in the sequence ensures that each of the n terms of the double sum are the same.

We might try to force these properties of Q to count as unity rather than symmetry properties, but Lange himself admits that there is not such a clear difference between the two:

> 'Strictly speaking, that a given result identifies a property common to every single case of a certain sort is just a symmetry in the result ... My view does not require that a theorem's displaying a striking symmetry be sharply distinguished from a theorem's being striking for its treating various cases alike' [Lange 2014: 508]

I suggest that, even if not strictly distinguished from unity, it is symmetry that plays an explanatory role in the Gaussian proof. After all, we could prove the same result using strong induction – which treats each case alike – and yet Lange himself claims that inductive proofs are never explanatory.[29] Similarly, even if a proof that handles all cases alike is generally preferable over a proof by cases, it's not clear that this is because of an increase in *explanatory* power: perhaps a unifying proof is neater or more elegant, or exhibits other epistemic virtues such as surveyability, as in the case of the Four Colour Theorem discussed in Chapter 2 (Section 2.2.2).

Now, I have already mentioned that Lange identifies symmetry as a salient

[28]I rejected the commutativity and associativity of addition as a characterizing property in Steiner's sense, but it could still serve as a salient symmetry on Lange's account, because Lange doesn't stipulate that the salient feature must be one that helps us to generalise the proof.

[29]More on this in Section 4.4, where I revisit Lange's views on induction.

feature as well as unity, so this might not seem particularly worrying. Let us therefore put unity aside for now, and focus on an example involving symmetry.

4.1.3 Complex conjugation

Lange presents the following theorem of complex analysis, which states that all of the nonreal roots of polynomials with real coefficients come in complex-conjugate pairs:

> **D'Alembert's theorem**
>
> If the complex number $z = a + bi$ (where a and b are real) is a solution to $a_n z^n + a_{n-1} z^{n-1} + \ldots + a_0 = 0$ (where the a_i are real), then z's "complex conjugate" $\bar{z} = a–bi$ is also a solution. [Lange 2014: 498]

One proof of the theorem 'pursues what mathematicians call a "brute force" approach, as follows:

> We can prove this theorem directly by evaluating $a_n \bar{z}^n + a_{n-1} \bar{z}^{n-1} + \ldots + a_0$.
>
> First, we show by calculation that $\bar{z}\bar{w} = \overline{zw}$:
>
> Let $z = a + bi$ and $w = c + di$. Then $\bar{z}\bar{w} = (a - bi)(c - di) = ac - bd + i(-bc - ad)$ and $zw = (a + bi)(c + di) = ac - bd + i(bc + ad)$ so $\overline{zw} = ac - bd - i(bc + ad) = \bar{z}\bar{w}$.
>
> Hence, $\bar{z}^2 = \bar{z}\bar{z} = \overline{zz} = \overline{z^2}$, and likewise for all other powers. Therefore, $\bar{z}^n + a_{n-1}\bar{z}^{n-1} + \ldots + a_0 = \overline{z^n} + a_{n-1}\overline{z^{n-1}} + \ldots + a_0$.
>
> Now we show by calculation that $\bar{z} + \bar{w} = \overline{z + w}$:
>
> Let $z = a + bi$ and $w = c + di$. Then $\bar{z} + \bar{w} = (a - bi) + (c - di) = a + c + i(-b - d)$ and $\overline{z + w} = \overline{a + bi + c + di} = a + c - i(b + d) = \bar{z} + \bar{w}$.
>
> Thus, $\overline{z^n} + a_{n-1}\overline{z^{n-1}} + \ldots + a_0 = \overline{z^n + a_{n-1}z^{n-1}} + ... + a_o$, which equals $\bar{0}$ and hence 0 if z is a solution to the original equation.' [Lange 2014: 498]

This proof 'simply calculates everything directly, plugging in everything we know and grinding out the result' [ibid.]. Yet, writes Lange, 'the striking feature of d'Alembert's theorem is that the equation's nonreal solutions all come in pairs ... Why does exchanging i for $-i$ in a solution still leave us with a solution? ... we are inclined to suspect that there is some reason for [this symmetry]' [ibid.: 498].

Lange suggests that the 'reason' for this symmetry is that

> '$-i$ could play exactly the same roles in the axioms of complex arithmetic as i plays. Each has exactly the same definition: each is exhaustively captured as being such that its square equals -1. ... Whatever the axioms of complex arithmetic say about one can also be truly said about the other. Since the axioms remain true under the replacement of i with $-i$, so must the theorems' [ibid.: 499].

So, we are presented with a striking fact: that complex roots of real polynomials come in symmetrical pairs, and an explanatory proof[30]: one which exploits the fact that i and $-i$ are an essentially interchangeable, and hence symmetric, pair.

Lange does not provide an explicit account of symmetry, but we can simply take the mathematical idea of symmetry as an operation or mapping under which some property or relevant formal structure remains invariant. This is a broad type that decomposes into more specific types, such as geometrical symmetries – reflection in a line, rotation by 90 degrees, etc.

We could take two symmetries to be *similar* when they are located in the same domain (e.g. both geometric), or when they are isomorphic (e.g. different presentations of the same operation). Lange writes that 'Admittedly, several fairly elastic notions figure in my idea of a proof's exploiting the same kind of feature in the problem as was salient in the result' [Lange 2014: 524]. So it seems he does not have a specific relation of similarity in mind.

[30]Strictly speaking, it is just a proof sketch. But Lange's conditions on explanation could be easily expanded to allow for a proof sketch; alternatively, Lange might simply say that a proof sketch is explanatory if it points towards a proof that would fulfil his three conditions (and, perhaps, indicates *how* the proof would fulfil the three conditions).

In the current example, the relevant symmetry is complex conjugation: an automorphism of the complex number plane (indeed, the only non-trivial automorphism of $\mathbb{C}$). Complex conjugation is clearly featured in the result. The proof also exploits complex conjugation, since the symmetric relation between i and $-i$ is precisely the reason that complex conjugation forms a structure-preserving map of $\mathbb{C}$. In this case, then the salient symmetries in the proof and in the result are not just similar, but the same.

So far, then, it seems that symmetry makes for a plausible salient feature on Lange's account. I agree with Lange that symmetry can add explanatory power.[31]

However, I will consider some problem cases for Lange's account in the next section.

4.2 Problem cases

Recall my reading of Lange's three conditions: a proof is explanatory only if (1) the result, R, exhibits some salient feature, S_R; (2) the proof, P_R, exploits a salient feature, S_{PR}; and (3) the two salient features, S_R and S_{PR}, are similar.

The potential problem cases fall into two kinds.

> Type A: The result exhibits a salient feature, but there is an explanatory proof that does not exploit a similar salient feature.
>
> Type B: There is an explanatory proof that exploits a salient feature, but where the result does not exhibit a similar salient feature.

These are genuine worries (if such cases exist), because we have seen that Lange's three conditions are intended to be necessary conditions on explanation. For example, Lange writes that:

> 'The distinction between proofs that explain why some theorem holds and proofs that merely establish that it holds exists only when some feature of the result being proved is salient. That feature's

[31] More on my view in Chapter 5.

> salience makes certain proofs explanatory. A proof is accurately characterized as an explanation (or not) only in a context where some feature of the result being proved is salient.' [Lange 2014: 507]

I will consider cases of type A and B in turn.

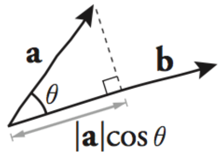
4.2.1 The scalar product

First, an example of type A. Mathematician Victor Blåsjö considers the result that the geometric and coordinate forms of the scalar product are equal, and claims that this result displays symmetry as a salient feature. Blåsjö writes:

> 'In my calculus book I prove this [result] in a cognisable way. I start with the geometrical idea of a projection, and I show through intuitive-visual reasoning how this leads to the coordinate form. This is an explanatory proof, in my opinion.
>
> The standard proof in other textbooks is to start with the coordinate form and derive the cosine form using the law of cosines. This is obviously extremely unsatisfactory, since the law of cosines is a "black-box," algebraic hocus-pocus result. This is a prototypically non-explanatory proof, in the sense of the quotations from Lange above.
>
> Yet according to Lange's proposal it is the latter proof that is the explanatory one. For the result is certainly symmetric, and the standard proof likewise respects this symmetry throughout.
>
> My proof, on the other hand, is most definitely asymmetrical: it involves the projection of one vector onto another, an asymmetrical relation. Thus my proof actually introduces an asymmetry that was not in the result itself. And it was precisely this move that made the proof cognisable, and thereby explanatory, in my view. Which is the exact opposite of what Lange's proposal says should happen.' [Blåsjö 2016]

Blåsjö's exposition runs as follows:

'Vectors can be used to express projection in a very convenient way.

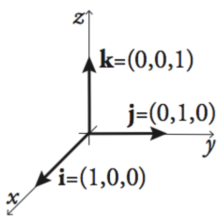


The length of **a**'s projection onto **b** is easily expressed trigonometrically. It is $|\mathbf{a}|cos\theta$, where $|\mathbf{a}|$ means the length of the vector a (just as absolute value always means distance to the origin, or simply magnitude). The remarkable thing is that the length of the projection can also be expressed in a very simple way in terms of the coordinates of the vectors. This is codified in the scalar product

$$\mathbf{a}.\mathbf{b} = |\mathbf{a}||\mathbf{b}|cos\theta = a_1b_1 + a_2b_2 + a_3b_3$$

When **b** is a unit vector (i.e., has length 1) the middle expression is precisely the length of the projection, so the formula tells us that we can find it simply by multiplying the vectors component-wise and adding the results. Nothing could be easier.

What is the reason behind this magical harmony of geometry and algebra? To see this it is useful to introduce the unit vectors **i**, **j**, **k** pointing in the direction of the axes:



Then by breaking up a and b into their coordinate components we get

$$\mathbf{a}.\mathbf{b} = (a_1\mathbf{i} + a_2\mathbf{j} + a_3\mathbf{k}).(b_1\mathbf{i} + b_2\mathbf{j} + b_3\mathbf{k})$$

> Now, the projection properties of **i**, **j**, **k** are particularly simple: any one of them projected onto itself gives 1, and projected onto each of the other two gives zero. Therefore when we multiple out the parenthesis all the cross terms go away and only the "like with like" terms survive. So the result is $a_1b_1 + a_2b_2 + a_3b_3$, as claimed.' [Blåsjö 2018: 62-3]

Lange could simply deny that Blåsjö's proof sketch counts as explanatory, but this would need some argument since Blåsjö is a practising research mathematician, and his suggestion that the geometrical proof adds explanatory value is backed up by other mathematicians, in pedagogical contexts at least. For example, mathematicians Dray and Manogue 'argue for pedagogical reasons that the dot and cross products should be defined by their geometric properties, from which algebraic representations can be derived, rather than the other way around' [Dray and Manogue 2006: 1].

Although Dray and Manogue cite pedagogical concerns, they provide mathematical reasons in support of their view, pointing out that 'the Law of Cosines follows immediately from the geometric definition of the dot product, in direct contrast to the traditional treatment, in which the order is reversed' [ibid.: 11].

Lange is interested (as am I) in an epistemic account of explanation that allows for the context and concerns of the observer, so it seems to me that Lange should take these supporting quotes seriously, despite the pedagogical context. So I doubt Lange would choose to deny that Blåsjö's asymmetrical proof counts as explanatory.

I think the best response for Lange here is to argue that Blåsjö's proof does in fact exploit a symmetry. Although Blåsjö says the proof is 'most definitely asymmetrical: it involves the projection of one vector onto another, an asymmetrical relation', the details of the proof in fact make use of the fact that i, j and k are defined to give 1 when projected onto themselves, and 0 when projected pairwise (they form an orthonormal basis).

As Dray and Manogue put it, 'The geometric definition is coordinate independent, and therefore conveys invariant properties of these products' [Dray and Manogue 2006: 1].

In particular, the result of the scalar product is invariant under rotation, and this is exactly the kind of symmetrical feature Lange could point to as a salient feature exploited by the proof.

So, I think Lange can get around this potential problem case fairly easily. But the next case I will consider is more problematic.

4.2.2 Gaussian sums

I will consider two examples of type B: explanatory proofs that exploit a salient feature, but where the results do not exhibit a similar salient feature.

We have already seen an example of the sum from 1 to n:

> **Theorem:** For any positive integer n, the sum from 1 to n is $n(n+1)/2$.
>
> **Proof:**
>
> $$\begin{array}{ccccccccccc} 1 & + & 2 & + & 3 & + & ... & + & n & = & S \\ n & + & (n-1) & + & (n-2) & + & ... & + & 1 & = & S' = S \\ \hline (n+1) & + & (n+1) & + & (n+1) & + & ... & + & (n+1) & = & n(n+1) \end{array}$$
>
> Since there are n terms in the sequence, we have $2S(n) = n(n+1)$ and hence $S(n) = \frac{n(n+1)}{2}$.

Consider additionally a very similar example involving co-prime integers, also due to Gauss:

> **Theorem:** For any integer $n > 1$, the sum of the positive integers less than and co-prime to n is $(\varphi(n).n)/2$, where $\varphi(n)$ is the number of integers less than and co-prime to n.
>
> **Proof:** Let $b_1, \ldots, b_{\varphi(n)}$ be the positive integers less than and co-prime to n (in order of size). Now $gcd(b, n) = 1$ iff $gcd(n-b, n) = 1$, and so the numbers $b_1, \ldots, b_{\varphi(n)}$ are (in the same order) the numbers $n - b_{\varphi(n)}, \ldots, n - b_1$. We can write the second representation of the sum backwards, and follow the same proof method as before:

$$\begin{array}{ccccccccccc}
b_1 & + & b_2 & + & b_3 & + & ... & + & b_{\varphi(n)} & = & S \\
n - b_1 & + & (n - b_2) & + & (n - b_3) & + & ... & + & n - b_{\varphi(n)} & = & S' = S \\
\hline
n & + & n & + & n & + & ... & + & n & = & n.\varphi(n)
\end{array}$$

Since there are n terms in the sequence, we have $2S(n) = n.\varphi(n)$ and hence $S(n) = \frac{n.\varphi(n)}{2}$.

In both cases, the double sums are symmetric in the sense that the outcome is invariant under a change in the order in which we add up the terms.

Both proofs also exploit a further symmetry in the double sum, namely that each vertical term in the double sum has the same value. In the first case this symmetry is underpinned by a constant difference between terms in the sequence, while in the second case the symmetry is underpinned by the fact that $gcd(b, n) = 1$ iff $gcd(n - b, n) = 1$.

However, it doesn't seem that symmetry is a salient feature in either of the results above; certainly, symmetry is not something that strikes or jumps out at me when I read the theorems. This is a problem because it seems plausible to me that we should count both proofs as explanatory, even though they do not seem to meet Lange's three conditions. This suggests that Lange's three conditions are not necessary conditions on mathematical explanation.

Now, Lange writes that

> 'My proposal predicts that if the result exhibits no noteworthy feature, then to demand an explanation of why it holds, not merely a proof that it holds, makes no sense. There is *nothing that its explanation over and above its proof would amount to* until some feature of the result becomes salient.' [Lange 2014: 507, emphasis added]

So it looks like Lange goes so far as to say that there is no such thing as an explanatory proof if the result exhibits no salient feature, such as symmetry.

Later, however, Lange allows that the proof itself can draw our attention to the relevant salient feature: 'A proof may focus our attention on a particular feature of the result that would not otherwise have been salient.' [Lange 2014: 511]

So his earlier claim was a bit too quick: presumably Lange thinks that even if a result exhibits no noteworthy feature, there is still something an explanation would amount to, namely a proof that both draws our attention to a noteworthy feature and exploits it.

Let us apply this concession to our potential problem case. For example, even though we were not initially struck by a symmetry in the sums-of-integers result, we might allow that Gauss's proof draws our attention to the relevant symmetries present in the entities featured in the result. Once our attention is drawn to this symmetry, an explanatory proof is one that exploits the symmetry, and this allows us to categorise Gauss's symmetry proof as explanatory on Lange's account.

Although this seems a promising response for Lange, I would still question whether the symmetries I mentioned are truly features of the result. I suggested that a symmetry present in both proofs was invariance under transformation of the order in which we add up the numbers. But what makes this a particularly salient feature of the fact that the sum is $n(n+1)/2$, for example? The same symmetry holds for any sum in $\mathbb{N}$, so the feature seems too broad to usefully mark out results that call for an explanation from those where no explanation is desired or possible.

Moreover, in the second example of co-prime integers it seems less plausible that the proof draws our attention to a symmetry in the result. What symmetry exactly is present in the claim that for any integer $n > 1$, the sum of the positive integers less than and co-prime to n is $(\varphi(n).n)/2$?

In response to this worry, it looks like Lange is happy to bite the bullet and simply deny that the symmetries present in the two cases above play an explanatory role. He writes that:

> 'my proposal entails that if an explanatory proof appeals to a symmetry in the setup, but the result being explained fails to exhibit any similar striking symmetry, *then the proof's explanatory power does not arise from its appeal to symmetry*' [Lange 2017: 267, emphasis added].

The idea is that the explanatory power of the proof must arise from its appeal to some other salient feature, one that is also displayed in the result.

But here again we run into a problem, as it is far from clear what other salient feature Lange might propose to handle these cases. I leave the burden of proof to Lange here.

In the next section, I consider a case where Lange does identify a specific alternative salient feature, arguing that his attempt fails.

4.2.3 Counting digits

Consider the following result, familiar from Chapter 1:

> 'Suppose you write down all of the whole numbers from 1 to 99,999. How many times would you write down the digit 7? The answer turns out to be 50,000 times. This is a striking result: 50,000 is almost exactly $\frac{1}{2}$ of 99,999'. [Lange 2017: 267]

Lange presents the following proof by Dreyfus and Eisenberg as an explanatory proof:

> 'Include 0 among the numbers under consideration – this will not change the number of times that the digit 7 appears. Suppose all of the whole numbers from 0 to 99,999 (100,000 of them) are written down with five digits each, e.g., 1306 is written as 01306. All possible five-digit combinations are now written down, once each. Because every digit will take every position equally often, every digit must occur the same number of times overall. Since there are 100,000 numbers with five digits each – that is, 500,000 digits – each of the 10 digits appears 50,000 times. That is, 100,000 x 5 / 10 = 50,000'. [Dreyfus and Eisenberg 1986: 3]

According to Dreyfus the explanatory proof 'appeals to the symmetric role played by all nonzero digits in the list of numbers', and this seems right to me: it would be natural to wonder, on first reading the Dreyfus-Eisenberg result, whether 7 is special in some way, and the explanatory proof shows that 7 is not special by appealing to a symmetry of invariance under choice of digit.

However, Lange claims that 'this appeal to symmetry is not responsible for the proof's explanatory power. That is because the result's striking feature lies elsewhere'. In this case, according to Lange, the proof is explanatory because it:

> '... explains why the number of 7's is almost exactly 1/2 of 99,999. It reveals where the 1/2 comes from: There are 5 digits in each number being written down (if 1306 is being written as 01306), there are 10 digits in base 10 (0 through 9), and 5/10 = 1/2. This proof explains, then, by virtue of tracing the result's striking feature to a similar feature of the setup.' [Lange 2017: 267-8]

Now, it is not clear what kind of feature involving $\frac{1}{2}$ in a surprising way is. It is not a matter of the specific number $\frac{1}{2}$, since the equivalent result for 1 to 9,999 features 4/10 instead:

> Suppose you write down all of the whole numbers from 1 to 9,999. How many times would you write down the digit 7? The answer turns out to be 4,000 times. This is a striking result: 4,000 is almost exactly 4/10 of 9,999.

