A Noetic Account of Explanation in Mathematics

William D'Alessandro and Ellen Lehet

Abstract

We defend a noetic account of intramathematical explanation. On this view, a piece of mathematics is explanatory just in case it produces understanding of an appropriate type. We motivate the view by presenting some appealing features of noeticism. We then discuss and criticize the most prominent extant version of noeticism, due to Inglis and Mejía Ramos, which identifies explanatory understanding with the possession of well-organized cognitive schemas. Finally, we present a novel noetic account. On our view, explanatory understanding arises from meeting specific explanatory objectives. We defend a cluster-concept account of explanatory objectives and identify four important subfamilies within the relevant network of resemblance relations. The resulting view is objectivist (in the sense that it takes explanatory success to be a matter of observer-independent fact), broader in scope than why-question-based accounts, compatible with empirical findings on experts' explanatory judgments, and capable of generalizing (with appropriate provisos) to scientific explanation as a whole. It thus fulfills Friedman's half-century-old demand for a general and objectivist theory which accounts for the link between explanation and understanding.

Keywords: explanation, understanding, noeticism, mathematical practice, mathematical explanation, proof

1 Introduction

It's become increasingly clear in recent years that the search for explanations is a important component of mathematical practice.[1] This realization has given rise to a puzzle: What could mathematical explanation possibly be?

The non-causal nature of such explanations immediately rules out many popular views, including the manipulationist and mechanist frameworks ascendant in much contemporary philosophy of science. Of the remaining accounts defended in the literature, most are either plagued by serious problems or limited to describing a special subclass of all mathematical explanations (or both).[2] This is an unsatisfying state of affairs. Is it possible to do better?

We think so. In this paper, we defend a *noeticist* account of mathematical explanation. Our view is a general theory that avoids problems with existing accounts, is independently plausible, and has many appealing features.

We begin in this section with motivation for noeticism. In §2, we discuss and criticize [Inglis & Mejía Ramos 2021], the most prominent noeticist theory currently on the market. We present our own account in §3. Unlike Inglis and Mejía Ramos, we deny that amassing or deploying a great deal of objectual understanding is sufficient for possessing explanatory understanding. We hold instead

[1] Recent overviews on explanation in mathematics include [Mancosu 2008], [D'Alessandro 2019] and [Pincock 2023]. The most important book-length discussion is [Lange 2016].

[2] See [D'Alessandro 2019] for an overview of the then-leading candidates. More recently, a counterfactual approach has been explored ([Baron et al. 2020]); see [Kasirzadeh 2021] and [Lange 2022] for what we take to be deep problems with such views.

that explanatory understanding is gained by meeting epistemic goals of certain types, which we call "explanatory objectives".

§3 lays out various aspects of this theory. Notably, we propose a cluster-concept account of explanatory objectives. On this view, explanation-directedness is a matter of standing in a network of family resemblance relations rather than satisfying a single set of necessary and sufficient conditions. We identify four subfamilies within the network in question, characterized by relatively strong internal affinities: these are Seeing the Big Picture (aimed at linking subject matters and achieving unification), Getting a Handle (aimed at delivering intuition, insight and cognitive fluency), Putting Things in Order (aimed at identifying dependence relations), and Accounting for Surprises (aimed at rationalizing striking or unexpected facts). §3 also addresses questions about objectivity, why-questions, empirical findings on mathematicians' explanatory judgments, and the unity of scientific and mathematical explanation. We conclude in §4.

1.1 A noeticist primer

According to the noetic theory of explanation, to possess an explanation of some fact is to be in an appropriate state of understanding with respect to that fact.[3] Call this state *explanatory understanding*. (The noeticist's biggest job is to say what this state is; §3 below presents our view.[4]) We think noeticism about mathematical explanation has many initially attractive features. Since the view is less often discussed and less well understood than many of its rivals, it's worth mentioning a few of these at the outset.