This sounds a little odd to me: it seems far less likely that we would be struck by the proportion 4/10 appearing in a result than the proportion 1/2. I suggest this is simply a contingent feature of human psychology, and yet the equivalent proof still seems to be explanatory:

> Consider the sequence of numbers from 0 to 9,999 when written down with four digits each (again, adding 0 will not change the number of times that the digit 7 appears). There are 10,000 numbers with four digits each – that is, 40,000 digits in total. Each of the ten digits must occur the same number of times overall, so each of the ten digits occurs 4,000 times. And 10,000 x 4 / 10 = 4,000.

Now, as we saw at the beginning of this chapter, Lange embraces the conclusion that 'if some extraterrestrials differ from us in which features of a given theorem they find salient, then it follows from my account that those extraterrestrials will also differ from us in which proofs they ought to regard as explanatory.' [Lange 2014: 525] So perhaps he would simply bite the bullet here and say that if we do not regard 'involving $^4/_{10}$' as a salient feature in the second result, we should not regard the second proof as explanatory.

But it seems far more plausible to me that both proofs are explanatory by virtue of exploiting a symmetry in the order of digits. Dreyfus and Eisenberg emphasise that 'there is only one important step, namely the conclusion from the symmetry between the digits to the equality of their number' [Drefus and Eisenberg 1986: 4].

I do see Lange's point that showing where the number $1/2$ 'comes from' helps to answer that why-question, if it arises. This feature may indeed add explanatory value to the proof. But my point here is that the proof's symmetry *also* plays an explanatory role: and yet symmetry does not feature as a salient feature in the result.

Therefore, I am sceptical about Lange's view that both the result and the proof must exhibit a similar salient feature for the proof to count as explanatory.

Let us now move on to look at the third feature Lange identifies: simplicity.

4.3 Simplicity

4.3.1 Euler's identity

Lange presents the following result involving partitions of n. Note that a partition of a positive integer, n, is a sequence of positive integers which sum to n. For example, $1 + 2 + 4$ is a partition of 7. An odd partition of n is a sequence of odd positive integers which sum to n. For example, $1+3+3$ is an odd partition of 7. A distinct partition of n is a sequence of distinct positive integers which sum to n. So $1+2+4$ is a distinct partition of 7, but $1+3+3$

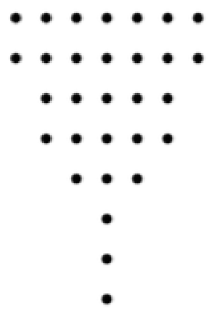
is not a distinct partition, because 3 is repeated.

Consider the following theorem.

> **Euler's identity**: The number of odd partitions of n, $P_O(n)$, is equal to the number of distinct partitions of n, $P_D(n)$.

Lange claims that a *simple* bijective proof of the identity is the most explanatory, a claim supported by mathematicians working in this area: 'a common feeling among combinatorial mathematicians is that a simple bijective proof of an identity conveys the deepest *understanding* of why it is true' [Andrews and Eriksson 2004: 9, emphasis in the original].

Lange's bijective 'proof', following Sylvester, runs as follows. We consider a special case, $n = 30$. One of the odd partitions of 30 is 7 + 7 + 5 + 5 + 3 + 1 + 1 + 1, which can be represented using a dot diagram.



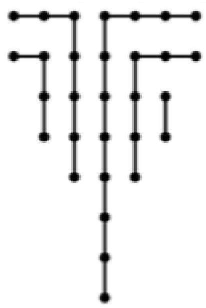
According to Lange:

> 'Each row has the number of dots in a part of the partition, with the rows weakly decreasing in length and their centers aligned. (Each row has a center dot since each part is odd.) Here is a simple way to transform this O-partition into a D-partition. The first part of the new partition [see figure overleaf] is given by the dots on a line running from the bottom up along the center column and turning right at the top – 11 dots. The next part is given by the dots on a line running from the bottom, up along a column one dot left of center, turning left at the top – 7 dots. The next part runs from the bottom upward along a column one dot right of center, turning

right at the last available row (the second row from the top) – 6 dots. This pattern leave us with a fish-hook diagram [see below].

The result is a D-partition (11 + 7 + 6 + 4 + 2), and the reverse procedure on that partition returns the original O-partition. With this bijection between O-partitions and D-partitions, there must be the same number of each. The key to the proof is that by "straightening the fish hooks,", we can see the same diagram as depicting both an O-partition and a D-partition.

Because the bijection is so simple, this proof traces the simple relation between $P_O(n)$ and $P_D(n)$ to a simple relation between the O-partitions and the D-partitions. Moreover, the simple feature of the setup that the proof exploits is similar to the result's strikingly simple feature: the result is that $P_O(n)$ and $P_D(n)$ are the same, and the simple bijection reveals that n's O-partitions are essentially the same objects as n's D-partitions, since one can easily be transformed into the other.' [Lange 2014: 518]



Now, Lange's presentation does not amount to a full proof, or even really a proof sketch, for two reasons: (i) he presents only a special case, $n = 30$ (and even then handles only one specific partition of 30); and (ii) the reverse procedure is not given: Lange presents only a way to get from an odd partition to a distinct partition, and not the other direction.

To be fair to Lange, mathematicians tend to present the same bijection in a similarly cursory way. For example, Pak presents the bijection as follows

[Pak 2006: 24]:

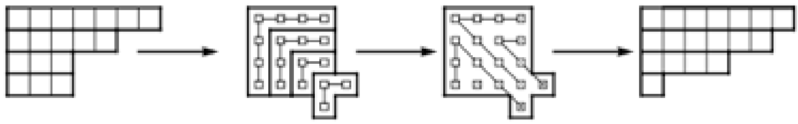


Fig. 12 Sylvester's bijection $\psi : (7, 5, 3, 3) \to (7, 6, 4, 1)$

However, it seems reasonable for Pak to leave the details up to the reader, while Lange is making a specific claim about the bijection – that it is simple – without actually presenting the full bijection.

One way to construct the reverse procedure is as follows:

Reverse procedure

Start with a distinct partition, i.e. a strictly decreasing and finite sequence of parts: $a_1, a_2, \ldots, a_k$.

If k is even:

- Re-order the sequence of parts as follows: $a_k, \ldots a_4, a_2, a_1, a_3, a_5, \ldots, a_{(k-1)}$
- Insert a_k as a row of dots on the left-hand side of the diagram. (We want it to be a fishhook without a vertical part, because if it had a vertical part then by symmetry there would have to be a subsequent part $a_{(k+1)}$ on the right-hand side.)
- Then insert $a_{(k-1)}$ as a fish-hook on the right-hand side of the diagram, making the bend such that the horizontal part of $a_{(k-1)}$ is one dot longer than the horizontal part of a_k and the horizontal parts lie on the same row. (This is possible because the sequence is strictly decreasing, so $a_{(k-1)}$ is at least one dot longer than a_k.)
- Then insert $a_{(k-2)}$ as a fish-hook on the left-hand side of the diagram, making the bend such that the vertical part of $a_{(k-2)}$ is one dot longer than the vertical part of $a_{(k-1)}$, and the horizontal part of $a_{(k-2)}$ lies one row higher than the horizontal part of $a_{(k-1)}$.

- Keep going, alternating these steps until you reach a_1, which you place as a fish-hook in the middle column, making the bend such that the horizontal part of a_1 is one dot longer than the horizontal part of a_2, and lies on the same row.
- We have then constructed a symmetrical diagram that represents an odd partition (it is odd because at each row we have ensured that the total row length is $2l + 1$ for some l).

If k is odd:

- Re-order the sequence of parts as follows: $a_{(k-1)}, \ldots a_4, a_2, a_1, a_3, a_5, \ldots, a_k$
- Insert a_k as a column of dots on the right-hand side of the diagram. (We want it to be a fishhook without a horizontal part, because if it had a horizontal part then by symmetry there would have to be a subsequent part $a_{(k+1)}$ on the left-hand side.)
- Then insert $a_{(k-1)}$ as a fish-hook on the left-hand side of the diagram, making the bend such that the vertical part of $a_{(k-1)}$ is one dot longer than the vertical part of a_k, and the horizontal part of $a_{(k-1)}$ lies one row higher than the horizontal part of a_k. (Again, this is possible because the sequence is strictly decreasing, so $a_{(k-1)}$ is at least one dot longer than a_k.)
- Then insert $a_{(k-2)}$ as a fish-hook on the right-hand side of the diagram, making the bend such that the horizontal part of $a_{(k-2)}$ is one dot longer than the horizontal part of $a_{(k-1)}$ and the horizontal parts lie on the same row.
- Keep going, alternating these steps until you reach a_1, which you place as a fish-hook in the middle column, making the bend such that the horizontal part of a_1 is one dot longer than the horizontal part of a_2, and lies on the same row.

- We have then constructed a symmetrical diagram that represents an odd partition (it is odd because at each row we have ensured that the total row length is $2l + 1$ for some l).

It took some work to construct this reverse procedure, and I think it is fair to say that Lange's presentation was not sufficient. Nevertheless, let us now consider the full bijection and examine what kind of feature simplicity is supposed to be.

4.3.2 Clarification

We need to distinguish some different notions here:

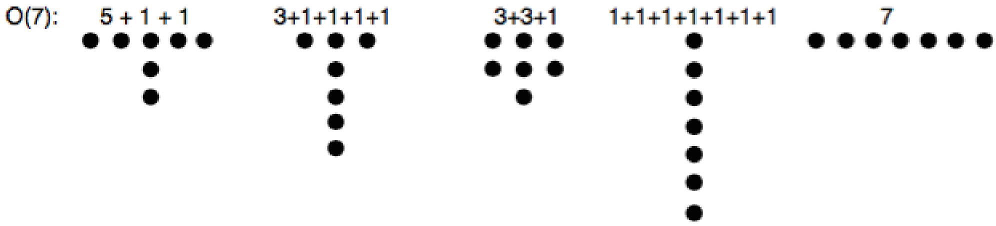
> 'A *correspondence* between two sets is a one-to-one function from one set into another ... A *bijection* is a correspondence between two sets together with its lexical description. Naturally, the same correspondence can be described in many different ways. Thus we can say that two bijections give the same correspondence. We also say that two bijections are *identical* if their descriptions are essentially the same or sufficiently close to each other.' [Pak 2006: 9].

It might help us to get clear on what simplicity is supposed to be if we think about what kind of thing potentially has the property. If we take Lange's claim that the bijection is simple at his word, then it's not the correspondence alone but the correspondence together with its description that is simple.

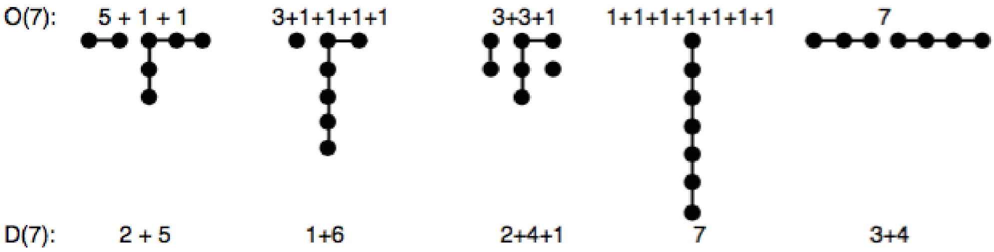
In fact, Lange's claims are slightly ambiguous since he says 'Because the bijection is so simple, this proof traces the simple relation between $P_O(n)$ and $P_D(n)$ to a simple relation between the O-partitions and the D-partitions' [Lange 2014: 518]; it's not clear here whether it's the description or the correspondence (the 'relation') that is simple.

I think simplicity must be a property of the bijection (the correspondence *and* description), rather than of the correspondence alone. This is because we could present the same correspondence in two different ways, with one presentation more explanatory than another. So simplicity of the correspondence alone would not be enough to account for explanatory power.

Take the following example, $n = 7$. There are five odd partitions of 7: $5+1+1$, $3+1+1+1+1$, $3+3+1$, $1+1+1+1+1+1+1$, and 7. There are five distinct partitions of 7: $1+6$, $2+5$, $2+4+1$, $3+4$, and 7. Following Lange's example, we present the five odd partitions as follows.



According to the fish-hook method Lange describes, we end up with the following diagram:



Alternatively, we could present the same correspondence as follows:

O(7)	<–>	D(7)
(5,1,1)	<–>	(2,5)
(3,1,1,1)	<–>	(1,6)
(3,3,1)	<–>	(2,4,1)
(1,1,1,1,1,1,1)	<–>	(7)
(7)	<–>	(4,3)

I think the first presentation is clearly more explanatory than the second, even though the same correspondence is presented. Hence simplicity of the correspondence alone can't account for the difference in explanatory value; as I've suggested previously, the description or presentation of the proof plays an important explanatory role.

This fits with Lange's own view: he writes that 'we can make a feature of some result more or less salient by the way we express it. What's salient is a feature of the result *under a certain representation*' [Lange 2017: 266, emphasis in the original].

So, let us take simplicity to be a property of the bijection – the correspondence together with its description – and let us now examine what kind of property this might be.

4.3.3 Directness

One natural thought is that simplicity describes something like directness. Both the correspondence and the description are potential bearers of directness, I think.

The correspondence in Lange's example is direct in the sense that we get straight from an element of one set to an element of the other, not going through multiple stages. This fits with Pak's concept of directness: 'We say that a bijection or involution is direct if it uses no intermediate steps in its constructions.' [Pak 2006: 10]. We can see directness as an objective feature here: it's independent of the reader whether or not a proof presents a direct correspondence.

But we must also have a notion of directness for the description, since we've just established that properties of the correspondence alone aren't sufficient to account for the proof's explanatory role.

The description is direct, I suggest, if the operations described are easily accessible to us. This is an epistemic aspect of directness.

The diagrammatic presentation gives us particularly direct access in Lange's example, while the second presentation of the correspondence I gave above does not give us access to the operations of the bijection at all: we are told which elements correspond, but we don't have access to the operation that connects them.

In Lange's presentation of the proof, the diagrams are essential to the reasoning. Sylvester, whose 1882 paper introduced the fish-hook diagrams, agrees that the diagrams can play both a facilitating and an essential role as an ‘instrument of transformation’:

> ‘The perception of the correspondence is in many cases greatly facilitated by means of a graphical method of representation, which also serves per se as an instrument of transformation’ [Sylvester 1882: 251].

However, I do not mean to claim that only diagrammatic proofs can provide direct access to the operations involved. For example, Andrews and Eriksson present a proof of Euler’s identity using a different procedure that does not involve a diagram:

> ‘From odd to distinct parts: If parts are distinct, there are no two copies of the same part. Hence, if the input to the bijection contains two copies of a part, then it must do something about it ... a natural thing to do is to merge the two parts into one part of double size. We can repeat this procedure until all parts are distinct – since the number of parts decreases at every operation, this must occur at the latest when only one part remains. For example,
>
> 3+3+3+1+1+1+1 —> (3+3)+3+(1+1)+(1+1) —> 6+3+2+2 —> 6+3+(2+2) —> 6+3+4
>
> Tracing our steps back to odd parts: The inverse of merging two equal parts is the splitting of an even part into two equal halves. Repeating this procedure must eventually lead to a collection of odd parts – since the size of some parts decreases at every operation, this must occur at the latest when all parts equal one. For example,
>
> 6+4+3 —> 6+3+(2+2) —> 6+3+2+2 —> (3+3)+3+(1+1)+(1+1) —> 3+3+3+1+1+1+1.
>
> It might seem that there is an arbitrariness in the order in which we choose to split (or merge) the parts. However, it is clear that splitting one part does not interfere with the splitting of other parts, so

> the order in which parts are split does not affect the result. Neither does the order of merging, since merging is the inverse of splitting.' [Andrews and Eriksson 2004: 8-9]

I suggest that this presentation of the correspondence is direct, as it makes the operations involved easily accessible to us. So, we can have directness without diagrams.

Now, to count as direct I don't think the presentation needs to give us *complete* access to the entity or property in question: in my earlier discussion of Steiner's account in Chapter 3, I suggested that someone could be justified in calling a proof explanatory if they have latched on to the right characterizing property, even if they don't latch on to it in full generality. Similarly, I claim that Lange's proof sketch of Euler's identity may correctly be judged explanatory even by readers who don't check the details of the reverse procedure as I did.

The idea here is that the proof sketch presents an objectively direct correspondence between even and distinct partitions, in the sense of not involving intermediate steps.[32] The proof sketch is also direct in the sense of describing the correspondence in a way that gives us epistemic access to the operations involved, even if it doesn't spell this out.[33]

So, understanding simplicity as directness in these two senses, we see that it is a property that holds of the bijection (the correspondence and its description), as required.

The second, epistemic aspect of directness – according to which a direct bijective proof gives us epistemic access to the details of the bijection – could be linked to the proof's constructive nature. Both explanatory proofs of Euler's identity – Sylvester's fishhooks version and Andrews and Eriksson's non-diagrammatic version – give us the construction needed to extend the corre-

[32]We might worry that my construction of the reverse procedure includes various steps: but these steps simply sketch out the details of the operation, rather than introducing new intermediate entities.

[33]Even though it took some time, I didn't need further information to construct the reverse procedure.

spondence to new cases, while the simple listing of pairs of partitions is non-explanatory precisely because it fails to give us this information: it provides only the correspondence, and not a way to construct it.

There is some support from mathematical practice to think that good bijective proofs of partition identities should involve construction. For example, Pak goes so far as to describe the theory of bijective proofs of partition identities as 'constructive partition theory' [Pak 2006: 7].

However, whether we link simplicity to directness or to the constructive nature of the proof, we run into a problem on Lange's account when we get back to his three conditions, as we haven't yet considered whether simplicity is present in the result.

Lange writes that:

> 'the simple feature of the setup that the proof exploits is similar to the result's strikingly simple feature: the result is that $P_O(n)$ and $P_D(n)$ are the same, and the simple bijection reveals that n's O-partitions are essentially the same objects as n's D-partitions, since one can easily be transformed into the other.' [Lange 2014: 518]

So in Lange's view, simplicity is $P_O(n)$ and $P_D(n)$ 'being the same' in the result, and n's O-partitions and D-partitions 'being the same' in the proof. In the proof, Lange describes 'being the same' in terms of ease of transformation, which seems to fit with my understanding of directness above.

In the result, 'being the same' is simply a statement of identity. It's not clear to me that this is an intrinsically simple property: before seeing the proof, we don't know whether $P_O(n)$ and $P_D(n)$ have the same value due to a multi-stage series of bijections, or even due to some mathematical coincidence.[34]

Again, we seem to have a problem case where the salient feature Lange proposes is present in the explanatory proof, but not in the result.

[34] For example, Lange suggests it is a mathematical coincidence that the thirteenth digits of π and e are both 9 [Lange 2010: 318].

I don't attempt to solve this problem with Lange's account, as indeed I'm not sure it can be solved in general.

In the next section, I return to examine the case of inductive proof. I present a way of extending Lange's account in order to make sense of some of his recent remarks about proof by induction.

4.4 Proof by induction (again)

As we saw in Chapter 2, Lange originally claimed that inductive proofs are never explanatory [Lange 2009]. In that chapter I argued that there is good reason to believe some inductive proofs are in fact explanatory, and indeed Lange has recently somewhat relaxed his view on induction to the effect that inductive proofs could fall 'somewhere *between* an explanation and a proof utterly lacking in explanatory power. (Explanatory power is a matter of degree)' [Lange 2014: 511, emphasis in the original]. This shift seems to have been prompted by his more recent account of mathematical explanation, analysed in the present chapter.

We might think that inductive proofs are a prime case of unifying proofs, since they show in one swoop that a property holds for every object in the domain. According to Lange, however, 'The inductive proof *nearly* treats every [case] alike. It gives special treatment only to the base case; all of the others receive the same treatment', and hence the inductive proof counts only as an almost-explanation [Lange 2014: 511, emphasis in the original].

I think this qualification is questionable, because as I argued earlier, cases of strong induction do not in fact require a base case to be handled separately from the rest of the domain. However, I don't wish to pursue this point here for two reasons: (1) Many cases of (weak) induction do involve a separate base case; and more importantly, (2) I have some doubts that the kinds of why-questions we have about inductively-proved theorems focus on unity. As I suggested in Section 4.1.2, even where a proof is unified it is not clear that its unity provides specifically explanatory power.

Instead, therefore, I want to explore another way of making sense of (some) explanatory inductive proofs by extending Lange's account to add a new salient feature. My suggestion is that we could think of induction as a proof strategy that tends to produce explanatory proofs in cases where the *recursive nature of the domain* is particularly salient.

Now, mathematical induction on $\mathbb{N}$ is just one kind of the more generally defined structural induction, which applies to any recursively defined set. In the case of the natural numbers, the recursive function that defines the set is the successor operation, and the set of natural numbers is the smallest set containing 0 and closed under succession:

1. 0 is an element of $\mathbb{N}$.

2. If n is an element of $\mathbb{N}$, then so is $n + 1$.

3. Nothing else is an element of $\mathbb{N}$.

We are familiar with the successor relation and with initial elements in the ordered set of natural numbers from learning to count as children. However, we are rarely explicitly aware at an early age that the set of natural numbers has a recursive structure, and inductive proofs are often introduced at a stage of education before students have explicitly entertained any thoughts about the structure of the set of natural numbers. So, I suggest, inductive proofs are often introduced at a time when the recursive nature of the domain is not salient for the average student.

It is interesting to examine the mathematics education literature on this point. For example, in a study on secondary school students' understanding of mathematical induction, a cohort of 213 students were asked questions about mathematical induction in connection with examples involving natural numbers and real numbers, respectively. The authors note that 'most students stated that MI [mathematical induction] holds only in the set of natural numbers, but they could not recognize the way this claim is justified with respect to the properties of this set. Only three students identified the lack of succession in the set of real numbers, a necessary property for applying the steps of MI' [Palla *et al.* 2012: 1041].

This could help to account for the difficulties many students face with understanding inductive proofs: the students have not yet appreciated the crucial feature – its recursive nature – that distinguishes the set of natural numbers from other number structures such as the real or rational numbers.

The authors go on to suggest that 'geometrical patterns can be an effective context for challenging secondary school students ... eventually, to attempt to produce a proof, mainly by appropriately corresponding the recursiveness of MI to the structure of the pattern' [Palla *et al.* 2012: 1043]. This would be interesting to investigate empirically in connection with my claim in Chapter 2 that pictorial inductive proofs are likely to be explanatory. My suggestion is that the diagrams draw our attention to the recursive nature of the domain by directly revealing a concrete way of constructing one case from the previous one. By drawing our attention to this feature of the domain, they make this feature salient for us. Once the recursive nature of the domain has been made salient, we are more likely to find the inductive proof explanatory, on my extension of Lange's account.

I think this suggestion could also account for the fact that proofs involving structural induction are more likely to be seen as explanatory. For example, in propositional logic the set of well-formed formulas is also recursively defined:

1. Every atom is a wff.
2. If ϕ is a wff, so is $\neg\phi$.
3. If ϕ and ψ are wffs, and $*$ is a binary connective (e.g. $\wedge, \vee, \rightarrow$), then $(\phi * \psi)$ is a wff.
4. Nothing else is a wff.

We can apply structural induction to prove facts about the set of well-formed formulas. In particular, to show that every element of the set has property P, we must:

- Show that each atom has property P.

- Show that if a wff ϕ has property P, then so does $\neg\phi$.
- Show that if wffs ϕ and ψ have property P, and $*$ is one of the binary connectives, then $(\phi * \psi)$ has property P.

As an example, consider the following theorem, covered in any introductory logic textbook.

> **Theorem:** Every well-formed formula of propositional logic has an equal number of right and left parentheses.
>
> **Proof:**
>
> Base case Holds vacuously, since the atoms do not contain any parentheses.
>
> Inductive step Suppose the wffs ϕ and ψ each have equal numbers of right and left parentheses (the inductive hypothesis). That is, suppose 'number of right parentheses in ϕ' $=$ 'number of left parentheses in ϕ'$= k$, for some k in $\mathbb{N}$, and that 'number of right parentheses in ψ' $=$ 'number of left parentheses in ψ'$= m$, for some m in $\mathbb{N}$.
>
> The wff $\neg\phi$ has the same number of parentheses as ϕ, so it also has an equal number of right and left parentheses. For each binary connective *, the wff $(\phi * \psi)$ has $k + m + 1$ right parentheses and $k + m + 1$ left parentheses, hence each wff has an equal number of right and left parentheses.

Here we have a trivial proof of the base case, and the inductive step takes us through each possible way of constructing wffs, showing that an equal number of right and left parentheses are added in each construction. I suggest that this counts as an explanatory proof.

Now, in cases like propositional logic our acquaintance with the set is mostly, or solely, via its definition, rather than acquaintance with its elements. Unlike the case of the natural numbers, we do not begin by counting early members of the set; rather, we are introduced to the set at the same time as we are introduced to the recursive operation that defines it. Plausibly, then, it is likely that our attention is drawn to the recursive nature of the domain.

My extension of Lange's account, which counts the recursive nature of the domain as a potential salient feature, therefore makes sense of the fact that the inductive proof just presented is likely to seem explanatory: the result displays the relevant salient feature, while the proof exploits it.

On this suggestion, Lange can make room for his modified view that inductive proofs can be explanatory.

4.5 Concluding remarks

In this chapter, I have considered Lange's recent account of intra-mathematical explanation in detail. I finished by suggesting an extension of the account to make room for explanatory inductive proofs.

However, I have argued that I don't think Lange's three conditions are met in general, as there are various cases where a proof gains explanatory power from exploiting a salient feature that is not salient in the result. I think both Lange and Steiner are too ambitious in trying to present an account that will cover all cases.

In the next chapter, I will instead carry out a careful case study of one specific example of an explanatory proof, aiming to analyse what makes that particular case explanatory. The case study will be Galois Theory.

5 Chapter 5: Galois theory

Why Galois theory? In earlier chapters, I argued that a good account of mathematical explanation should respect mathematical practice, at least by paying attention to what reflective mathematicians say about their practice. In this chapter, we will see that mathematicians frequently describe Galois theory as explanatory or illuminating, or enabling understanding, or providing the reason behind a result.

The Galois theory case provides a nice example, because a philosopher with a reasonable background in mathematics can access the explanation, so the example does not involve blindly accepting whatever mathematicians say. On the other hand, the proof is fairly complex and it does require a high level of mathematical skill: it is not an easy proof and the explanation it provides is not 'merely' pedagogical, as might be a worry for some of my earlier examples like the visual versions of simple inductive proofs.

Moreover, both of the philosophers whose accounts I have examined (Steiner and Lange) say something about the Galois theory case. I will argue that the Galois theory example provides further evidence that their proposed accounts are unable to cover all cases of mathematical explanation.

First, I must introduce the concepts needed to understand the case. For the mathematical background, I draw heavily on the canonical textbook in this field [Stewart 2015].

5.1 Mathematical background

5.1.1 Solving polynomials by radicals

Roughly speaking, a *polynomial of degree n* has the following form: $f(t) = a_0 + a_1t + ... + a_nt^n$. We obtain the corresponding *polynomial equation of degree n* by setting $f(t) = 0$. There are some subtleties about identifying the polynomial directly with the function f, and other considerations to do with how we define the generic polynomial, which I will discuss in Section 5.1.9.

A polynomial equation is *solvable by radicals* iff there is a formula for the roots of the equation using only the field operations of addition, subtraction, multiplication and division, n^{th} roots, and the coefficients of f. For example, the formula for the roots of the quadratic equation, $quad(t) = at^2 + bt + c$, is often learned by rote in school mathematics classes and is given by $t = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

It is harder to solve the general polynomial for degrees 3 and 4, the cubic and quartic. However, this was eventually achieved by, among others, Scipio del Ferro, Niccolo Fontana, Girolamo Cardano and Ludovico Ferrari in the sixteenth century. For example, Ferrari showed that the quartic equation could be solved by reducing it to a cubic. [See Stewart 2015, pp. 3-4]

At first, mathematicians hoped to extend the technique and provide a formula for the roots of the general polynomial equation of degree 5, the quintic. But various attempts over the next two hundred years, by great mathematicians such as Leonhard Euler and Joseph-Louis Lagrange, were unsuccessful, and mathematicians started to suspect that no such formula exists for the quintic.

In the late eighteenth and early nineteenth century, Paolo Ruffini showed that the quintic was not solvable by a certain restricted type of radical now known as Ruffini radicals (though he himself apparently did not realise that his proof was conditional on this restriction). Niels Henrik Abel further improved the situation in 1824 by showing that if a polynomial is soluble by radicals, then those radicals are all expressible in terms of rational functions of the roots, which means they are Ruffini radicals. So by reductio the quintic can't be soluble by radicals, since Ruffini had already showed it is not soluble by Ruffini radicals.

The 'Abel-Ruffini' proof hence shows us that the general quintic polynomial is not solvable by radicals.[35] However, as Pesic puts it, 'Though Abel's argument shows the impossibility of solving the quintic in general, it still seems opaque. The question remains: *why* this impossibility?' [Pesic 2003: 94].

[35] I will focus on the Galois theory proof here; for the full Abel-Ruffini proof, see e.g. [Pesic 2003].

The later proof based on the work of Evariste Galois answers this question, and moreover, helps us 'to appreciate the deep, general reason *why* quadratics, cubics, and quartics *can* be solved using radicals' [Stewart 2015: 108, emphasis in the original].

To understand the proof, we will first need to understand various crucial concepts, which I will introduce in the next few sections. The concepts are not straightforward to grasp, and in order to provide some context I will present the new concepts using the same worked example each time. This will also help to pinpoint some of the problems with the abstract way in which the Galois theory case is usually handled by philosophers.

I will make use of the example Stewart presents, since as his description shows, it's a good one to pick:

> 'The extension that we discuss is a favourite with writers on Galois theory, because of its archetypal quality. A simpler example would be too small to illustrate the theory adequately, and anything more complicated would be unwieldy. The example is the Galois group of the splitting field of $t^4 - 2$ over $\mathbb{Q}$.' [Stewart 2015: 155]

That is, we're going to consider the polynomial $g(t) = t^4 - 2$ over the field of rational numbers. I will describe what a 'splitting field' is in the next section, and introduce the concept of a 'Galois group' in Section 5.1.4.

5.1.2 Fields and field extensions

In the last section I spoke loosely of polynomials, without mentioning the importance of the field over which the polynomial is defined. Roughly speaking, a field is a non-empty collection of elements closed under addition and multiplication. More precisely, a field K can be defined as a set together with two operations, addition $(a + b)$ and multiplication (ab), satisfying the following conditions:

1. *Commutativity of addition*: $a + b = b + a$ for all $a, b \in K$.

2. *Associativity of addition*: $(a + b) + c = a + (b + c)$ for all $a, b, c \in K$.

3. *Additive identity*: There exists $0 \in K$ such that $0 + a = a$ for all $a \in K$.

4. *Additive inverse*: For any $a \in K$, there exists $-a \in K$ such that $a + (-a) = 0$.

5. *Commutativity of multiplication*: $ab = ba$ for all $a, b \in K$.

6. *Associativity of multiplication*: $(ab)c = a(bc)$ for all $a, b, c \in K$.

7. *Distributivity of multiplication*: $a(b + c) = ab + ac$ for all $a, b, c \in K$.

8. *Multiplicative identity*: There exists $1 \in K$ such that $1a = a$ for all $a \in K$.

9. *Multiplicative inverse*: For any $a \in K$ such that $a \neq 0$, there exists $a^{-1} \in K$ such that $aa^{-1} = 1$.

10. $1 \neq 0$.[36]

Common examples of fields are the number fields $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$: the rational, real and complex numbers.

When we describe a polynomial 'over' a field K, we simply mean that all of its coefficients are elements of K. Note that when a polynomial is over a field K, it can also be considered over any field in which K is contained: for example, if the coefficients of a polynomial are all rational numbers, then the polynomial can be considered over $\mathbb{Q}$, over $\mathbb{R}$ or over $\mathbb{C}$. The same polynomial will have different properties over different fields.

For example, take the polynomial from the worked example I will be discussing: $g(t) = t^4 - 2$. The coefficients of the polynomial are rational, hence real and also complex numbers. Over $\mathbb{R}$, the polynomial can be factorised into two polynomials of degree 2, $g(t) = (t^2 - \sqrt{2})(t^2 + \sqrt{2})$, because each of these 'smaller' polynomials also have real coefficients. However, the smaller polynomials do not have rational coefficients, so the original polynomial $g(t)$ is *irreducible* over $\mathbb{Q}$: it is not the product of two polynomials over $\mathbb{Q}$ with

[36] Condition 10 is not strictly necessary, but is usually included to avoid the problem of the field with one element.

smaller degree. So we can see that it is important to mention the field we are working in.[37]

There are many instances of more abstract fields, but we will be dealing exclusively with $\mathbb{C}$ and its subfields. A *subfield* of $\mathbb{C}$ is a subset of the elements of $\mathbb{C}$ that satisfies the field axioms just described, under the same field operations.

Now, a field L is a *field extension* of K if K is a subfield of L.[38] So, for example, $\mathbb{R}$ is a field extension of $\mathbb{Q}$, and $\mathbb{C}$ is a field extension of $\mathbb{R}$. We will be interested in one type of field extension in particular: the *splitting field* of a polynomial over K, which involves adjoining the roots of the polynomial to K.

First, note that the roots (or zeros) of a polynomial $f(t)$ over K are the values of t for which $f(t) = 0$. These are not relative to the choice of K, in the sense that the roots don't change depending on the field the polynomial is over: the roots are either elements of K or they are not. (Of course, whether the polynomial has solutions in the field it's considered over will depend on the field chosen.) In our example, the polynomial $g(t) = t^4 - 2$, the roots are $\sqrt[4]{2}$, $-\sqrt[4]{2}$, $i\sqrt[4]{2}$, and $-i\sqrt[4]{2}$, since $g(t) = 0$ for each of these values. None of these roots are in $\mathbb{Q}$, but all of them are in $\mathbb{C}$.

Now, we know by the Fundamental Theorem of Algebra that any polynomial $f(t)$ over $\mathbb{C}$ of degree n has a full set of n roots in $\mathbb{C}$ (allowing for repeated roots). Often the full set of n roots appears 'earlier' than that, in the sense that the roots are elements of a proper subfield of $\mathbb{C}$. For example, take one of the factors of $g(t)$, the polynomial $h(t) = t^2 - \sqrt{2}$. This polynomial, of degree 2, has two roots: $\sqrt[4]{2}$ and $-\sqrt[4]{2}$. These roots are both in $\mathbb{R}$, which is a proper

[37] The formal definition: A non-constant polynomial over a subring R of $\mathbb{C}$ is reducible if it is a product of two polynomials over R of smaller degree. Otherwise it is irreducible [Stewart 2015: 51, Definition 3.10]. A ring has the same conditions as a field except for condition 9, the existence of a multiplicative inverse. We don't need to worry about the difference between rings and fields for our purposes.

[38] Formally: A field extension is a monomorphism $\iota : K \to L$, where K and L are subfields of $\mathbb{C}$. We say that K is the small field and L is the large field. [Stewart 2015: 64, Definition 4.1]

subfield of $\mathbb{C}$.[39]

We capture this difference by saying that $g(t)$ splits over $\mathbb{C}$ (and not over $\mathbb{R}$), while $h(t)$ splits over $\mathbb{R}$ (and not over $\mathbb{Q}$): in general,

> 'If K is a subfield of $\mathbb{C}$ and f is a nonzero polynomial over K, then f *splits* over K if it can be expressed as a product of linear factors
>
> $$f(t) = k(t - \alpha_1)...(t - \alpha_n)$$
>
> where $k, \alpha_1, ..., \alpha_n \in K$. If this is the case, then the zeros of f in K are precisely $\alpha_1, ..., \alpha_n$. ' [Stewart 2015: 129, Definition 9.1]

This definition in turns helps us to identify the smallest subfield of $\mathbb{C}$ in which the polynomial has a full set of roots, namely the splitting field:

> 'A subfield Σ of $\mathbb{C}$ is a *splitting field* for the nonzero polynomial f over the subfield H of $\mathbb{C}$ if $H \subseteq \Sigma$ and
>
> (1) f splits over Σ;
>
> (2) If $H \subseteq \Sigma' \subseteq \Sigma$ and f splits over Σ' then $\Sigma' = \Sigma$.' [Stewart 2015: 130, Defn. 9.3]

Although Galois did not use this terminology, I will argue later that identifying the splitting field is one of the crucial insights of Galois theory. Indeed, the splitting field is also known as the Galois extension field, which hints to its importance.

Now, let us examine a concrete example of a splitting field, using our example $g(t) = t^4 - 2$. We have already seen that the roots of $g(t)$ are $\sqrt[4]{2}$, $-\sqrt[4]{2}$, $i\sqrt[4]{2}$, and $-i\sqrt[4]{2}$, and hence that $g(t)$ splits over $\mathbb{C}$:

$$g(t) = (t - \sqrt[4]{2})(t + \sqrt[4]{2})(t - i\sqrt[4]{2})(t + i\sqrt[4]{2})$$

However, although $g(t)$ does not split over $\mathbb{R}$, it does split over other subfields of $\mathbb{C}$: it will split over any subfield that contains the roots of $g(t)$ as

[39] A subfield K is *proper* if it is strictly smaller than the field L in which it is contained, that is, there is some element of L which is not an element of K.

elements. The splitting field we seek is the smallest such subfield. That is, the splitting field is the intersection of all subfields of $\mathbb{C}$ that contain the roots of $g(t)$. By definition, this is known as the subfield *generated by* the roots of $g(t)$.[40]

In our example, this is $\varSigma = \mathbb{Q}(\sqrt[4]{2}, -\sqrt[4]{2}, i\sqrt[4]{2}, -i\sqrt[4]{2})$, which simplifies to $\varSigma = \mathbb{Q}(\sqrt[4]{2}, i)$. This is just $\Sigma = \{a + b\sqrt[4]{2} + ci + di\sqrt[4]{2} \mid a, b, c, d \in \mathbb{Q}\}$.