Our first point is a simple one. Virtually everyone agrees that explanation and understanding are closely related, and it's widely recognized that a theory of explanation should be able to account for this link. These claims have been doctrine in good standing for at least some fifty years, since Friedman asked "that a theory of scientific explanation tell us what it is about the explanation relation that produces understanding" ([Friedman 1974], 6). Noeticism does so in the most direct possible way. By contrast, other views often have a hard time showing how explanations in their sense are supposed to play the roles they do in our cognitive and epistemic lives.[5]

Second, noeticism can explain its rivals' successes. Alternative theories of explanation work as well as they do because they've correctly identified various possible sources of explanatory understanding. So

[3]We follow, e.g., [Dellsén 2021] in using 'noetic' to describe an understanding-based theory.

[4]Noeticists can identify an explanation itself with a body of information that could, typically would, or actually does produce explanatory understanding.

[5]As [Friedman 1974] argues, focusing on Hempel's D-N account, Bridgman's reduction-of-unfamiliarity account and Toulmin's historicist approach. More recently, opponents of the "ontic conception of explanation" championed by Salmon and his followers have questioned how explanations in the ontic sense (conceived of as mind-independent parts of the external world) could be thought to promote understanding and other epistemic goods ([Wright 2015], [Sheredos 2016], [Bokulich 2018]).

it's no surprise that these theories often make true predictions. While non-noeticist views go wrong in mistaking a *common route* to explanation for a *necessary and sufficient condition*, they contain genuine insights which the noeticist can embrace with open arms. Rival theories of explanation have more trouble here—it's not obvious what a counterfactual theorist should say, for example, about the appeal of unificationism. No doubt some story can be told, but it's likely to be relatively complex, and in the end the proponent of one theory may have to dismiss much of her opponents' views as confusion or error.

Third, noeticism has no trouble handling different types of explanantia. It gives a unified account of theorems, proofs, diagrams and other mathematicalia: any type of information that confers the appropriate sort of understanding is fit to serve as an explanation, and all count as explanatory for the same reason.

Alternative theories often struggle on this front. Some, for instance, focus exclusively on explanatory proof, and what they say about proofs is hard to generalize to other cases. (The views of [Steiner 1978] and [Lange 2014] are in this boat.) Meanwhile, the counterfactual approach to mathematical explanation ([Baron et al. 2020]) only applies straightforwardly to theorems explained by other theorems. This is a bad sign for these views. As [D'Alessandro 2017] argues, mathematical practice recognizes several types of object and content as potentially explanatory. It's implausible that each of these types corresponds to a fundamentally different kind of explanation, following its own *sui generis* rules. A theory that suggests such a disunified picture—or worse, one that has to deny the existence of explanations embraced by working mathematics—is a theory deserving of skepticism.

So noeticism has some appealing features. But appealing features can be outweighed by damning flaws. Fortunately, noeticism is short on the latter. Indeed, some of the most common objections can be easily seen to fail when the view is understood properly.

For instance, some have complained that our "sense of understanding" is nothing more than an unreliable subjective feeling ([Trout 2002]). This is irrelevant even if true. The noeticist isn't primarily concerned with the sense of understanding, but rather with understanding itself—a substantive cognitive state that manifests in observable traits and abilities.

Relatedly, it's sometimes thought that if noeticism were true, we'd be unable to say anything interestingly general about what sorts of things are explanatory and why (because different people will respond differently to the same information, and some might gain understanding where others wouldn't). But this worry is overstated. The cognitive similarities between people are much more extensive than the differences. In spite of individual variations in memory, processing power, reasoning style and so on, most of us are built on the same cognitive template, and so most of us will tend to gain understanding from the same sources and for the same reasons. Philosophers and psychologists will have plenty of interesting work to do in characterizing these patterns.

Finally, it's sometimes insinuated that "psychologistic" answers to philosophical questions are somehow inherently bad. ("It is as though to classify a theory as psychologistic is enough automatically to disprove it, as though everyone already understands and agrees that psychologism is a defect or mistake, and that we should all be grateful for its having been thoroughly exploded in a heroic past age of philosophy" ([Jacquette 2003], 4).) This ideology seems to be inspired in part by the power of Frege's arguments against psychologism in logic and mathematics. However compelling those particular arguments might be, though, there's nothing intrinsically unacceptable about psychologistic theories in general.