The splitting field, or Galois extension field, is directly related to the Galois group, which I will introduce in Section 5.1.4. First, I will run through some properties of the splitting field, since some of these details will be important and I think they help us to get a grip on the concept of a splitting field.

5.1.3 Properties of the splitting field

We are interested in the following five properties of the splitting field, as defined above. First, the splitting field exists; second, it is unique. Additionally, it is finite, normal, and separable. I will present each property in turn.

Existence. Any polynomial $f(t)$ over (a subfield of) $\mathbb{C}$ has a splitting field, because either $f(t)$ has some proper subfield of $\mathbb{C}$ as a splitting field, or it has $\mathbb{C}$ as a splitting field (by the Fundamental Theorem of Algebra).

Uniqueness. For any polynomial $f(t)$ over (a subfield of) $\mathbb{C}$, its splitting field Σ is unique up to isomorphism.

The splitting field is finite. We need the following definition and supporting lemma from Stewart:

> Definition: Let K be a subfield of $\mathbb{C}$ and let $\alpha \in \mathbb{C}$. Then α is *algebraic* over K if there exists a non-zero polynomial p over K such that $p(\alpha) = 0$. Otherwise, α is *transcendental* over K. [Stewart 2015: 71, Defn. 5.1]

[40] Note on notation: If $L : K$ is a field extension and Y is a subset of L, then the subfield of $\mathbb{C}$ generated by $K \cup Y$ is written $K(Y)$ and is said to be obtained from K by adjoining Y [Stewart 2015: 65, Defn. 4.7]. Moreover, any subfield of $\mathbb{C}$ generated by a subset of $\mathbb{C}$ contains $\mathbb{Q}$, so we use the notation $\mathbb{Q}(Y)$ for the subfield of $\mathbb{C}$ generated by Y [ibid.]. Note also that by 'smallest' we mean smallest with respect to the subset relation, not with respect to cardinality.

> Lemma: An extension $L : K$ is finite if and only if $L = K(\alpha_1, ..., \alpha_r)$ where r is finite and each α_r is algebraic over K. [Stewart 2015: 84, Lemma 6.11]

We saw earlier that the splitting field of a polynomial $f(t)$, Σ, is generated by the finitely many roots of $f(t)$. Each of these roots is algebraic, since $f(\alpha_i) = 0$ for each root α_i, so Σ is generated by finitely many algebraic elements. Hence, Σ is a finite field extension.

The splitting field is normal. Recall from before that a polynomial is *irreducible* over a field K if it is not the product of two polynomials over K with smaller degree. Stewart provides the definition and result we need:

> Definition: An algebraic field extension $L : K$ is *normal* if every irreducible polynomial f over K that has at least one zero in L splits in L. [Stewart 2015: 132, Defn. 9.8]
>
> Theorem: A field extension $L : K$ is normal and finite if and only if L is a splitting field for some polynomial over K. [ibid., Theorem 9.9]

The splitting field is separable. We define separability of a polynomial as follows:

> Definition: An irreducible polynomial f over a subfield K of $\mathbb{C}$ is *separable* over K if it has [only] simple zeros [non-repeated roots] in $\mathbb{C}$, or equivalently, simple zeros in its splitting field. [Stewart 2015: 133, Defn. 9.10]

As we are only working over $\mathbb{C}$, the following result gives us all we need:

> Proposition: If K is a subfield of $\mathbb{C}$, then every irreducible polynomial over K is separable. [Stewart 2015: 135, Proposition 9.14]

Since Σ is a subfield of $\mathbb{C}$, the splitting field is separable.[41]

[41] The details are not important for our purposes, but formally we need to add the following definitions from [Stewart 2015]:

We have now seen some properties of the splitting field of a polynomial in $\mathbb{C}$, which will be important when we discuss the relation between field extensions and Galois groups in Section 5.1.7. In the next section, I will introduce the concept of the Galois group of a polynomial.

5.1.4 Automorphisms and the Galois group

First, we need to define the general concept of a group:

> 'A *group* is a set G equipped with an operation of 'multiplication' written $(g, h) \mapsto gh$. If $g, h \in G$ then $gh \in G$. The associative law $(gh)k = g(hk)$ holds for all $g, h, k \in G$. There is an identity $1 \in G$ such that $1g = g = g1$ for all $g \in G$. Finally, every $g \in G$ has an inverse $g^{-1} \in G$ such that $gg^{-1} = 1 = g^{-1}g$.' [Stewart 2015: 19]

A *subgroup* of G is simply a subset of the elements of G that satisfies these group axioms under the same group operation as G.

Now, an isomorphism is a structure-preserving map between two fields, and an automorphism is a structure-preserving map from a field to itself. Formally:

> Definition. Suppose that K and L are subfields of $\mathbb{C}$. An *isomorphism* between K and L is a map $\phi : K \to L$ that is one-to-one and onto, and satisfies the following condition
>
> $$\phi(x + y) = \phi(x) + \phi(y) \qquad \phi(xy) = \phi(x)\phi(y)$$
>
> for all $x, y \in K$. [Stewart 2015: 20, Defn. 1.3]
>
> ...

Definition 5.4. A polynomial $f(t) = a_0 + a_1t + ... + a_nt^n$ over a subfield K of $\mathbb{C}$ is *monic* if $a_n = 1$.

Definition 5.5. Let $L : K$ be a field extension, and suppose that $\alpha \in L$ is algebraic over K. Then the *minimal polynomial* of α over K is the unique monic polynomial m over K of smallest degree such that $m(\alpha) = 0$.

Definition 6.10. An extension $L : K$ is algebraic if every element of L is algebraic over K.

Definition 17.19. If $L : K$ is an extension then an algebraic element $\alpha \in L$ is *separable* over K if its minimal polynomial over K is separable over K. An algebraic extension $L : K$ is a *separable extension* if every $\alpha \in L$ is separable over K.

> An isomorphism of K with itself is called an *automorphism* of K. [ibid.: 21]

We then specify a special kind of automorphism:

> Definition. Let $L : K$ be a field extension, so that K is a subfield of the subfield L of $\mathbb{C}$. A K-automorphism of L is an automorphism α of L such that $\alpha(k) = k$ for all k in K. We say that α fixes $k \in K$. [Stewart 2015: 112, Defn. 8.1]

Finally, we can define the Galois group:

> Definition. The *Galois group* of a field extension $L : K$ is the group of all K-automorphisms of L under the operation of composition of maps. [Stewart 2015: 113, Defn. 8.3][42]

Since this is quite abstract, let's go back to our example, $g(t) = t^4 - 2$.

In this example, as we saw earlier, we have $g(t) = (t - \sqrt[4]{2})(t + \sqrt[4]{2})(t - i\sqrt[4]{2})(t + i\sqrt[4]{2})$, and the splitting field is $\Sigma_g = \mathbb{Q}(\sqrt[4]{2}, i)$, or $\Sigma_g = \{a + b\sqrt[4]{2} + ci + di\sqrt[4]{2} \mid a, b, c, d \in \mathbb{Q}\}$.

Here $\Sigma_g : \mathbb{Q}$ is the field extension, and we are looking for the group G_g of all $\mathbb{Q}$-automorphisms of Σ_g; that is, the group of all automorphisms of Σ_g which leave $\mathbb{Q}$ fixed.

Now any element, s, of the splitting field Σ_g is a combination of rational numbers; i; and $\xi = \sqrt[4]{2}$ (for ease of notation). An automorphism of the splitting field that leaves the rationals fixed is one that has no effect on the rational part of s; so all we need to know is the effect that the mapping has on the 'i' and 'ξ' parts of s.

In order for the mapping to be an automorphism, the mapping ϕ needs to be one-to-one and onto, and must satisfy the conditions $\phi(x+y) = \phi(x)+\phi(y)$ and $\phi(xy) = \phi(x)\phi(y)$.

It is fairly straightforward to spot two such mappings, $\sigma : \Sigma_g \mapsto \Sigma_g$ and $\tau : \Sigma_g \mapsto \Sigma_g$, defined as follows:

---

[42] This is well-defined because we know that the set of all K-automorphisms of L always forms a group under composition of maps (see [Stewart 2015: 113, Theorem 8.2].

$\sigma(i) = i \quad \sigma(\xi) = i\xi$

$\tau(i) = -i \quad \tau(\xi) = \xi$

Why are these mappings of the right kind? Well, take σ first. For any element s of the splitting field Σ_g, we have $s = a + b\xi + ci + di\xi$ for some $a, b, c, d \in \mathbb{Q}$,

Then $\sigma(s) = a + ib\xi + ci - d\xi$. This is an element of the splitting field, so the image of the splitting field Σ_g under $\sigma$is certainly contained in Σ_g. To check that σ is onto (its image is the whole of Σ_g), we must check that for any s in Σ_g, $\exists r \in \Sigma_g$ such that $\sigma(r) = s$. Take $s = a + b\xi + ci + di\xi$; we just need to set $r = a + d\xi + ci - bi\xi$. Then $\sigma(r) = a + id\xi + ci + b\xi = s$, as required.

For the mapping σ to be one-to-one, it must be the case that for any two elements s, t of Σ, if $\sigma(s) = \sigma(t)$ then $s = t$. Let $t = e + f\xi + gi + hi\xi$ for some $e, f, g, h \in \mathbb{Q}$. Then $\sigma(t) = e + if\xi + gi - h\xi$. If $\sigma(s) = \sigma(t)$ then we have $a = e$, $b = f$, $c = g$, and $d = h$ and so $s = t$, as required.

The addition and product conditions can be easily checked, as can the conditions for τ.

Now, we note that σ and τ are not the only such mappings. Products of σ and τ yield further distinct $\mathbb{Q}$-automorphisms, as listed in the table provided by Stewart [2015: 156]:

Automorphism	*Effect on* $\xi = \sqrt[4]{2}$	*Effect on* i
1	ξ	i
σ	$i\xi$	i
σ^2	$-\xi$	i
σ^3	$-i\xi$	i
τ	ξ	$-i$
$\sigma\tau$	$i\xi$	$-i$
$\sigma^2\tau$	$-\xi$	$-i$
$\sigma^3\tau$	$-i\xi$	$-i$

Stewart notes that:

> ‘Other products do not give new automorphisms, since $\sigma^4 = 1$, $\tau^2 = 1$, $\tau\sigma = \sigma^3\tau$, $\tau\sigma^2 = \sigma^2\tau$, $\tau\sigma^3 = \sigma\tau$. (The last two relations follows from the first three.) Any $\mathbb{Q}$-automorphism of $[\Sigma_g]$ sends i to

some zero of t^2+1, so $i \mapsto \pm i$; similarly ξ is mapped to $\xi, i\xi, -\xi$, or $-i\xi$. All possible combinations of these (eight in number) appear in the above list, so are precisely the $\mathbb{Q}$-automorphisms of $[\Sigma_g]$.' [ibid.][43]

So, the Galois group G_g contains all of (and exactly) the automorphisms listed in the table above. The relations *between* the automorphisms give us the abstract structure of the group; as Stewart puts it,

> 'The generator-relation presentation
>
> $G = \langle \sigma, \tau : \sigma^4 = \tau^2 = 1, \tau\sigma = \sigma^3\tau \rangle$
>
> shows that G is the dihedral group of order 8, which we write as $\mathbb{D}_4$.' [ibid.]

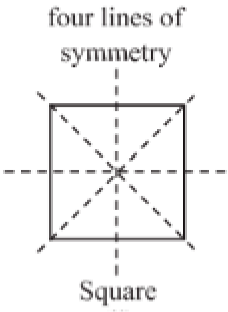
A *dihedral group* is a group of symmetries of a regular polygon. To understand the nature of this group, we will need to look at groups of symmetries in more detail, which I will do in the next section.

5.1.5 Groups of symmetries

Recall from Chapter 4 that an automorphism is simply the mathematical generalisation of the geometrical notion of a symmetry: an automorphism is a structure-preserving bijection of a set onto itself, and a geometrical symmetry is a specific instance of an automorphism where the relevant set is a collection of points in the plane. In the case of mirror reflection, for example, the set of points in the plane is bijected onto itself, with the structure of geometrical figures preserved (e.g. distance ratios, angles, etc.). For a specific geometric figure, a mirror symmetry is a mirror reflection that not only maps the whole plane onto itself, but that maps the geometrical figure onto itself (preserving not just distance ratios and angles, but position in the plane).

[43] To unpack this a little: $f(t) = t^2+1$ is the minimal polynomial of i over $\mathbb{Q}(\sqrt[4]{2})$, because $i^2+1=0$ but $i \notin \mathbb{R} \supseteq \mathbb{Q}(\sqrt[4]{2})$. Note that $f(\sigma(i)) = \sigma(f(i)) = f(i)$, since the coefficients of f are all fixed by σ. So the automorphism σ must send i to another zero of f, that is, either to i or $-i$. Similarly $g(t)$ itself is the minimal polynomial for ξ.

For example, the square has four lines of mirror symmetry, that is, lines in which a mirror reflection maps the square onto itself. These are the dotted lines in the diagram:



Note that applying the mirror reflection twice from any starting position brings all points of the square back to their starting position. Call the automorphism of mirror reflection τ_r. We can see that it has one of the same properties as the automorphism τ described above: $\tau_r^2 = 1$.

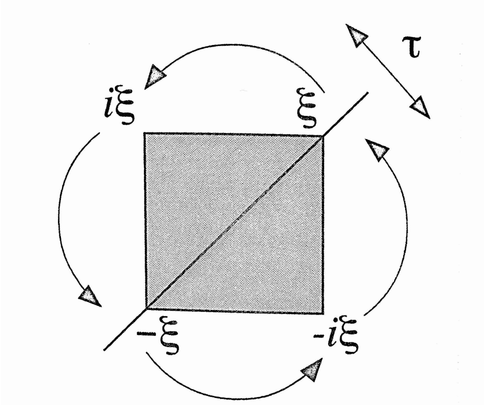
The square is also mapped onto itself by a rotation through 90 degrees. Note that applying this rotation mapping three times brings all points of the square back to their starting position. Call the automorphism of rotation through 90 degrees σ_r, and again we see that it has one of the same properties as the automorphism σ described above: $\sigma_r^3 = 1$.

Another property is also matched: namely $\tau_r\sigma_r = \sigma_r^3\tau_r$, since reflecting and then rotating once brings the square to the same position as rotating three times and then reflecting. And in fact, these were the three crucial relations from which the other relations between τ and σ followed in the previous section. Here too, we have $\tau_r\sigma_r^2 = \sigma_r^2\tau_r$ and $\tau_r\sigma_r^3 = \sigma_r\tau_r$. So the eight symmetries, or automorphisms, of the square are as follows: $\{1, \sigma_r, \sigma_r^2, \sigma_r^3, \tau_r, \sigma_r\tau_r, \sigma_r^2\tau_r, \sigma_r^3\tau_r\}$. The dihedral group of order 8 is this collection of automorphisms or symmetries, together with the operation of composition of maps.

This is all it means to say that the Galois group of $\Sigma_g : \mathbb{Q}$ in the previous section 'is the dihedral group of order 8'; it has exactly the same structure as the symmetry group of the regular square.[44] Indeed, we can explicitly label

[44]As Stewart puts it, the Galois group 'has a geometric interpretation as the symmetry group of a square' [Stewart 2015: 156].

the vertices of the square to make the correspondence obvious [Stewart 2015: 156]:



In the case of the square, the relevant symmetries map the whole plane to itself, with special attention to where the points of the square end up: they must get mapped to another point of the square. In the case of the polynomial, the relevant symmetries map the whole of the splitting field to itself, with special attention to where the rational numbers (the coefficients of the polynomial) end up: they must be fixed, i.e. mapped to themselves.

Now, remember that the splitting field contains only the rationals and those numbers that need to be adjoined to the rationals in order to get all roots of the polynomials. The process of fixing the rationals – the coefficients of the polynomials – allows us to focus on where the roots of the polynomial end up: each is mapped to another root. That is, we are focusing on symmetries of the roots of the polynomial. This is one of the crucial insights of Galois theory: as Stewart puts it, 'In modern terms, Galois's main idea is to look at the symmetries of the polynomial $f(t)$.' [Stewart 2015: 108]

Now, since the group of symmetries of the polynomial $g(t)$, G_g, has the same structure as the dihedral group $\mathbb{D}_4$, the two groups also have the same structural properties. This is important because we already know about some of the structural properties of $\mathbb{D}_4$. That is, we can use $\mathbb{D}_4$ as a 'counterpart', in what Steiner describes as 'the contemporary style ... to study domain X by assigning a counterpart in domain Y to each object in X' [Steiner 1987: 149-50].

Once we have established the structural properties of G_g, it 'turns out' that we can translate these properties to properties of the splitting field Σ_g, and onwards to properties of the polynomial $g(t)$. In particular, we are interested in the corrrespondence between subgroups of G_g and subfields of Σ_g. It 'turns out' there is a one-to-one correspondence between them.

The Fundamental Theorem of Galois Theory shows that this 'turning out' is no series of coincidences or 'fortunate accidents' [Stewart 2015: 124].

5.1.6 The Fundamental Theorem of Galois Theory

The Fundamental Theorem states that there is a one-to-one correspondence between (normal) subgroups of a Galois group and (normal) subfields of the relevant Galois field extension. To describe the correspondence, we need one further definition. Suppose (as in the Galois case) that group G is a group of field automorphisms on a field L, and H is a subgroup of G. The *fixed field* of H in L is the set of elements of L that are fixed by all operations in the subgroup H. That is, $L^H := \{x \in L | \sigma(x) = x \; for \; all \; \sigma \in H\}$. [Stewart 2015: 114-5]

The correspondence we are after is given by two maps, as described by Stewart [2015: 151]:

> 'Let $L : K$ be a field extension in $\mathbb{C}$ with Galois group G, which consists of all K-automorphisms of L. Let $\mathcal{F}$ be the set of intermediate fields, that is, subfields M such that $K \subseteq M \subseteq L$, and let $\mathcal{G}$ be the set of all subgroups H of G. We have defined two maps
>
> $* : \mathcal{F} \to \mathcal{G}$
>
> $\dagger : \mathcal{G} \to \mathcal{F}$
>
> as follows: if $M \in \mathcal{F}$, then M^* is the group of all M-automorphisms of L. If $H \in \mathcal{G}$, then $H^\dagger$ is the fixed field of H.'

That is, if M is one of the intermediate fields, it gets mapped by $*$ to the group of all M-automorphisms of L: the group of all automorphisms of L that leave M fixed. If H is one of the subgroups of G, then it gets mapped by $\dagger$ to the fixed field of that subgroup. The one-way maps $*$ and $\dagger$ are clearly defined carefully so that together they create a two-way map.

Indeed, the Fundamental Theorem of Galois Theory states that if $L : K$ as defined above is a *finite normal* field extension in $\mathbb{C}$, then 'The maps $*$ and $\dagger$ are mutual inverses, and set up an order-reversing one-to-one correspondence between $\mathcal{F}$ and $\mathcal{G}$'. We call this the *Galois correspondence*.