Consider the case of Hume. Past authors had assumed that the problem of causation was that of elucidating the special worldly connection between cause and effect. Hume proposes, instead, that "a cause is an object precedent and contiguous to another, and so united with it, that the idea of the one determines the mind to form the idea of the other, and the impression of the one to form a more lively idea of the other" (*Treatise* 1.3.14.31).[6] Causation, in other words, isn't just an objective relation between things in the world, but also involves the tendency of those things to produce certain mental states in us. Whatever one thinks of Hume's psychologistic account, it's at least a serious and intelligible theory that might have been true rather than a simplistic conceptual error.

On the whole, then, noeticism about explanation has a lot going for it. A view with so many virtues and few obvious vices deserves a careful look.

2 Inglis and Mejía Ramos's account

We aren't the first to advocate an understanding-based view of explanation in mathematics. Adopting the terminology of [Wilkenfeld 2014], Matthew Inglis and Juan Pablo Mejía Ramos—hereafter IMR—defend what they call a *functional theory* of explanation ([Inglis & Mejía Ramos 2021]). On this view, what makes something an explanation is that it functions as a source of understanding. IMR's account is therefore a version of noeticism. (We prefer 'noeticism' over 'functionalism' because the latter gives no hint about the nature of the relevant function.) In this section, we discuss IMR's theory and explain where we think it goes wrong. §3.2.7 below comments on [Frans & Van Kerkhove 2023]'s alternative version of noeticism.

A basic component of IMR's account is the concept of a *schema*, introduced by Piaget and still widely used in cognitive and educational psychology. A schema is, roughly speaking, a structured mental representation of a subject matter S that facilitates remembering, recognizing, predicting and reasoning about information related to S.

Schemas are thought to play a role in many cognitive tasks. For instance, you have a *house* schema,

[6] Whether or not Hume really endorsed this psychologistic view, and how it relates to his first, seemingly non-psychologistic definition of causation, is a topic of much debate. It's enough for our purposes that many Hume scholars accept the psychologistic reading.

which encapsulates your understanding of the important general attributes of houses (e.g. that they have parts, are made of materials, and have functions) as well as some default specific values for those attributes (e.g. that the parts of a house are rooms, that the materials are wood or brick, and that the function is human habitation). Accessing this schema can provide various benefits. Since its components are strongly cognitively interlinked, the schema allows you to quickly classify something as a house on the basis of a few key characteristics. And once classified, the schema helps you recognize other expected house-features, make inferences about features you didn't directly perceive, and remember what you saw later.

We also have schemas for more abstract subjects, like chess games and mathematical problems, which play cognitive roles similar to schemas for everyday objects. A skilled chess player may have a detailed *Sicilian Defense* schema, for instance, encoding the situations in which this opening is likely to be played, its most important variations and the counterplays for each, typical middle-game positions to which the Sicilian tends to lead, and so on. A player with many such schemas at her disposal will be able to choose moves quickly and accurately. A player with none will have to painstakingly note the positions of each piece on the board, call to mind their possible moves and calculate the consequences of each, leading to slower and more error-prone performance.

On IMR's view, the creation and consolidation of schemas is constitutive of mathematical understanding. As they write, "one can be said to have understood something when a sufficiently well-organised schema... has been encoded into long-term memory" (S6381). For noeticists like IMR, the task of a theory of explanatory proof is therefore to say what kinds of proofs allow us to generate and store high-quality mathematical schemas.

We claim on the contrary that possessing well-organized schemas is neither necessary or sufficient for explanatory understanding. IMR's proposal is therefore a step in the wrong direction.

Consider the necessity claim first. Suppose that T is a theorem belonging to subject $\mathcal{S}$. Suppose also that we want to understand why T is true, and hence to explain T's truth. We claim that one can often do this without having an extensive, detailed or well-structured $\mathcal{S}$-schema.

Example: Consider Lange's well-known "calculator number" case. A calculator number is a six-digit number of the form *abccba*, where the digits *abc* are the numbers forming some row, column or main diagonal of a square 1-9 calculator keyboard. There are sixteen calculator numbers in total, and it turns out that each one is divisible by 37. Why does this curious fact hold? Lange offers the following explanatory proof ([Lange 2014], 488; originally from [Nummela 1987]):

The three digits from which a calculator number is formed are three integers a, $a+d$, $a+2d$ in arithmetic progression. Take any number formed from three such integers in the manner of a calculator number—that is, any number of the form $10^5 a + 10^4(a+d) + 10^3(a+2d) + 10^2(a+2d) + 10(a+d) + a$.

Regrouping, we find this equal to

$$a\left(10^5 + 10^4 + 10^3 + 10^2 + 10 + 1\right) + d\left(10^4 + 2 \cdot 10^3 + 2 \cdot 10^2 + 10\right)$$
$$= 1221\,(91a + 10d)$$
$$= (3 \cdot 11 \cdot 37)\,(91a + 10d)\,.$$

We agree that this proof explains why calculator numbers are divisible by 37. But it's not clear that the proof either provides or exploits any suitable well-organized schema. The list of concepts, facts and techniques in play is quite small—the notion of an arithmetic progression, multiplication and addition, factoring, and so on. And the argument does nothing very profound or ingenious with these tools. So the average reader is unlikely to upgrade her basic arithmetical and algebraic schemas as a result of working through the proof.

One could try to claim that the proof confers understanding by contributing to a more specialized schema, say *calculator number*. But there's a risk of ad hocness here. Given any proof whatsoever, one can typically concoct some highly problem-specific schema to which the proof can be said to contribute. We need a stronger condition to distinguish explanatory proofs from the rest.

Even setting aside this worry, it's doubtful that the above strategy could work. Does Lange's proof really leave us with a well-organized *calculator number* schema? Surely not. The only interesting fact about calculator numbers appearing in the proof is that the digits *abc* must be in arithmetic progression. Perhaps it's hard to say in general what counts as a well-organized schema, but this isn't a borderline case. Someone who knows basic definitions and the facts used in the proof may still be almost completely clueless about calculator numbers. Such a person will have only the most threadbare schemas. And yet the proof is explanatory (for them).

Could an appeal to *degrees* of explanatoriness help IMR here?[7] The idea would be that Lange's proof, while not highly explanatory in absolute terms, is at least *somewhat* explanatory, and in particular is *more* explanatory than salient alternatives (such as the brute-force proof which merely enumerates the sixteen calculator numbers and checks that 37 divides each one). Perhaps these comparative facts are enough to make sense of our judgment that Lange's proof explains the result.

We acknowledge that IMR's account has the resources to distinguish more and less explanatory proofs in this way. And we're willing to grant that Lange's proof does somewhat better by IMR's criteria than the brute-force proof. Still, we don't think these concessions are enough to vindicate IMR, at least on what we take to be the most natural interpretation of their view.

Presumably IMR want to say that the explanatoriness of a proof is proportional to its schema-enhancing power: a highly explanatory proof will enrich our schemas extensively, while a proof that offers

[7]We thank an anonymous referee for raising this issue.

minor additions or upgrades will count as only marginally explanatory at best. Indeed, on their view, if a proof "did develop a reader's objectual understanding, but only by some trivial amount, then it is unlikely that it would exceed the increased-understanding threshold for explanatoriness. In such a scenario, we could coherently describe the proof as being non-explanatory" (S6386).

It's clear, we think, that Lange's calculator-number proof doesn't profoundly add to anyone's objectual understanding. Rather, it contributes just a few basic facts and ideas. Considered on its own, then, it should either count as unexplanatory (if it fails to clear the relevant contextual threshold) or at most only weakly explanatory for IMR. But neither of these verdicts seem right. The proof, we take it, is perfectly successful as an explanation, and not just a little better than a very poor alternative.[8] So the appeal to degrees of explanatoriness doesn't obviously go in IMR's favor here.

We conclude that it's possible to possess understanding and grasp an explanation without having anything resembling a high-quality schema. Our next task is to show that the reverse implication fails too.