Now, we know from Section 5.1.3 that a field extension $L : K$ is normal and finite if and only if L is a splitting field for some polynomial over K. We are interested precisely in field extensions that are splitting fields of polynomials, so the field extensions $L : K$ we are interested in are indeed normal and finite, as required in the set-up of the Fundamental Theorem.[45]

Furthermore, the Galois correspondence preserves normality: for any intermediate field $M \in \mathcal{F}$, M is a normal field extension of K if and only if M^* is a normal subgroup of G, where $H \subset G$ is a *normal* subgroup of G if H is a subgroup of G and $gH = Hg$ for all $g \in G$.

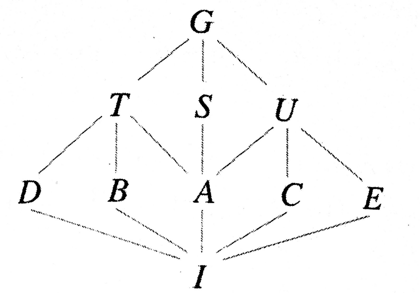
I will not go into the full details of the proof of the Fundamental Theorem here; for that see Stewart [2015: 151-2]. The important thing for our purposes is to get a sense of how the Galois correspondence allows us to translate useful properties between the Galois group and the splitting field.

Let us examine the Galois correspondence in the case of our particular example, $g(t) = t^4 - 2$. First, we need to find the subgroups of G_g. These are as follows [Stewart 2015: 157]:

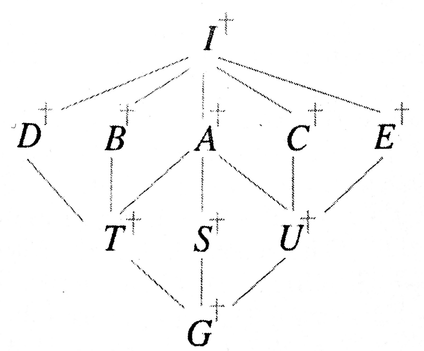
[45] As Stewart puts it, 'Galois's results can be interpreted as giving conditions under which $*$ and $\dagger$ are mutual inverses, setting up a bijection between $\mathcal{F}$ and $\mathcal{G}$. The extra conditions needed are called separability (which is automatic over $\mathbb{C}$) and normality.' [Stewart 2015: 115]

Order of subgroup	*Elements of subgroup*	*Name of subgroup*
Order 8 :	$\{1, \sigma, \sigma^2, \sigma^3, \tau, \sigma\tau, \sigma^2\tau, \sigma^3\tau\}$	G
Order 4 :	$\{1, \sigma, \sigma^2, \sigma^3\}$	S
	$\{1, \sigma^2, \tau, \sigma^2\tau\}$	T
	$\{1, \sigma^2, \sigma\tau, \sigma^3\tau\}$	U
Order 2 :	$\{1, \sigma^2\}$	A
	$\{1, \tau\}$	B
	$\{1, \sigma\tau\}$	C
	$\{1, \sigma^2\tau\}$	D
	$\{1, \sigma^3\tau\}$	E
Order 1 :	$\{1\}$	I

There are various 'inclusion relations' among these subgroups. For example, I is contained in all of the other subgroups, A is contained in both T and S, and E is contained in U. We can represent these inclusion relations graphically using a *lattice diagram* [Stewart 2015: 157]:



We then use the map † to give us the corresponding subfields, $G^\dagger$, $S^\dagger$, $T^\dagger$, $U^\dagger$, $A^\dagger$, $B^\dagger$, $C^\dagger$, $D^\dagger$, $E^\dagger$ and $I^\dagger$. Now, the subfields are related to each other in the same lattice pattern, though the inclusion relations are reversed:



Take for example the subfield $I^\dagger$. This is the fixed field of I in the splitting field Σ_g, that is, the elements of Σ_g that are fixed by all automorphisms in the subgroup I. The only operation in I is the identity automorphism, and all of Σ_g is fixed by the identity automorphism, so $I^\dagger = \Sigma_g$.

The fixed field $G^\dagger$ contains the elements of Σ_g fixed by all automorphisms in G, and G contains all automorphisms of Σ_g. The only element of Σ_g fixed by all automorphisms is the identity element, 1.

The fixed field $A^\dagger$ contains the elements of Σ_g fixed by all automorphisms in A, namely 1 and σ^2. The field containing all and only elements which are fixed by 1 and σ^2 is an intermediate field between $I^\dagger$ and $G^\dagger$. Indeed, it is an intermediate field between $I^\dagger$ and $S^\dagger$, since $S^\dagger$ contains all and only those elements of Σ_g fixed by 1, σ^2, and the additional automorphisms σ and σ^3.

So, we can appreciate that the Galois correspondence gives the subgroups of the Galois group and the subfields of the splitting field the same relational structure. But how does this help us to find out whether a particular polynomial is solvable by radicals? From the lattice diagram, we can draw a chain of inclusion from I to G and from $G^\dagger$ to $I^\dagger$. This is an important feature in the concept of *solvability of groups*, which will help us to tackle the question of solvability of polynomials and which I will discuss in the next section.

5.1.7 Two concepts of solvability

We have already introduced the concept of solvability for polynomials, namely solvability by radicals: we said that a polynomial f is *solvable by radicals* (call this *solvable*$_R$) iff there is a formula (some *radical expression*) for the roots of the equation using only the field operations of addition, subtraction, multiplication and division, n^{th} roots, and the coefficients of f. We now make this precise by introducing a few further definitions, using the concepts we have already encountered.

> Definition. A [field] extension $L : K$ in $\mathbb{C}$ is *radical* if $L = K(\alpha_1, ..., \alpha_m)$ where for each $j = 1, ..., m$ there exists a [positive] integer n_j such that

$$\alpha_j^{n_j} \in K(\alpha_1, ..., \alpha_{j-1}) \quad (j \geq 2)$$

The elements α_j form a *radical sequence* for $L : K$. The radical degree of the radical α_j is n_j. [Stewart 2015: 121, Defn. 8.12]

Definition. Let f be a polynomial over a subfield of $\mathbb{C}$, and let Σ be the splitting field for f over K. We say that f is *soluble by radicals* if there exists a field M containing Σ such that $M : K$ is a radical extension. [Stewart 2015: 172, Defn. 15.1]

Two points are important here. First, a radical extension $L : K$ has a chain of intermediate subfields leading from K to L. Namely, $K = K_0 \subset K_1 \subset \cdots \subset K_m = L$, where for $j = 1, \cdots, m$, $K_j = K_{j-1}(\alpha_j)$ for $\alpha_j^{n_j} \in K_{j-1}$. (This is what it means to say the elements α_j form a radical sequence for $L : K$.) That is, $L : K$ is built up from K by adding n^{th} roots, together with the usual field operations of addition, subtraction, multiplication and division. This makes it easy to see that any radical element $\sqrt[n]{\beta}$ must be contained in some radical extension.

Second, note that the splitting field of the polynomial need not itself be a radical extension. As Stewart puts it, 'We want everything in the splitting field Σ to be expressible by radicals, but it is pointless to expect everything expressible by the same radicals to be inside the splitting field.' [Stewart 2015: 172]

Recall that the splitting field Σ contains all of the roots of the polynomial. Stewart's point here is that the splitting field does not necessarily contain everything expressible by the radical expressions needed to express the roots of the polynomial. But as long as there exists a field M containing Σ such that $M : K$ is a radical extension, then M contains both the roots of the polynomial and the radical expressions that express the roots. This is what we need for the polynomial to be solvable by radicals.

We can now move on to look at the concept of solvability for groups. Again, we need a few further definitions. First, for a subgroup H of a group G and an element g in G, we define the *left coset* of H, gH, to be the set $\{gh : h \in H\}$, and the right coset of H, Hg, to be the set $\{hg : h \in H\}$.

Recall from the previous subsection that a subgroup $H \subset G$ is a *normal* subgroup of G if $gH = Hg$ for all $g \in G$; that is, the left and right cosets of H are the same.

A group G is *abelian* if all of its elements commute: that is, for any g and h in G, $gh = hg$. Note that all subgroups of an abelian group are normal.

Finally, for a group G and a normal subgroup N of G, the quotient group of N in G, written G/N and read "G modulo N", is the set of cosets of N in G.[46]

Using these definitions, we can define a concept of solvability for groups:

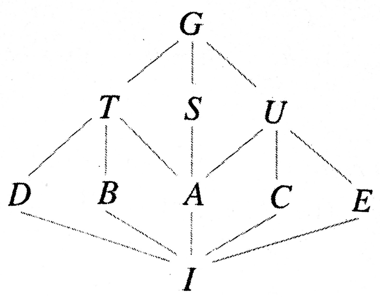
> A group G is *solvable*$_G$ if it has a finite series of subgroups: $1 = G_0 \subset G_1 \subset G_2 \subset \ldots \subset G_n = G$, such that G_i is normal in G_{i+1} and the quotient group G_{i+1}/G_i is abelian for $i = 0, \ldots, n{-}1$ [Stewart 2015: 61, Defn. 14.1].

Already, we can see that both concepts of solvability involve chains, of subfields and subgroups respectively.

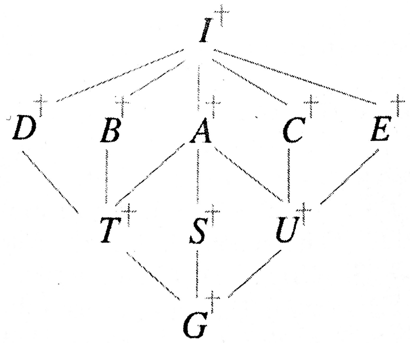
In our example $g(t) = t^4 - 2$, we can trace out these chains using the lattice diagrams. Take the lattice diagram for subgroups first. We have for example $I \subset A \subset S \subset G$. Recall that the elements of these groups are as follows: I contains only 1, A is made up of the elements $\{1, \sigma^2\}$, S of $\{1, \sigma, \sigma^2, \sigma^3\}$, and G of $\{1, \sigma, \sigma^2, \sigma^3, \tau, \sigma\tau, \sigma^2\tau, \sigma^3\tau\}$, in each case together with the group operation of composition of maps.

As an example, take the first step of this chain. The left and right cosets of I in A are $1I$ and $\sigma^2 I$, and $I1$ and $I\sigma^2$ respectively. Since $1I = I1$ and $\sigma^2 I = I\sigma^2$, the left and right cosets are equal. That is, I is normal in A.

[46]See e.g. http://mathworld.wolfram.com/QuotientGroup.html



Now consider the lattice diagram for subfields.



Here the relevant chain of subfields is $G^\dagger \subset S^\dagger \subset A^\dagger \subset I^\dagger$. Because the relational structure of subgroups and subfields is the same in both domains, a suitable chain exists in one domain iff it exists in the other. This helps us to understand why solvability coincides in the two domains, i.e. why a polynomial f is $solvable_R$ if and only if its Galois group G is $solvable_G$.

5.1.8 The generic polynomial

So far, we have focused on a specific example of a polynomial, $g(t) = t^4 - 2$, which is solvable by radicals. But we are interested in polynomials that are *not* solvable by radicals, in particular the result that the general polynomial of degree ≥ 5 is not solvable by radicals. Before proving this result, we will first need to go back to the subtlety I mentioned at the start of the chapter.

I said that a *general polynomial of degree n* has the following form: $f(t) = a_0 + a_1 t + ... + a_n t^n$, for some $n \in \mathbb{N}$ and $a_o, ..., a_n \in \mathbb{Q}$. This 'general' polynomial is in fact just a schema for specific polynomials like the example we have been working with: we insert some specific degree n (in our example, 4), and some specific coefficients $a_0, ..., a_n$ (in our example, $-2, 0, 0, 0, 1$), to get the specific polynomial (in our example, $g(t) = t^4 - 2$).

Now, working in $\mathbb{C}$, we follow Stewart [2015: 116-17] in defining the *generic* polynomial of degree n as follows:

$F_n(t) = (t-t_1)(t-t_2)...(t-t_n) = t^n - s_1t^{n-1} + s_2t^{n-2} + ... + (-1)^n s_n$, where the s_j are the *elementary symmetric polynomials*:

$$\begin{aligned} s_1 &= t_1 + ... + t_n \\ s_2 &= t_1t_2 + t_1t_3 + ... + t_{n-1}t_n \\ &... \\ s_n &= t_1...t_n \end{aligned}$$

The elementary symmetric polynomials are expressed here as rational functions of the roots, $t_1, ...t_n$, and they are symmetric in the sense that each of them is invariant under any change of the order of the roots.[47] The outcome is that the generic polynomial of degree n is invariant under any change of the order of its roots.

Now, there is some danger of ambiguity here as $F_n(t)$ is sometimes referred to as the 'general' polynomial of degree n. This is because, as Stewart puts it, 'this polynomial has a universal property. If we can solve $F_n(t) = 0$ by radicals, then we can solve any *specific* complex polynomial equation of degree n by radicals.' [Stewart 2015: 117]. I will be careful to distinguish the use of the terms 'general' and 'generic', since it is clear the latter 'has a paradoxically special meaning in this context' [ibid.: 116].

I have drawn up the table below to help illustrate the difference between the terms.

[47] Note, importantly, that any symmetric polynomial expressed in the roots t_i can also be written as a rational expression in s_i; for any polynomial g, symmetric expressions in its roots are also rationally expressible in terms of its coefficients. See Stewart [2015: 208-211].

	specific	general	generic
n=1	e.g. $g_1(t) = 3t - 2$	$f_1(t) = a_1t + a_0$, where a_i are rational numbers.	$F_1(t) = t - s_1$, where s_1 is the elementary symmetric polynomial of degree 1.
n=2	e.g. $g_2(t) = 7t^2 + 3t - 3$	$f_2(t) = a_2t^2 + a_1t + a_0$, where a_i are rational numbers.	$F_2(t) = t^2 - s_1t + s_2$, where s_j are the elementary symmetric polynomials.
n=3	e.g. $g_3(t) = t^3 - 2t^2 + 3t$	$f_3(t) = a_3t^3 + a_2t^2 + a_1t + a_0$, where a_i are rational numbers.	$F_3(t) = t^3 - s_1t^2 + s_2t - s_3$, where s_j are the elementary symmetric polynomials.
n=4	e.g. $g_4(t) = 9t^4 + t^2 + 1$	$f_4(t) = a_4t^4 + \ldots + a_1t + a_0$, where a_i are rational numbers.	$F_4(t) = t^4 - s_1t^3 + s_2t^2 - s_3t + s_4$, where s_j are the elementary symmetric polynomials.
n=5	e.g. $g_5(t) = 2t^5 + 3t$	$f_5(t) = a_5t^5 + a_4t^4 + \ldots + a_1t + a_0$, where a_i are rational numbers.	$F_5(t) = t^5 - s_1t^4 + s_2t^3 - s_3t^2 + s_4t - s_5$, where s_j are the elementary symmetric polynomials.
n>5	e.g. $g_n(t) = nt^n + (n-1)t^{n-1} + \ldots + 2t^2 + t$	$f_n(t) = a_nt^n + a_{n-1}t^{n-1} + \ldots + a_1t + a_0$, where a_i are rational numbers.	$F_n(t) = t^n - s_1t^{n-1} + \ldots + (-1)^ns_n$, where s_j are the elementary symmetric polynomials.

For our purposes, the crucial feature of the generic polynomial of degree n is that it is invariant under any change of the order of its roots, as already mentioned. This means that its Galois group (that is, the Galois group of $\Sigma_{F_n} : \mathbb{Q}$) is isomorphic to the full symmetric group of n elements, as I will explain in the next section.

5.1.9 The symmetric group and its subgroups

Earlier, I talked about groups of symmetries, giving specific examples like the group of rotations and reflections of the square. The rotations and reflections are automorphisms of the set of points of the square: that is, permutations with structure-preserving properties. For example, they preserve the distance between points.

We generalize from this to the full *symmetric group,* S_n, which consists of *all* permutations of the set $\{1, 2, ..., n\}$ under the operation of composition (not only the structure-preserving ones). [Stewart 2015: 19]. Elements of S_n include, for example, the permutation $(12345...n)$, which moves each element along by one place and takes n to position 1.

For each n, S_n has various subgroups, and one of these subgroups is always the *alternating group*, A_n: the group of *even* permutations of $\{1, 2, ..., n\}$. An even permutation of $\{1, 2, ..., n\}$ is simply one that contains precisely an even number of swaps or transpositions between pairs of elements. For example, (12)(34) is an even permutation of $\{1, 2, 3, 4\}$, while (1)(23)(4) is an odd permutation. The collection of even permutations forms a subgroup because the composition of two even permutations is even (unlike the collection of odd permutations, which is not closed since the composition of two odd permutations is also even).

Now for $n \geq 5$, the alternating group A_n is quite different from the alternating groups for earlier n. For $n = 1$ and $n = 2$, only the identity permutation has an even number of swaps (that is, zero), and so A_1 and A_2 are isomorphic to the trivial group. A_3 is abelian (so all of its subgroups are normal), while A_4 is not abelian but has a normal non-trivial subgroup. This means we can get the chain of subgroups needed to show that A_4 is *solvable*$_G$, as are A_1, A_2 and A_3.[48]

A_5, on the other hand, is the smallest non-abelian group that is *simple*: it has no normal subgroups except the trivial group and A_5 itself. This means that A_5 is not *solvable*$_G$: it cannot be broken down into a chain of normal subgroups for which each quotient group is abelian, as required in the definition.

Now, any subgroup of a solvable group is solvable.[49] Since A_5 is a subgroup of S_5 that is not solvable, S_5 is not solvable either.

Having built up all of the theory required, we can now present the proof we are interested in very quickly.

5.1.10 The unsolvability of the quintic

The result we are interested in is the claim that the generic quintic polynomial, F_5, is not solvable by radicals.

1. (Fundamental Theorem of Galois Theory) There is a one-one Galois correspondence between (normal) subgroups of a Galois group and (normal)

[48]For ease of reference, I repeat the definition here: A group G is *solvable*$_G$ if it has a finite series of subgroups: $1 = G_0 \subset G_1 \subset G_2 \subset \ldots \subset G_n = G$, such that G_i is normal in G_{i+1} and the quotient group G_{i+1}/G_i is abelian for $i = 0, \ldots, n–1$ [Stewart 2015: 161, Defn. 14.1].

[49]See for example [Rotman 1994: 102, Theorem 5.15].

subfields of the relevant Galois field extension.

2. (Translation of solvability) It follows from the Galois correspondence that if the Galois group of F_n is not $solvable_G$, then F_n is not $solvable_R$.

3. (Fact about the quintic) The Galois group of F_5 is S_5, the symmetric group of degree 5.

4. (Fact about the symmetric group of degree 5) S_5 has A_5 as a subgroup.

5. (Result from group theory) It follows from S_5's having A_5 as a subgroup that S_5 is not $solvable_G$.

6. (Conclusion) Hence F_5 is not $solvable_R$.

Before giving my own analysis of the explanatory value of this proof, I will first examine how Lange, Steiner and Pincock attempt to handle the case.

5.2 Salient features and characterizing properties

In this section, I'll examine Lange and Steiner's accounts of the Galois theory case. Both authors briefly address the example and take it to be a genuine case of explanation in mathematics.