Our basic objection is that we often know quite a lot about a subject or problem without grasping why some of the associated facts are true. In such cases, it seems we have a well-organized schema without possessing explanatory understanding.

To take a well-known example, consider the four color theorem. Planar graphs and graph colorings have been major areas of study for well over a century, and graph theorists have amassed a great deal of detailed, sophisticated and intricately structured knowledge about these subjects. Appel and Haken's computer-assisted 1976 proof of the four-color theorem makes masterful use of the techniques developed to date. (As they themselves note: "Over the past 100 years, a number of authors including A. B. Kempe, G. D. Birkhoff, and H. Heesch have developed a theory of reducibility to attack the problem. Simultaneously, a theory of unavoidable sets has been developed and the fusion of these has led to the proof" ([Appel & Haken 1976], 711). The body of knowledge embedded in Appel and Haken's work surely amounts to a well-organized schema if anything does.

But the proof of the four color theorem is also paradigmatically unexplanatory. As the graph theorist Paul Seymour writes, "We would very much like to know the 'real' reason the [four color theorem] is true; what exactly is it about planarity that implies that four colours suffice? Its statement is so simple and appealing that the massive case analysis of the computer proof surely cannot be the book proof" ([Seymour 2016], 417).[9] Similarly, according to Cristopher Moore and Stephan Mertens, the Appel-Haken proof leaves us with "an unsatisfying state of affairs. The best proofs do not simply certify that something is true—they illuminate it, and explain *why* it must be true. While the Four Color theorem may be

[8]To be sure, the explanatory inquiry involved in the calculator number case is thoroughly shallow, trifling and a mathematical dead end; this is a quintessential piece of idle recreational math and doesn't pretend to be anything else. But it's important to distinguish a *good explanation of a trivial fact* from a *bad explanation*. We discuss issues of explanatory value further in §3.2.3.

[9]"The book" is God's book, which contains the best possible proof of every theorem. (The idea is Paul Erdős's.)

proved, it is not well understood" ([Moore & Mertens 2011], 125; emphasis in original).

So we seem to have a proof that turns on various well-organized schema, and which synthesizes those schemas to impressive effect, but which nevertheless fails to explain or produce understanding.

In response to this sort of case, IMR might want to say that, although the Appel-Haken proof *draws upon* high-quality schemas, it doesn't substantially *add to* or *improve* those schemas. Indeed, IMR sometimes seem to claim that a proof never counts as explanatory unless it contributes something new to the reader's cognitive resources. For instance, they criticize a certain digit-counting proof in the following way:

> Can we account for why [the proof] is seen as non-explanatory...? The answer... seems clear. [The proof] is a simple counting argument. For the vast majority of readers it will activate only existing schemas that are exceptionally well understood (those concerned with natural numbers represented in base 10). Since most readers who encounter this proof will already have extremely well-developed schemas concerning these matters, there is little room for the proof to develop new understanding. Instead it is a simple matter of applying existing understanding to a particular context. (S6385)

But this suggestion is hard to swallow. IMR are saying that a proof can only be explanatory for a given person if it generates understanding which that person previously lacked. This seems to entail, counterintuitively, that proofs lose their explanatory value after they're grasped for the first time. It also entails that someone with maximal understanding of a subject couldn't possibly find any proofs on that subject explanatory. Our experiences contradict the first of these implications and cast serious doubt on the second: it seems clear that a proof whose ins and outs I've thoroughly mastered can nevertheless continue to be explanatory for me, in virtue of my appreciation of its present personal cognitive value.[10]

IMR would do better, then, to count a proof as explanatory if it makes appropriate use of one's existing schemas, even if it doesn't add to or improve them. Doing so would make their view more plausible in general. But it also prevents them from dismissing the Appel-Haken proof as unexplanatory simply because it fails to enhance the typical reader's schemas. (We set aside the factual question of whether the proof actually does fail in this way, which seems far from straightforward.)

To summarize, the moral of this example is that proofs can make essential use of well-organized schemas without explaining their results or conferring understanding. Although IMR suggest that explanatory proofs must always *add to* one's schemas—a potential defense against our counterexample—this view is implausible. We conclude that possessing or using high-quality schemas isn't sufficient for understanding.

[10]For our own treatment of this issue, see §3.2.4, especially footnote 14 and the surrounding discussion.

This discussion has been somewhat involved, so we'd like to end with a broad-brush account of our perspective. IMR claim, to a first approximation, that understanding or explaining why P is nothing more than having a body of well-structured knowledge relevant to P from which P can be inferred. In other words, IMR believe that "the distinction between... objectual understanding and understanding-why/explanatory understanding... is a matter of degree rather than of kind" (S6377).

We disagree. Explanatory understanding isn't merely a threshold on the continuum of objectual understanding. As the above examples show, one can have good explanatory understanding while possessing virtually no objectual understanding (as in the calculator number case). Conversely, one can have abundant objectual understanding while totally lacking explanatory understanding (as in the four color theorem case).

Experience has shown that these claims are liable to misinterpretation, so let us try to make our position clear. We don't deny that explanatory understanding normally tends to increase along with objectual understanding. In general, the better one's $\mathcal{S}$-schemas, the more relevant information one has, and so the more $\mathcal{S}$-related explanations one may be in a position to grasp. We also agree, of course, that someone with zero objectual understanding (i.e., no relevant information at all) necessarily has zero explanatory understanding. And perhaps someone with "maximal objectual understanding" would necessarily also have maximal explanatory understanding. Our claim is just that explanatory and objectual understanding are different in kind, and possessing a great deal of the latter is neither necessary nor sufficient for possessing the former.

An analogy: according to the *wealth theory of automobile ownership*, being sufficiently rich is just what it is to own a car. The wealth theory is obviously false. But one can reject it while acknowledging that wealth and car ownership are correlated, that one needs at least a little money to obtain a car, that extremely wealthy people almost always have some cars, and so on. The point is that modest (or even fairly strong) correlation isn't sufficient for dependence, identity or other metaphysically heavyweight relations.

We conclude that focusing exclusively on objectual understanding—that is, on the quality of an inquirer's schemas—is a false start for a noeticist theory of explanation.

3 A different style of noeticism

As noeticists, we agree with IMR and others that a mathematical explanation is nothing but a piece of mathematics that provides an appropriate type of understanding. We think existing accounts haven't yet succeeded at identifying this type of understanding. We try to do so below.

3.1 Motivating the view

As our discussion of IMR has shown, possessing information relevant to P—even abundant, well-structured and cognitively tractable information—is in general neither necessary nor sufficient for explaining P. So the first task for an adequate noeticist theory is to avoid equating explanatory understanding with a mere accumulation of knowledge.

The key insight needed to solve this problem, we take it, is as follows. As seekers of explanatory understanding, we typically aim to resolve specific confusions, clear up specific mysteries, uncover specific sorts of reasons-why, comprehend things from specific viewpoints. In a word, we aim to meet some *explanatory objective* or other. A proof is therefore explanatory when, and only when, it's capable of meeting a relevant objective. Conferring or drawing upon a vast store of knowledge is no use if the knowledge in question doesn't help achieve our explanatory goals.

This picture makes sense of the cases discussed above. Consider the calculator number phenomenon. We were initially puzzled by the fact that all calculator numbers are divisible by 37; our objective was to understand why this occurs. A proof would count as explanatory relative to our objective just in case it intelligibly presents a reason that accounts for the divisibility fact. The proof in question did so by observing that calculator numbers have the form $10^5a + 10^4(a+d) + 10^3(a+2d) + 10^2(a+2d) + 10(a+d) + a$, which simplifies to $(3 \cdot 11 \cdot 37)(91a + 10d)$. Meeting our explanatory objective in this case didn't require deep or extensive knowledge about calculator numbers, but just a couple of elementary facts of the right kind.