5.2.1 A salient feature of $n \geq 5$

Lange discusses the Galois theory case only briefly in an endnote. He says:

> 'Regarding various proofs of the fact that the quintic equation is not generally solvable in radicals, the proof using Galois theory is often cited by mathematicians ... and philosophers ... as going beyond previous proofs to *explain why* the quintic is not generally solvable (and hence why, despite great efforts, no general algebraic formula for solving it has ever been found). I suggest that the distinction between an explanation and a mere proof of the quintic's unsolvability is grounded in this result's most striking feature: that

> in being unsolvable, the quintic differs from all lower-order polynomial equations (the quartic, cubic, quadratic, and linear), every one of which is solvable. In the context of this salient difference, we can ask why the quintic is unsolvable. Accordingly, an explanation is a proof that traces the quintic's unsolvability to some other respect in which the quintic differs from lower-order polynomials. Therefore, Galois's proof (unlike Abel's, for example) explains the result. That the distinction between an explanation and a mere proof depends upon the salience of the difference in solvability between quintics and lower-order polynomials is strongly suggested by the why questions's typical form: "Why? What happened to the pattern of algebraic equations each having solutions? What is it about the fifth degree that causes the problem? Why does it then go on to affect all higher degrees of equations?"'. [Lange 2017: 440, quoted material in Pesic 2003: 3]

Now, it seems right that the Galois theory proof 'traces the result' to the fact that the quintic differs from lower-order polynomials in being unsolvable. But this is precisely the fact we were trying to prove; any successful proof will identify a property that the quintic has and lower-order polynomials lack. Lange's point here is that the Galois theory proof identifies a certain associated property, P, which holds precisely when solvability by radicals holds and fails precisely when solvability by radicals fails. As we saw in the previous section, property P is $solvability_G$. By identifying property P, the Galois theory proof provides a condition that can be applied to each n in order to determine whether the generic polynomial of degree n is solvable. This helps us to understand both why solving by radicals worked for early cases, and why it failed at $n = 5$.

By contrast, the Abel proof deals with each case in a different way, obscuring the crucial shift at $n = 5$ (or so, presumably, thinks Lange). In fact, though, it is plausible that we can also understand the Abel proof as providing an 'associated property' – in this case the satisfiability of the solution formula, as D'Alessandro has recently pointed out.

D'Alessandro writes:

> 'The main novel claim here is that Galois's proof of the unsolvability of the quintic, though not Abel's, works by identifying and exploiting some difference between fifth-degree and lower-degree equations. Hence only Galois's proof is explanatory. This is, I think, mistaken. I can't discuss Abel's proof in detail here – for that, see [Pesic 2013] – but the general idea is as follows. First Abel shows that any solution formula for an equation of degree m must have the form
>
> $x = p + R^{\frac{1}{m}} + p_2 R^{\frac{2}{m}} + ... + p_{m-1} R^{\frac{m-1}{m}}$,
>
> where 'p, p_2, ... are finite sums of radicals and polynomials and $R^{\frac{1}{m}}$ is in general an irrational function of the coefficients of the original equation' ([Pesic 2013], p. 90). (The already-known formulas for lower-degree equations can easily be seen to have this form.) Hence, if there were a solution formula for the general quintic, it would look like
>
> $x = p + R^{\frac{1}{5}} + p_2 R^{\frac{2}{5}} + p_3 R^{\frac{3}{5}} + p_4 R^{\frac{4}{5}}$.
>
> The rest of the proof shows that this can't occur. Abel's main tool here is a theorem of Cauchy, which implies that $R^{\frac{1}{m}}$ can have only two or five distinct values when the roots of the quintic are permuted. Abel shows that each of these possibilities leads to a contradiction; it follows that the general quintic is unsolvable.
>
> *Pace* Lange, this argument strikes me as 'a proof that traces the quintic's unsolvability to some other respect in which the quintic differs from lower-order polynomials'. The respect in question is the satisfiability of the equation $x = p+R^{\frac{1}{m}}+p_2 R^{\frac{2}{m}}+...+p_{m-1} R^{\frac{m-1}{m}}$ for $m < 5$, as compared to its satisfiability for $m \geq 5$. By Lange's criterion, then, it's unclear why Galois's proof should count as any

> more explanatory than Abel's.' [D'Alessandro 2018: 19. The reference should be to Pesic 2003].

I agree with D'Alessandro here, taking the relevant associated property P' to be satisfiability of Abel's formula. That is, Lange's analysis of the Galois theory case does not succeed in drawing the distinction he is after between the Galois and Abel proofs of the result.

It is curious that Lange does not make use of the salient features he himself proposes in order to analyse the Galois theory example: it seems to me that symmetry in particular plays a crucial explanatory role. The Galois theory proof explicitly focuses on symmetries in the roots of polynomials, as we saw in the previous sections, and the Galois group for the generic polynomial of degree n is precisely the symmetry group S_n.

The reason Lange doesn't propose symmetry as the salient feature in this case is presumably because it doesn't look like symmetry is 'contained' in the result about the unsolvability of the quintic. Certainly, symmetry does not seem like a salient feature in the statement of the theorem. So we have a potential counterexample to Lange's account just like the ones I suggested in Chapter 4, where the proof exploits symmetry – and symmetry seems to play an explanatory role – and yet where symmetry is not a salient feature of the result.

By insisting that a salient feature must be present in both the proof and the set-up of the theorem, Lange is unable to draw on the most natural way to fit the Galois theory case into his account of mathematical explanation. I suggest that Lange's focus on providing a general theory of mathematical explanation that covers all cases has prevented his account from providing insight into the Galois theory case.

I will discuss the role of symmetry in the Galois theory proof in more detail in Section 5.4. For now, I will move on to look at how Steiner handles the case.

5.2.2 The Galois group as a characterizing property

Steiner presents the example as follows:[50]

> 'Galois theory explains the prior results that the [generic] polynomial equation is solvable in radicals if and only if the equation is of degree less than five (special cases of this had been proved by Cardan and Abel). The explanation assigns a group of automorphisms to each equation, and studies the groups instead of the equations. Let E be an equation with coefficients in a field F. We can demonstrate the existence of a smallest field K containing F in which E can be factored and thus solved (not necessarily in radicals) – call this the splitting field of E. The group G of automorphisms of K which leave F alone is called the Galois group of E. Now an automorphism of K (i.e. any member of G) is determined by its action on the roots of E (for K is the smallest field containing F and the n roots). Also, it is obvious that any member of G maps a root of E onto (the same or another) root of E. Thus the Galois group of E is a certain subset of the permutations of the roots of E. A final definition from group theory: G is solvable if it is the culmination of a finite chain of groups
>
> $0 \subset G_1 \subset G_2 \subset \ldots \subset G_n = G,$
>
> in which each group is a normal subgroup of the next (H is a normal subgroup of G if $GHG^{-1} = H$). The explanation of Cardan's and Abel's results consists of the following triad:
>
> 1. An equation is solvable in radicals if and only if its Galois group is solvable.
>
> 2. The Galois group of the [generic] polynomial equation of degree n consists of every possible permutation on n different objects. This group is called the symmetric group on n letters.

[50] I have substituted 'generic' where Steiner uses 'general' for clarity, since it is clear that this is what his proof actually refers to.

3. The symmetric group on n letters is solvable if and only if n is less than five (this is the whole reason for this detour into group theory; groups are easier to study than equations). But this example, in fact, supports our view. If we take the family $E(n)$ of equations of degree n, then the Galois group of $E(n)$ – the symmetric group on n letters – characterizes each equation as required.' [Steiner 1978: 148-9]

Steiner moves very quickly through the mathematics here, and he has missed out one of the crucial conditions on the chain of groups, namely that the quotient group G_{i+1}/G_i must be abelian for $i = 0, \ldots, n–1$. However, this need not prevent us from analysing Steiner's philosophical interpretation of the result.

Recall from Chapter 3 my reading of Steiner's three conditions on explanation:

1. The proof makes reference to a characterizing property of an entity or structure mentioned in the theorem.

2. If we substitute an entity from the same domain which lacks the characterizing property, then the theorem collapses.

3. The same argument applied to other objects with the same characterizing property is a proof of the modified proposition that is now the conclusion.

Steiner suggests that 'the Galois group of $E(n)$ – the symmetric group on n letters – characterizes each equation as required' [Steiner 1978: 149].

Recall from Chapter 3 that Steiner's account allows for partial characterization. As Steiner puts it,

> '... an arbitrary equation with rational coefficients has not a unique Galois group, in the sense that no other equation has it ... . The concept of 'characterization' will have to be weakened to allow for partial characterization. The Galois group of E characterizes it in that the properties of the Galois group tell us much about E' [ibid.: 149-50].

Now, the generic polynomial F_5 is certainly mentioned in the theorem. Second, it is clear that if we substitute in a generic polynomial of degree 4 or less, the theorem will collapse, as the same proof method shows that the Galois group in these cases is $solvable_G$, and hence that the polynomial is $solvable_R$.

Finally, if we substitute in a generic polynomial with degree higher than 5, the proof generalizes: we can apply the same reasoning to get to the relevantly modified theorem, namely that the polynomial is not solvable by radicals. We can take, for example, the generic polynomial of degree 6 and adapt the proof from Section 5.1.10: we simply need to amend steps 3 to 5 by considering the symmetric group S_6 instead of S_5.

However, Pincock disagrees with me that the Galois theory proof generalises according to Steiner's three conditions. Pincock writes:

> 'Steiner's notion of dependence requires the existence of an appropriate family of proofs. The problem with this proposal is that the explanatory power of the Galois theory proof ... does not hinge on the existence of this family of proofs. As Steiner notes, there is no effective means to move to a new Galois group and determine the polynomial equations associated with it. So there is no family of proofs that can be generated along the lines that Steiner requires. I conclude that this sort of explanatory proof is explanatory in its own right. Whatever dependence relations it describes are not relations that require the existence of a family of proofs.' [Pincock 2015: 8]

The worry seems to be that the Galois theory proof does not generalize to cover a family of proofs, because we can't simply change the Galois group and come up with a new result that can be proved by the same method.

However, on my analysis above we are actually interested in substituting in a different polynomial equation to the Galois theory proof, rather than substituting in a new Galois group. It is important to bear in mind here that there are two different senses in which we might hope for a proof to solve the case of a particular polynomial equation:

(1) Given a Galois group H that is not *solvable*$_G$, the proof gives you a way to determine coefficients for a polynomial f that has H as its Galois group (and hence is not *solvable*$_R$). Or more generally, the proof gives you a way to pick coefficients so that the polynomial with those coefficients is not *solvable*$_R$.

(2) Given a polynomial f, the proof gives you a way to determine whether its Galois group is *solvable*$_G$, and hence whether the polynomial is *solvable*$_R$.

It is true that the Galois theory proof presented in Section 5.1.10f does not provide a solution of type (1). This is simply a fact about the proof: 'there is no obvious way to find an equation with rational coefficients which has a given group as its Galois group – an unsolved Hilbert problem', as Steiner points out [Steiner 1978: 149]. As Pincock puts it, the Galois theory proof 'suffices to show that unsolvable polynomials exist, but leaves their coefficients a mystery'. [Pincock 2015: 7][51]

Since the Galois theory proof does not provide a solution of this type, and given that Steiner recognises this, it seems reasonable to suppose that Steiner's account does not rest on the availability of a solution of this type. Rather, the proof must generalise for Steiner in the sense that it provides a general solution of type (2).

As discussed above, it looks like the proof does indeed generalize for generic polynomials in this second sense: the Galois theory proof can be straightforwardly adapted (or 'deformed') for every n to identify the Galois group, S_n, for the generic polynomial of degree n. For each n we can find out whether S_n is *solvable*$_G$ by looking at its subgroups, in particular A_n. Hence, for each n we can find out whether the generic polynomial of degree n is *solvable*$_R$. The 'deformation' here is straightforward.

However, Steiner's account does run into some difficulty when we consider specific polynomials like $g(t) = t^4 - 2$. In this case, as we saw in Section 5.1.4, finding the Galois group for $g(t)$ took some effort. This becomes ever more

[51]Pincock's worry here is similar to that of the original referees of Galois' paper: as Stewart puts in, 'The referees wanted some kind of condition on the *coefficients* that determined solubility; Galois gave them a condition on the *roots*. The referees' expectation was unreasonable. No simple criterion based on the coefficients has ever been found, nor is one remotely likely.' [Stewart 2008: 106]

difficult when we take more complex cases (as Stewart puts it, ‘anything more complicated [than $g(t)$] would be unwieldy’ [Stewart p. 155]).

The worry is that the Galois theory proof that some *specific* polynomial of degree n is not $solvable_R$ is not a straightforward ‘deformation’ of the Galois theory proof that the *generic* polynomial of degree n is not $solvable_R$. Further work is needed to modify the proof in these cases, which Steiner’s three conditions do not seem to allow for.

Now, Steiner could simply respond that his account was only ever intended to provide for the explanatory value of the general proof. But this is a problem because handling the specific case is one of the crucial ways in which the Galois theory proof improves upon Abel’s earlier version. As Stewart points out, ‘for all Abel’s methods could prove, every particular quintic equation might be soluble, with a special formula for each equation’ [Stewart 2015: 5], while ‘[Galois’s] method went much further: it applies not just to the general polynomial $F(t)$, but to any polynomial whatsoever. And it provides necessary *and sufficient* conditions for solutions by radicals to exist’ [ibid.: 120, emphasis in original].

Neither Steiner nor Lange successfully account for this feature of the Galois theory proof, and as we have seen, neither of them spend much time on the example. Pincock provides the most detailed analysis so far, and I will look at his account of the case in the next section.

5.3 Pincock on Galois

5.3.1 A worry about proof chauvinism

Pincock sets out to identify ‘what is special about abstract explanations and their explanatory power’, as he believes that the Galois theory proof is an instance of an *abstract mathematical explanation*, at least in the sense that abstract explanations ‘share some core features with the Galois theory proof’ [Pincock 2015: 2]. For Pincock, the explanatory value of the Galois theory proof comes from displaying an ontological dependence between less and more abstract objects. For example, he claims that 'the Galois theory proof is an explanatory proof because it invokes a special sort of ontological dependence between distinct mathematical domains', and 'the Galois theory proof explains because there is a special sort of dependence relation between facts about groups and facts about polynomial equations.' [Pincock 2015: 3, 7]

In a recent paper, D'Alessandro worries that Pincock focuses too heavily on the idea that the Galois theory *proof* is explanatory. This is part of D'Alessandro's general worry that too much of the literature on mathematical explanation has been too quick to assume something like the following: 'Proof Chauvinism: All or most cases of mathematical explanation involve explanatory proofs in an essential way.'

D'Alessandro writes:

> 'Pincock takes it for granted that the case involves an explanatory proof, and he fixes on Galois's proof of the unsolvability of the quintic as the relevant item. ...
>
> I think this is a mistake. Galois theory isn't explanatorily successful by virtue of containing an explanatory proof. Instead, I claim that the explanantia in this case are theorems, notably Galois's criterion itself and his results about the solvability of symmetric groups.' [D'Alessandro 2018: 16-17]

D'Alessandro presents a number of quotes from mathematicians suggesting that their view of the Galois case is indeed that the result or theory explains, and not the proof.

However, even if it's right that the theorem itself does explanatory work, this does not rule out that *both* the proof and the theorem play an explanatory role.

On the one hand, we might think the Fundamental Theorem of Galois Theory (which says that there is a one-one correspondence between (normal) subgroups of a Galois group and (normal) subfields of the relevant Galois field extension) explains why there is a one-one correspondence in the particular case $n = 5$ (at least in the sense of explanation as instantiation, for example as proposed by Räz [2018]).

On the other hand, it also seems to me that the proof does explanatory work: it is the details of the specific construction of the bijection that explain why group-theoretic properties translate into facts about fields. The mere existence of the bijection alone does not explain this.

I will develop this suggestion in more depth in Section 5.4. For now, note that even if D'Alessandro accepts that the Galois theory proof can do explanatory work, he seems to take particular issue with the thought that Galois himself presented an explanatory proof (rather than an explanatory theorem or theory).

D'Alessandro writes:

> 'In any case, the suggestion that a proof was Galois's main explanatory contribution is problematic by itself. For one, as Leo Corry observes, 'Galois's writings were highly obscure and difficult, and his proofs contained many gaps that needed to be filled' ([Corry 2004], p. 26). Presumably, an argument as problematic and incomplete as the one Galois actually gave is a poor candidate for an explanatory proof. Presumably, also, this is why Pincock presents a modern field-theoretic version of the proof of Galois's criterion, even though the reasoning is substantially different from what Galois himself could have had in mind.' [D'Alessandro 2018: 18]

But Pincock surely isn't claiming that 'a proof was Galois's main explanatory contribution'. Pincock's paper is not a historical treatise; rather, I think Pincock's idea is that the Galois(ian) proof contains ideas (or displays dependence relations) that explain the result in a way that the earlier Abel-Ruffini proof didn't. This is independent of whether Galois's own presentation was sufficiently rigorous to count as a proof. So, I don't think D'Alessandro's worry presents a serious problem for Pincock's account.

In the next section, I will explore Pincock's suggestion that the Galois theory case is an instance of an abstract mathematical explanation.

5.3.2 Abstractness as a partial order

According to Pincock, a mathematical proof can be more explanatory than another proof 'by invoking a more abstract kind of entity than the topic of the theorem' [Pincock 2015: 1].[52] I think Pincock is right to focus on abstraction as a crucial feature of the Galois theory proof, and this has further support in the literature. As Pesic puts it:

> 'What is new in Galois is a turn toward abstraction in an essentially modern way, leading to a complete understanding of solvability, which Abel lacked. As he reformulated and extended Abel's work, Galois found it expedient to discuss permutations of the roots of equations not case by case but through a master abstraction that would encompass many permutations at once.' [Pesic 2003: 108-9]

The important question for Pincock's account is whether (and how) this abstraction plays an explanatory role in the proof. To analyse this, we need to look at Pincock's understanding of abstraction in more detail. Pincock writes:

> 'Schematically, I will represent a fact of type X with constituent x using '$X_i(x)$', where 'i' is an index used to distinguish that fact from another fact of that type. In special cases it turns out that (i) for any x, y, given $R(x, y)$, $((X_i(x) \leftrightarrow Y_j(y))$. That is, for a suitably constrained set of facts of type X and type Y, once the R relation obtains between two appropriate objects, facts of those sorts are paired up in the following sense: one such fact obtains just in case a fact of the other type obtains.

[52] Pincock also calls the 'topic' of the theorem the 'subject-matter', bringing to mind Steiner's condition that an explanatory proof must refer to entities 'mentioned' in the theorem. Steiner focuses on entities that are mentioned, though, while Pincock is interested in entities that are not mentioned, but which illustrate a certain kind of dependence relation that helps to make a proof explanatory.

> ...
>
> First, suppose that we wish to fix the explanatory grounds of facts of type Y with constituents y. I will say that each collection of type X facts that are mentioned in the true biconditionals of form (i) noted above constitute some *potential* explanatory grounds for the Y facts. To choose between these potential grounds I invoke a notion of abstractness. That is, one object can be more abstract than another. The explanatory grounds will be given by facts of type X whose constituents x are *more* abtract than y but *less* abstract than objects x drawn from any other potential explanatory grounds. So, the idea is that the unique explanatory grounds for facts of type Y are the least more abstract domain of facts where an appropriate biconditional obtains.' [Pincock 2015: 12, emphasis in the original]

How does this work for the Galois theory proof? It looks like there are a number of relevant biconditionals to choose from. Let K be a subfield of $\mathbb{C}$ (such as $\mathbb{Q}$).