In other situations, however, our objectives may be quite different, and a suitable explanation will have to meet a distinct set of needs. For instance, many areas of contemporary mathematics aim to transfer problem-solving resources between domains, or to unify seemingly disparate problems within a common abstract framework. A number theorist might seek a specifically geometric understanding of why the Riemann hypothesis is true. Or it might occur to a category theorist to wonder whether dissimilar-looking results in linear algebra, set theory and group theory have a common explanation in terms of adjoints and limits. In cases like these, developing rich and extensive schemas will be crucial to obtaining explanatory understanding. But again, however, not just any problem-relevant knowledge will do. All the analytic number theory in the world won't help the number theorist achieve her goal, even if it leads to a perfectly good proof of the Riemann hypothesis. And mastering linear algebra, set theory and group theory individually won't deliver the category theorist's unifying perspective. Whether or not a given piece of mathematics counts as explanatory, then, depends sensitively on the type of explanation one is after.

This conception of explanation has some affinities with erotetic accounts like Lange's, which identify mathematical explanations with answers to antecedently specified why-questions. We find Lange's work insightful, but our view departs from his in important ways, as we'll discuss below.

3.2 Explanatory objectives

3.2.1 Preliminaries

We've suggested that a piece of mathematics confers explanatory understanding when it meets an appropriate explanatory objective. An explanatory objective, in turn, is a demand for insight of a certain sort—insight into why a theorem is true, what a phenomenon means, how a result should be interpreted, or how one piece of mathematics relates to others, for instance.

The literature on mathematical explanation is replete with general theories as well as accounts of specific types of explanation. As we've suggested, the noeticist can accommodate these views by treating them as descriptions of possible sources of understanding—that is, as possible explanatory objectives. We adopt this viewpoint here. Hewing closely to notions previously discussed in the literature, then, here we give a partial list of explanatory objectives:

- Identifying a difference-making feature without which a result wouldn't hold (i.e., mechanistic explanation in the sense of [Frans & Weber 2014], or counterfactual-based explanation in the sense of [Baron et al. 2020])
- Identifying an object or structure upon which a mathematical phenomenon metaphysically depends (i.e., abstract explanation, in the sense of [Pincock 2015])
- Deriving a result using tools from another part of mathematics, thereby revealing shared structure or intertheoretic connections (i.e., impure explanation, in the sense of [Ryan 2023])
- Deriving a striking feature of a theorem from a feature of the same type in the problem setup (i.e., salience-based explanation, in the sense of [Lange 2014])
- Finding a proof that clearly presents a compelling reason for a theorem's truth (i.e., transparent explanation, in the sense of [D'Alessandro 2021])
- Finding a proof that "shows what a theorem means" by linking it with other results in a broader theoretical context (i.e., deep explanation, in the sense of [D'Alessandro 2021])
- Subsuming a diverse collection of facts under a single general framework (i.e., unifying explanation, in the sense of [Kitcher 1989] or [Lehet 2021a])
- Showing how to obtain a family of proofs of related theorems by varying a characteristic property of a relevant object (i.e., deformability-based explanation, in the sense of [Steiner 1978])

We think this list captures some especially noteworthy types of explanatory objectives, although we don't presume it to be exhaustive. One could debate the distinctness, independence, or relative importance of the items on the list, or even whether all of them ought to count as genuine routes to explanation. These are legitimate questions, but for simplicity we take the list at face value here.

The remaining subsections below will lay out several aspects of the theory of explanatory objectives in more detail. For now we'll address a few general preliminaries.

First, on our view, it's a matter of determinate fact whether or not a purported explanation meets a given explanatory objective. So it's possible for inquirers to mistakenly fail to recognize correct explanations, or to mistakenly accept incorrect ones. Understanding is only gained when one grasps a correct explanation that actually meets one's objectives. (What's more, one has to recognize *that* the explanation is correct and that it meets the relevant objectives. If I seek a unifying explanatory proof of theorem T, and in fact I know such a proof but somehow fail to notice its unifying character, I haven't yet properly fulfilled my objective or fully grasped T's explanation.)

Second, explanatory objectives may be more or less specific. One could, for instance, be seeking *just any* impure explanation of a given fact, or one might demand a topological explanation in particular. (The first objective would be appropriate for an inquirer wishing to understand a result from an illuminating alternative viewpoint, without knowing which viewpoint this might turn out to be.) Complex objectives are also fair game: one might want a unifying explanation that's also highly transparent, say. Of course, the more requirements one imposes, the harder it will be to find a suitable explanation, and the greater the chance that no such explanation exists (more on this in §3.2.3 below).