1. Facts about polynomials $\leftrightarrow$ facts about Galois groups. For example, for any polynomial equation y over K, Galois group x, and relation $R(x, y)$ which holds exactly when x is the Galois group of y over K, we have $R(x, y) \rightarrow (X_i(x) \leftrightarrow Y_j(y))$: x is a solvable group if and only if y is solvable by radicals.

2. Facts about polynomials $\leftrightarrow$ facts about field extensions. For example, for any polynomial equation y over K, field extension x, and relation $R(x, y)$ which holds exactly when x is the splitting field of y over K, we have $R(x, y) \rightarrow (X_i(x) \leftrightarrow Y_j(y))$: x is contained in a radical extension of K if and only if y is solvable by radicals.

3. Facts about field extensions $\leftrightarrow$ facts about Galois groups. For example, for any field extension y of K, Galois group x, and relation $R(x, y)$ which

holds exactly when x is the group of all K-automorphisms of y, we have $R(x, y) \rightarrow (X_i(x) \leftrightarrow Y_j(y))$: x is a solvable group if and only if y is contained in a radical extension of $\mathbb{Q}$.

Now, we need a way to label one side of each biconditional as more abstract than the other, in order to identify the explanatory grounds on Pincock's account. As Pincock acknowledges:

> 'For this proposal to work, objects must be partially ordered by their abstractness. To clarify this ordering, I draw on the broadly structuralist thought that some objects can have other objects as instances. When a has b as an instance, I say that a is more abstract than b. The instantiation relation here is irreflexive, asymmetric and transitive[53]. Perhaps the clearest case of what I have in mind is the type-token relationship [where the type is more abstract than the token].' [Pincock 2015: 12]

Pincock takes biconditional (1) to be the best description of the relevant explanatory correspondence in the Galois theory proof. That is, he takes facts about polynomials (for example, that the generic quintic is not solvable by radicals) to be explained by facts about Galois groups (for example, that S_5 is not a solvable group).

However, recall from the definition in Section 5.1.4 that strictly speaking, a Galois group is defined of a field extension, not of a polynomial.[54] In fact, then, to get from the domain of facts about polynomials to facts about Galois groups we must proceed via the domain of facts about field extensions – in other words, we must proceed via biconditionals (2) and (3).

[53]This is a bit curious, because instantiation is not transitive in general: for example, a particular token of the letter A might instantiate the type *A*, and the type *A* might instantiate the property of 'being an English letter type', but the token A does not instantiate this property, since it is a token and not a type. Similarly, I am a token of the human species, and the human species instantiates the property of 'being a type of biological species', but I do not instantiate this property as I am not a type of species. Thanks to Marcus Giaquinto for this point.

[54]For ease of reference: The Galois group of a field extension $L : K$ is the group of all K-automorphisms of L under the operation of composition of maps.

At first glance, this doesn't seem to be a big problem for Pincock: after all, biconditional (2) is 'merely' definitional[55], and biconditional (3) is guaranteed by the Galois correspondence. Pincock's instantiation relation is supposed to be transitive, so we simply need to check that the order of abstractness increases in the right direction through the chain of biconditionals.

But in fact this is not straightforward, as there is no clear instantiation relation in any of the three biconditionals. Polynomials are not instances of Galois groups; polynomials are not instances of field extensions; and field extensions are not instances of Galois groups.

Pincock admits this much, writing that

> '... we cannot say that polynomial equations are instances of groups. Each polynomial equation gives rise to a field extension, and each field extension determines a collection of automorphisms. I claim that each collection of automorphisms is an instance of a group. So, in our stipulated sense, groups like S_5 are more abstract than the associated collection of field automorphisms.' [Pincock 2015: 13]

It is true that a suitable collection of field automorphisms – together with a group operation – is an instance of a group. However, this suffices only to show that facts about collections of field automorphisms are explained by facts about Galois groups, on Pincock's account.[56] We haven't yet got to the result that facts about *polynomials* are explained by facts about Galois groups. But surely it is precisely facts about polynomials that the Galois theory proof is thought to explain.

The worry here is that the technical details of Pincock's account of abstraction don't quite seem to work when analysing the Galois theory proof. This

[55]Recall the definition: Let f be a polynomial over a subfield of $\mathbb{C}$, and let Σ be the splitting field for f over K. We say that f is soluble by radicals if there exists a field M containing Σ such that $M : K$ is a radical extension.

[56]In fact, there is an added complication that in each biconditional the explanatory grounds must not simply be more abstract, but the *least* more abstract – Pincock acknowledges that 'there still might be ... a kind of object that is more abstract than the collections of field automorphisms and yet less abstract than groups' [Pincock 2015: 13].

worry also arises when we look at the details of his mathematical claims. For example, Pincock writes:

> '... the crucial property that all these S_n groups have in common, when $n \geq 5$, is that they have a subgroup A_n that is not the size of a prime number. The size of each S_n group is $n!$, and the size of each subgroup A_n is $n!/2$. If $n \geq 5$, then $n!/2$ is not prime. In some sense, this is the ultimate reason that [every polynomial of degree n is solvable by radicals if and only if $n = 2, 3, 4$]'. [Pincock 2015: 6]

But this can't be right, because $n!/2$ is not prime for $n = 4$ either: the size of A_4 is 12. Pincock has hit upon the wrong feature here. The crucial properties of A_n for $n \geq 5$ are that A_n is simple and non-abelian; it follows from this that A_n has non-prime order for $n \geq 5$, but it does not follow that A_n has prime order for $n < 5$. That is, the (non-)primeness of the order of A_n is not the difference maker between $n = 2, 3, 4$ and $n \geq 5$.

I suggest that Pincock has not paid careful enough attention to the details of the Galois theory proof. Nevertheless, I am inclined to agree with his central idea that the abstraction in the Galois proof in some way adds to the proof's explanatory power, and I will continue to explore this idea in my own way in the next section.

5.4 A positive suggestion

I suggest that abstraction and symmetry both play an explanatory role in the Galois theory proof.

Abstraction plays an explanatory role, I claim, by allowing us to ignore certain irrelevant properties of the field extension when it is convenient to do so. In particular, the proof abstracts away from all of the non-structural properties of the field extension: the group representation includes only information about the relations between the elements (such as the existence of a chain of subgroups), and no extraneous information about the elements themselves (whether they are interpreted as field automorphisms or rotations of a square).

By focusing in on these structural relations, the group representation allows us to pinpoint the crucial difference-making features that determine whether the corresponding polynomial is solvable by radicals. The polynomial is $solvable_R$ exactly when the group representation of the collection of relevant field automorphisms – the Galois group – is $solvable_G$, that is, exactly when the Galois group has a certain structure (having a chain of solvable subgroups).

The properties of solvability in both domains are abstract in the sense of being structural and not dependent on the natures of the elements (whether groups or fields). This, I claim, is part of what makes the Galois proof explanatory: the group representation allows us to identify and zoom in on the relevant structural properties.

This interpretation is supported by Stewart's note that:

> 'Galois introduced a new point of view into mathematics ... he took a necessary but unfamiliar step into abstraction. In Galois's hands, mathematics ceased to become the study of numbers and shapes – arithmetic, geometry, and ideas that developed out of them like algebra and trigonometry. It became the study of *structure*. ... he was the first person seriously to appreciate that mathematical questions could sometimes be best understood by transporting them into a more abstract realm of thought.' [Stewart 2015: 110-11, emphasis in the original]

I suggest that by allowing us to identify the relevant difference-making features, the Galois theory proof helps us to understand the following two features of polynomials: (1) why solvability fails at $n = 5$, and not earlier in the series $n = 2, 3, 4, ...$ (or indeed later); (2) why some specific quintic polynomials are solvable and some are not. Abel's proof fails to account for these two features because of his failure to push towards abstraction.[57]

[57] This is not to say that Abel's ideas would not eventually have led him to a similar

Stewart's words confirm my view of the importance of the first feature just mentioned: 'Not only does [Galois theory] prove that the general quintic has no radical solutions, it also explains why the general quadratic, cubic and quartic *do* have radical solutions and tells us roughly what they look like' [Stewart 2008: 116, emphasis in the original].

The Galois theory proof identifies a crucial change in structure at $n = 5$ which, I suggest, explains the difference between the cases $n = 2, 3, 4$ and $n \geq 5$ for the generic polynomial. This change in structure occurs because A_5 is non-abelian and simple – that is, A_5 has no chain of normal subgroups with abelian quotients. The breakdown occurs precisely here because A_5 is the 'first' (as in smallest) such group. By allowing us to zoom in on the crucial structural properties of A_5 as compared to A_4, abstraction plays a crucial facilitating role in aiding our cognitive access to the relevant features and helping us to understand why solvability fails precisely at $n = 5$.

Turning now to the second claim: abstraction also helps us to understand the difference between specific solvable and unsolvable quintic polynomials, I claim. We have already seen that being able to handle the specific case was a crucial improvement of the Galois theory proof over the Abel-Ruffini proof. In Stewart's words again: 'for all Abel's methods could prove, every particular quintic equation might be soluble, with a special formula for each equation' [Stewart 2015: 5], while '[Galois's] method went much further: it applies not just to the [generic] polynomial $F(t)$, but to any polynomial whatsoever. And it provides necessary *and sufficient* conditions for solutions by radicals to exist' [ibid.: 120, emphasis in the original]. Similarly, Pesic writes 'This was Galois's great advance over Abel, specifying exactly which equations are solvable and which not.' [Pesic 2003: 130]

It seems right to think that this improvement is an explanatory one – one of the explananda we started with was the observation that certain specific polynomials were solvable by radicals, while for others no suitable formula

account: as Pesic points out, he had crucial early insights about the importance of the property of being abelian, and of course this property bears his name. But unfortunately Abel died young, like Galois.

could be found.

Let us therefore examine the specific case and see exactly how the Galois theory proof helps us to handle it.

First, consider a specific quintic polynomial, $h(t) = t^5 - 2$. The splitting field Σ_h of $h(t)$ is $\mathbb{Q}(\sqrt[5]{2}, \omega)$, where ω is the fifth root of unity, and the Galois group G_h of h consists of the identity automorphism and the automorphism $\phi : \Sigma_h \mapsto \Sigma_h$ such that $\phi(\sqrt[5]{2}) = -\sqrt[5]{2}$ and all rationals remain fixed, together with the operation of composition of maps. The Galois group is $solvable_G$: the required chain of subgroups is simply the identity and G_h itself. Hence, the polynomial h is $solvable_R$. As before, the abstraction to group theory has allowed us to focus only on the relevant structural properties of the Galois group and by extension the collection of field automorphisms, aiding our epistemic access to the difference-making features.

In this example, we could in fact have gone almost straight to the conclusion that h is solvable by radicals, because $\omega^5 = 1 \in \mathbb{Q}(\sqrt[5]{2})$ and $(\sqrt[5]{2})^5 = 2 \in \mathbb{Q}$: the splitting field itself is a radical extension of $\mathbb{Q}$. This doesn't mean that Galois theory wasn't involved – after all, the splitting field is also known as the Galois extension field because it is so closely linked to the Galois group.[58]

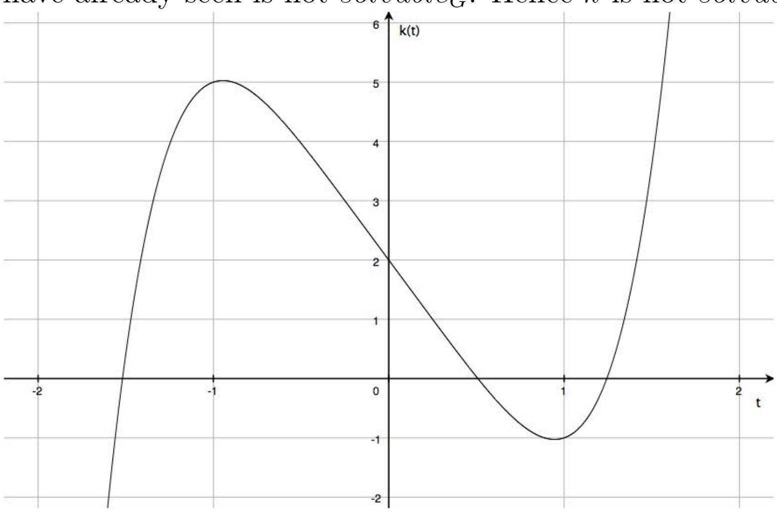
Let us now examine an example of an unsolvable quintic, $k(t) = t^5 - 4t + 2$. First, consider the following theorem:

> **Theorem:** If f is an irreducible rational polynomial of degree 5 with exactly two non-real roots, then the Galois group of f is the full symmetric group S_5.
>
> **Proof sketch:** The Galois group of f is a subgroup of S_5 (this is true in general). Complex conjugation fixes the three real roots of f, and the two non-real roots of f are complex conjugates (otherwise there would be another non-real root, since the complex conjugate of any root is also a root). This means that the Galois

[58] The problem with taking a more complicated example is that the Galois group quickly becomes 'unwieldy', in Stewart's terms. This provides further support for my claim that abstracting away from the irrelevant details aids our epistemic access.

group contains a transposition of two of the roots of f. Since f is irreducible, its Galois group contains a permutation of order 5 and hence contains a 5-cycle.[59] The only subgroup of S_5 that contains a transposition and a 5-cycle is S_5 itself. Hence the Galois group of f is the full symmetric group S_5.

By examining the graph of $k(t) = t^5 - 4t + 2$ we can see[60] that it crosses the t-axis three times, so k has exactly three real roots (that is, exactly two non-real roots). So the Galois group of k is the full symmetric group S_5, which we have already seen is not $solvable_G$. Hence k is not $solvable_R$.



Here we have directly appealed to the structure of the Galois group of the specific polynomial, k, to establish that k is not solvable: its Galois group is S_5, which has no suitable chain of solvable subgroups. We can therefore explain the difference in solvability between the two specific quintic polynomials h and k in the same way as we explained the break between $n = 4$ and $n = 5$ for the generic polynomial.

Interestingly, Steiner also considers polynomial k, but reasons that we can draw conclusions about a specific polynomial without actually having to de-

[59] A 5-cycle is simply a permutation of S_5 that maps all elements to each other in a cyclic fashion. For example, $1 \to 2 \to 3 \to 4 \to 5 \to 1$, or (12345).

[60] This 'seeing' is not tied to the diagram; we can also consider changes in sign of the derivative. However the diagram is helpful here, and since t and $k(t)$ range over all complex numbers, we need not worry about $k(t)$ failing to intersect with the axis (as we would if they ranged only over the rationals).

termine its Galois group at all. Steiner writes:

> 'If E is a solvable, irreducible equation of [odd] prime degree q with [rational] coefficients, then it either has exactly one real root, or all its roots are real. (For example, consider $x^5 - 4x + 2 = 0$. Sketching $x^5 - 4x + 2$, we can see that it crosses the x-axis three times. Thus the equation cannot be solved.)' [Steiner 1978: 149]

The fact that solvable and irreducible polynomials of odd prime degree have either one or all real roots is known as Kronecker's polynomial theorem.[61] The proof for $n = 5$ follows from the theorem I sketched above: we already know that if the polynomial has exactly two non-real roots, it is not solvable by radicals. The polynomial must have an even number of non-real roots, since the non-real roots come in complex conjugate pairs. So if it is solvable, it must have either four or zero non-real roots. That is, it must have either exactly one real root, or all of its roots are real. So in fact, I suggest that Steiner's method of dealing with k does make use of the specific Galois group S_5, albeit indirectly.

Let me summarise the current state of progress: I have claimed that abstraction plays an explanatory role in the Galois theory proof in an epistemic sense, in affording us access to relevant explanatory features. As Saatsi puts it, we might describe abstraction as playing a *thin explanatory role* [Saatsi 2016: 1056].

I now suggest that symmetry plays what we might call a *thick explanatory role*, in the sense that it is an objective feature of the mathematical entities involved and plays an explanatory role independently of whether we are able to appreciate that it does so. To clarify, symmetry is an objective feature of mathematical entities in the sense that its presence is independent of our cognitive capacities and background, but of course some people are better at recognising symmetry than others. Trained mathematicians will be quick to see symmetries in group representations and isomorphisms, while a lay person

[61] See e.g. http://mathworld.wolfram.com/KroneckersPolynomialTheorem.html

might find the symmetry to be obscured behind the symbolic notation. I will be interested here in the objective symmetry, however it is represented.

There is ample support in the literature for the idea that symmetry is crucially involved in the Galois proof. As mathematician Edward Frenkel puts it, 'What Galois had done was bring the idea of symmetry, intuitively familiar to us in geometry, to the forefront of number theory. What's more, he showed symmetry's amazing power' [Frenkel 2013: 76]. Stewart confirms that 'In modern terms, Galois's main idea is to look at the symmetries of the polynomial $f(t)$' [Stewart 2015: 208], and in the mathematics education literature Galois theory is described as a 'theory of symmetry' [Leuders 2016].

To explore my idea that symmetry plays an explanatory role in the Galois theory proof, I will first need to tease out the different ways in which symmetry is involved in the proof. Second, I will need to give some reason to believe that at least one of these ways adds to the explanatory power of the proof.

Now, the Galois group for the generic polynomial of degree n is the full symmetric group, S_n, as we have seen, and the Galois group for any specific polynomial of degree n is a subgroup of S_n. The elements of S_n are permutations and each represents a specific symmetry of the roots of the polynomial; here then is one example of symmetry in the Galois theory proof.[62] The abstraction allows us to ignore the 'detail' of the elements of the Galois group when it is convenient to do so, but the detail is explanatorily relevant: it is because each element of the Galois group represents a symmetry of the roots of the polynomial that the solvability of the Galois group exactly corresponds to the solvability of the polynomial.

Furthermore, the Galois correspondence itself displays another case of symmetry. As we saw from the lattice diagrams in Section 5.1.6, the Galois correspondence preserves structure, although inverting the inclusion relations: the relations between the (normal) subgroups of the Galois group exactly mirror

[62]This symmetry is not present in just any group – as Frenkel notes, 'It turns out there are many examples of groups that have nothing to do with symmetries, [even though this] was our motivation to introduce the concept of a group in the first place.' [Frenkel 2013: 47]

the relations between the (normal) subfields of the relevant Galois field extension. This clear case of mirror symmetry is explanatorily relevant because its underpins a correspondence that is objectively *direct* in the sense discussed in Section 4.3.3, namely not involving any intermediate entities. The failure of solvability for the polynomial is explained here by a directly related mathematical entity. Here I suppose I agree with Pincock's intuition that the 'closest' entity is explanatorily relevant, though I do not describe this in terms of instantiation.

On my positive proposal, the important insights afforded by the Galois theory proof are as follows: (1) abstraction can play an epistemically explanatory role, allowing us to focus on the crucial structural properties of the mathematical entities involved; (2) symmetry can play an ontically explanatory role, underpinning the direct Galois correspondence and explaining why solvability for polynomials and groups coincides. I conclude that it has been philosophically fruitful to play close attention to Galois theory as an example of mathematical explanation.

6 Conclusion

In this thesis, I have taken a constructive attitude towards existing philosophical accounts of explanation in the hope of making some progress towards a successful account of explanation in mathematics.

In Chapter 1, I examined four existing accounts of scientific explanation: grounding, counterfactual, unification, and why-question accounts. I presented a few simple examples from mathematics to see whether these accounts might transfer over to intra-mathematical explanation. I argued that each of the accounts faces some problems. The grounding account was seen to be problematic because it focused on ontic features and did not fit well with epistemic aspects of explanation, such as whether a proof helps us to understand why the theorem is true. The unification account was seen to be problematic because it did not allow for fine-grained differences in explanatory value between proof types which are equivalent, like the principle of induction and the well-ordering principle.

The counterfactual account did not transfer neatly to the intra-mathematical case, because it is hard to make sense of the antecedent of the relevant counterfactual in cases where a statement is necessarily true. Nevertheless, the central idea behind the counterfactual account that explanation involves identifying a difference-making feature is intuitively attractive from an ontic perspective on explanation, and I examined Steiner's extension of this intuition to the mathematical case in Chapter 3.

The why-question account involved a relevance relation which would need to do a lot of work in the intra-mathematical case, and it was difficult to restrict the relevance relation in a non ad hoc way to prevent the account overgenerating and counting all proofs as explanatory. Nevertheless, the central idea behind the why-question account that explanation is context-relative is intuitively attractive from an epistemic perspective on explanation, and I examined Lange's extension of this intuition to the mathematical case in Chapter 4.

In Chapter 2, I took a quick detour to consider inductive proof as a case study. I selected inductive proof because I noticed a fairly common view in the philosophical literature that inductive proofs are not explanatory. This view is sometimes so strong that inductive proofs are seen as a suitable test case for an account of mathematical explanation, in the sense that some philosophers believe no successful account of explanation should count inductive proofs as explanatory. However, the view is rarely given much evidential support beyond a strong feeling or intuition, so I felt it was important to examine this issue in more detail. In Chapter 2, therefore, I argued against Lange's argument for the claim that there are no explanatory inductive proofs, and I put forward some plausible examples of explanatory inductive proofs. I also defended the methodological point that attention to examples is important, and that intuitions about particular cases (of candidate mathematical explanations) should not be dispensed with in favour of exclusive reliance on intuitions about general principles (of mathematical explanation).

In Chapter 3, I provided an in-depth analysis of one of the earliest modern accounts of mathematical explanation, given by Mark Steiner. Although this account is often taken to be quite puzzling and hence rejected in the philosophical literature, I argued that Steiner's account repays deeper analysis by providing a sympathetic reading that makes sense of his puzzling remarks and draws out some important questions. I presented what I take to be the best reading of Steiner's three key conditions on explanation in mathematics, and I showed how one of his central examples fits this schema. Although Steiner's account is ontically motivated, I showed how my extension of his proposal can make room for an epistemic component that accounts for what I take to be the primary epistemic function of an explanation, namely, to help us see why the fact to be explained is true. This is an important achievement in my thesis, because I claim that any successful account of mathematical explanation should include an epistemic component.

In Chapter 4, I provided an in-depth analysis of one of the most recent accounts of mathematical explanation, given by Marc Lange. Lange puts forward a huge number of putative examples, and I examined some of these in detail, considering a case study for each of the three salient features Lange identifies: unity, symmetry and simplicity. I presented what I take to be the best reading of Lange's three key conditions on explanation in mathematics, and discussed some problem cases. I argued that unity and symmetry are not significantly distinct on Lange's account, and that his account of simplicity is unclear. I identified symmetry as the most plausible explanatory feature among those proposed by Lange. In this chapter I also considered Lange's recent comment that inductive proofs might after all count as somewhat explanatory, which seems to vindicate my view on inductive proof presented in Chapter 2. I proposed a novel way to make sense of this shift in perspective on Lange's behalf.

Although Lange's account is epistemically motivated, I suggested that we should see features like symmetry as objective mathematical properties that are independent of our interests, even when their salience is context-relative. This is an important point, because I claim that any successful account of mathematical explanation should include an objective or ontic component.

In Chapter 5, I conducted an in-depth case study of Galois theory, which I claim provides an excellent candidate for an explanatory proof. Of the relatively small number of philosophers active in the area of intra-mathematical explanation, a relatively large proportion consider the Galois theory case. I examine each of their proposals, suggesting that they all miss important insights owing to a lack of attention to the mathematical detail of the proof. On my positive proposal, the important insights afforded by the Galois theory proof are as follows: (1) abstraction can play an epistemically explanatory role, allowing us to focus on the crucial structural properties of the mathematical entities involved; (2) symmetry can play an ontically explanatory role, underpinning the direct Galois correspondence and explaining why solvability for polynomials and groups coincides. I conclude that it is philosophically fruitful to follow my strategy of careful consideration of case studies.

Now, in one sense the outcome of my thesis is negative. I have argued that a number of existing proposals do not satisfactorily account for explanation in mathematics. However, although I have provided a careful and critical analysis of existing accounts by identifying and discussing a number of problems in each case, I have maintained a positive perspective throughout. My aim has been to recognise and draw on insights from each account to inform a new proposal. Additionally, I have made a number of positive suggestions on each author's behalf. My guiding thought, given the relative ease in finding counterexamples to any proposed set of conditions, is that we should remain open to the idea that there may be no general or universal account of explanation in mathematics. Our progress towards an account of mathematical explanation should proceed by careful analysis of cases, as I have exemplified in my final chapter.

Finally, recall that at the very beginning of the thesis, I suggested that an account of intra-mathematical explanation might be useful to philosophy of science.

I was fascinated to discover that the specific symmetries in the Galois theory example are also taken to be explanatory features in some applications in chemistry. In particular, the Buckminster Fullerene molecule, $C60$, has exactly the same structure as the alternating group A_5. By allowing us to focus on the molecule's symmetries, as in the case of the quintic polynomial, the Galois group theoretic representation turns out to be crucial in helping us to understand the structure and properties of the Buckminster Fullerene molecule.

For example, Chung and Sternberg write that:

> 'Of course mathematical methods are indispensable in the study of all molecules, but buckminsterfullerene is a special case, where mathematics has proved extraordinarily effective in illuminating the structure and the properties of the molecule'. [Chung and Sternberg 1993: 56]
>
> ...

> 'The branch of mathematics called group theory is the most central to an understanding of the buckyball; group theory describes the symmetries of the molecule and thereby determines some of its most distinctive properties. Specifically, the symmetries of the buckyball govern its spectrum, which was the means by which the molecule was first detected and identified.' [Chung and Sternberg 1993: 56]
>
> ...
>
> 'The symmetries of the truncated icosahedron may seem quite remote from the physics and chemistry of carbon molecules. But in fact they are closely connected. The nature of the link can be made clear by an anecdote. Throughout the later years of the 1980s, when Curl and Smalley were struggling to make more than microscopic quantities of buckminsterfullerene, they described their quest as "the search for the yellow vial." When Krätschmer and Fostiropoulos finally produced a measurable quantity of fullerenes, their initial vial of dissolved material was red, but purified films of buckminsterfullerene were indeed yellow. The question arises how anyone knew what color to look for in a material that had never been seen before. The answer is that the color could be predicted from knowledge of the molecule's absorption spectrum. The spectrum in turn depends on the symmetries of the molecule.' [Chung and Sternberg 1993: 63]

My initial suggestion, drawing on insights gained from the intra-mathematical case, is that the object itself (the molecule) has certain symmetries, and its properties are dependent on these symmetries. The group representation serves as an abstraction, affording us epistemic access to the crucial structural features of the molecule.

It would be very interesting to examine this example in the context of indispensability arguments about mathematical explanation in science. For example, it looks like a prime example of genuinely mathematical explanation in science if Steiner is correct that 'only in [genuinely] mathematical explanation is this the case: when we remove the physics, we remain with a mathematical explanation – of a mathematical truth!' [Steiner 1978b: 19].

I leave this open as an avenue for future research.

7 References

Andrews, George and Eriksson, Kimmo (2014) *Integer Partitions*, New York: Cambridge University Press.

Baker, Alan (2009) 'Mathematical Accidents and the End of Explanation'. In Bueno, O. and Linnebo, Ø. *New Waves in the Philosophy of Mathematics*, Palgrave Macmillan, 137–159.

Baker, Alan (2010) 'Mathematical induction and explanation', *Analysis*, 70(4): 681–689.

Baker, Alan (2012) 'Science-Driven Mathematical Explanation', *Mind*, 121(482): 243–267.

Baker, Alan (2017) 'Mathematics and Explanatory Generality', *Philosophia Mathematica*, 25(2): 194–209.

Baker, Alan and Colyvan, Mark (2011) 'Indexing and Mathematical Explanation', *Philosophia Mathematica* (III) 19: 323–334.

Baldwin, John (2016) 'Foundations of Mathematics: Reliability and Clarity: the explanatory role of mathematical induction'. In: Väänänen J., Hirvonen Å., de Queiroz R. (eds) *Logic, Language, Information, and Computation. WoLLIC 2016. Lecture Notes in Computer Science*, vol 9803, 68-82. Springer, Berlin, Heidelberg.

Blåsjö, Victor (2016) 'Critique of Lange's theory of explanation', blog post on 4/2/2016, last accessed on 8/9/18 at http://intellectualmathematics.com/blog/critique-of-langes-theory-of-explanation/

Blåsjö, Victor (2018), *Intuitive Infinitesimal Calculus*, online textbook accessed on 8/9/18 at http://intellectualmathematics.com/dl/calculus.pdf

Brown, James R (1997) 'Proofs and pictures', *British Journal for the Philosophy of Science*, 48: 161–80.

Burge, Tyler (2010) *Origins of Objectivity.* Oxford: Oxford University Press.

Cariani, Fabrizio (ms) 'Mathematical Induction and Explanatory Value in Mathematics', self-archived paper available at http://cariani.org/files/MIEV.cariani.pdf

Chung, Fan and Sternberg, Shlomo (1993) 'Mathematics and the Buckyball', *American Scientist*, 81(1): 56–71.

Corry, Leo (2004) *Modern Algebra and the Rise of Mathematical Structures*, Basel: Birkhäuser.

D'Alessandro, William (2018), 'Mathematical Explanation beyond Explanatory Proof', forthcoming in *The British Journal for the Philosophy of Science*, online preprint: 1–23.

Dougherty, John (2017) 'What inductive explanations could not be', *Synthese*, 195(12): 5473–5483.

Dray, Tevian and Manogue, Corinne (2006) 'The Geometry of the Dot and Cross Products', *Journal of Online Mathematics and Its Applications*, published by the Mathematical Association of America, 6: 1–13.

Dreyfus, Tommy and Eisenberg, Theodore (1986) 'On the Aesthetics of Mathematical Thought', *For the Learning of Mathematics* 6: 2–10.

Dubeau, Francois (1991) 'Cauchy and mathematical induction', *International Journal of Mathematical Education in Science and Technology*, 22(6): 965–969.

Erdős, Paul (1934). 'On the density of the abundant numbers'. *Journal of the London Mathematical Society*, s1-9(4): 278–282.

Feigl, H (1970) 'The 'Orthodox' View of Theories: Remarks in Defense as well as Critique', in Radner, M and Winokur, S (eds.), *Minnesota Studies in the Philosophy of Science*, vol IV, Minneapolis: University of Minnesota Press.

Fine, Kit (2001), 'The question of realism', *Philosophers' Imprint*, 1: 1–30.

Frenkel, Edward (2013), *Love and Math*, Basic Books.

Gale, David (1990), 'Proof as Explanation', *The Mathematical Intelligencer*, 12(1): 4.

Gale, David (1991), 'Mathematical Entertainments', *The Mathematical Intelligencer*, 13(1): 40–42.

Giaquinto, Marcus (2011) 'Crossing Curves: A Limit to the Use of Diagrams in Proofs', *Philosophia Mathematica*, 19(3), 281–307.

Gowers, Timothy and Nielsen, Michael (2009), 'Massively collaborative mathematics', *Nature*, 461(15): 879–881.

Grosholz, Emily (2000). 'The partial unification of domains, hybrids, and the growth of mathematical knowledge'. In Grosholz, E and Breger, H (Eds.), *The growth of mathematical knowledge* (81–91). Dordrecht: Kluwer Academic Publishers.

Hafner, Johannes, and Paolo Mancosu (2005) 'The Varieties of Mathematical Explanation'. In Mancosu, P., Jørgensen, K. F., Pedersen, S. A. (eds) *Visualization, Explanation and Reasoning Styles in Mathematics*, 215–50. Dordrecht: Springer.

Hafner, Johannes, and Paolo Mancosu (2008) 'Beyond Unification', *The Philosophy of Mathematical Practice*, Oxford University Press.

Hanna, Gila (1990). 'Some pedagogical aspects of proof', *Interchange* 21: 6–13.

Hoeltje, Miguel, Schnieder, Benjamin, and Steinberg, Alex (2013) 'Explanation by induction?', *Synthese*, 190(3): 509–524.

Hopkins, B., and R. Wilson (2004) 'The Truth about Königsberg', *College Mathematics Journal*, 35(3): 198–207.

Illari, Phyllis (2013). 'Mechanistic explanation: Integrating the ontic and epistemic'. *Erkenntnis*, 78: 237–255.

Kitcher, Philip (1975). Bolzano's ideal of algebraic analysis'. *Studies in History and Philosophy of Science* 6: 229–69.

Kitcher, Philip (1989) 'Explanatory unification and the causal structure of the world', in *Scientific explanation. Minnesota studies in the philosophy of science*, 13: 410–505.

Kitcher, Philip (1981) 'Explanatory Unification', *Philosophy of Science*, 48(4): 507–531.

Lange, Marc (2009) 'Why proofs by mathematical induction are generally not explanatory', *Analysis*, 69(2): 203–211.

Lange, Marc (2010) 'What Are Mathematical Coincidences (and Why Does It Matter)?', *Mind*, 119(474): 307–340.

Lange, Marc (2013) 'What Makes a Scientific Explanation Distinctively Mathematical?', *The British Journal for the Philosophy of Science* 64(3): 485–511.

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

Lange, Marc (2017) *Because without Cause: Non-Causal Explanations in Science and Mathematics*, New York: Oxford University Press.

Lange, Marc (ms) 'Ground and Explanation in Mathematics', unpublished manuscript.

Leuders, Timo (2016) 'Subject Matter Analysis with a Perspective on Teacher Education – The Case of Galois Theory as a Theory of Symmetry', *Journal für Mathematik-Didaktik*, Suppl 1: 163–191.

Miller, Steven and Montague, David (2012). ‘Picturing irrationality’, *Mathematics Magazine*, 85(2): 110–114, 2012.

Minio, Roberto (1984). ‘Interview with Michael Atiyah’. *The Mathematical Intelligencer*, 6(1): 9–19.

Pak, Igor (2006), ‘Partition bijections, a survey’, *The Ramanujan Journal*, 12(1), 5–75.

Palla, Marina, Potari, Despina and Spyrou, Panagiotis (2012) ‘Secondary school students’ understanding of mathematical induction: Structural characteristics and the process of proof construction’, *International Journal of Science and Mathematics Education*, 10(5), 1023-1045.

Pesic, Peter (2003), *Abel’s Proof: An Essay on the Sources and Meaning of Mathematical Unsolvability*, The MIT Press.

Pincock, Christopher (2015) ‘The Unsolvability of the Quintic: A Case Study of Abstract Mathematical Explanation’, *Philosophical Quarterly*, 15(3): 1–19.

Raman, Manya (2002) *Proof and Justification in Collegiate Calculus*, PhD Thesis, University of California, Berkeley.

Räz, Tim (2018), ‘Euler’s Königsberg: the explanatory power of mathematics’, *European Journal for Philosophy of Science*, 8(3): 331-346.

Resnik, Michael and Kushner David (1987). ‘Explanation, independence and realism in mathematics’. *British Journal for the Philosophy of Science*, 38: 141–158.

Reutlinger, Alexander. 2013. *A Theory of Causation in the Biological and Social Sciences.* New York: Palgrave Macmillan.

Reutlinger, Alexander (2016) ‘Is There A Monist Theory of Causal and Noncausal Explanations? The Counterfactual Theory of Scientific Explanation’, *Philosophy of Science*, 83: 733–745.

Rotman, Joseph (1994) *An Introduction to the Theory of Groups* (Fourth edition), New York: Springer Verlag.

Saatsi, Juha (2016) 'On the 'Indispensable Explanatory Role' of Mathematics', *Mind*, 125(500): 1045–1070.

Salmon, Wesley (1984). 'Scientific explanation: Three basic conceptions'. *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 2: 293–305.

Salverda, Josephine (2017) 'Characterizing properties and explanation in mathematics', *Logique et Analyse*, 237: 25-46.

Salverda, Josephine (2018) 'Can proofs by mathematical induction be explanatory?', in Piazza, Mario and Pulcini, Gabriele (Eds.) *Truth, Existence and Explanation*, Springer: Boston Studies in the History and Philosophy of Science.

Sandborg, David (1997) *Explanation and Mathematical Practice*, PhD thesis, University of Pittsburgh.

Sandborg, David (1998) 'Mathematical Explanation and the Theory of Why-Questions', *Brit. J. Phil. Sci.*, 49: 603–624/

Siegmund-Schultze, Reinhard (2014) 'Euclid's Proof of the Infinitude of Primes: Distorted, Clarified, Made Obsolete, and Confirmed in Modern Mathematics', *The Mathematical Intelligencer*, 36(4): 87-97.

Singh, Simon (1997) *Fermat's Last Theorem*, HarperPress.

Steiner, Mark (1978) 'Mathematical explanation', *Philosophical Studies*, 34, 135–151.

Steiner, Mark (1978b) 'Mathematics, explanation and scientific knowledge', *Nous*, 12: 17–28.

Stewart, Ian (2008), *Why Beauty is Truth: A History of Symmetry*, Basic Books.

Stewart, Ian (2015), *Galois Theory* (Fourth Edition), CRC Press, Taylor & Francis Group.

Sylvester, J. J. (1882), with insertions by F. Franklin, 'A Constructive Theory of Partitions, Arranged in Three Acts, an Interact and an Exodion', *American Journal of Mathematics*, 5(1): 251–330.

Thompson, Naomi (2016) 'Grounding and Metaphysical Explanation'. *Proceedings of the Aristotelian Society* 116(3): 395–402.

Van Fraassen, Bas (1980) *The Scientific Image*, New York: Oxford University Press.

Wagon, Stan (1987). 'Fourteen proofs of a result about tiling a rectangle'. *The American Mathematical Monthly*, 94(7): 601–617.

Weber, Erik and Verhoeven, Liza (2002). 'Explanatory proofs in mathematics'. *Logique et Analyse*, 179-180: 299–307.

Woodward, James (2003) *Making Things Happen: A Theory of Causal Explanation*, New York: Oxford University Press.

Zelcer, Mark (2013) 'Against Mathematical Explanation', *Journal for General Philosophy of Science*, 44(1): 173–192.