Third, we need to account for the fact that we commonly seek, evaluate and accept explanations without specifying any explanatory objective. This might be taken to suggest that there exist explanations *simpliciter* which answer to no objective in particular. Our diagnosis is different. Just as one can have a complex conjunctive objective, disjunctive objectives are possible, and the limiting case is the maximally disjunctive objective which simply seeks understanding of some form or other. We suggest that this is our default objective much of the time. Unless we have specific reasons to favor a narrower goal—because we're novices struggling for intuitive clarity, or experienced researchers in search of a specific explanatory vantage point—we're often happy to take any kind of insight we can get. In this standard situation, it's harmless to think and talk about explanation in an unqualified way.[11]

On the other hand, it's worth noting that mathematicians often *do* specify particular explanatory objectives. They might state a preference for a deep proof of a particular theorem over an elementary transparent proof, for instance ([D'Alessandro 2021]). Or they might repudiate impure proofs as epistemically unsatisfying, as Aristotle, Bolzano, Frege and others did ([Detlefsen & Arana 2011]). So our notion of explanatory objectives isn't a foreign imposition on mathematical practice. On the contrary, we need

[11]We hypothesize that inquirers in this open-minded state are often especially attuned to easily recognizable, general-purpose kinds of explanatory virtues, such as transparency. See §3.2.6 for further discussion.

some such notion to make sense of the specificity and diversity of mathematicians' stated explanatory aims.

Fourth, our account can accommodate intermediate degrees of explanation. Degree structures can manifest in various ways, depending on the kinds of objectives at issue. For instance, some objectives come with built-in degree metrics of an obvious sort—a proof can be highly transparent or only moderately so, and the more transparent, the more explanatory (relative to that objective). Alternatively, degrees can measure the extent to which an explanation succeeds at meeting multiple objectives concurrently. A proof that's impure, abstract, deep and unifying all at once will have greater explanatory value (for someone concerned with those objectives) than a proof that's merely impure.

Finally, Inglis and Mejía Ramos—our noeticist foils—give a detailed account of the psychological phenomena which they take to be involved in understanding. Can we do the same? As appealing as we find this goal, it's unlikely, given our view, that there's any single characterization of the relevant mental processes. Our version of noeticism countenances a variety of explanatory objectives, each of which calls for its own type of information with its own epistemic and cognitive roles to play. We can't point to a simple psychological equation (like "better understanding = better schemas") that fully captures this diversity.

3.2.2 What makes an objective explanatory?

Mathematicians have many epistemic aims. They seek to gain knowledge, to make fruitful conjectures, to gather evidence, to develop reliable heuristics, and so on. Many of these desiderata don't involve explanatory understanding. What, then, distinguishes explanatory objectives from epistemic interests of other kinds?

On our view, *explanatory objective* is a cluster concept. Thus one can't point to a single property that all explanatory objectives have in common. Instead, there's a set of features which such objectives tend to share; all explanatory objectives will possess at least a few features from the list, with varying degrees of overlap between one objective and another.

This picture is similar to Collin Rice and Yasha Rohwer's cluster-concept account of scientific explanation ([Rice & Rohwer 2021]). Rice and Rohwer are motivated by two observations. First, there seems to be no single theory that fully captures the diversity of types of explanation while avoiding counterexamples. Second, the pluralist's response to this situation is unsatisfying: if we merely say that there exist many types of explanation $E_1, E_2, \ldots, E_n$ sharing nothing interesting in common, we're at a loss to explain why the same label applies to all the E_i. A cluster-concept theory can accommodate both insights.

We depart slightly from Rice and Rohwer, in that we defend monism about *mathematical explanation* but propose a cluster theory of *explanatory objectives.* That is, we think all explanations work in essentially the same way—by providing explanatory understanding, and hence by meeting explanatory objectives—but we hold that the variety of explanatory objectives are related to one another only by family resemblance. (Still, it's harmless enough to regard our view as a cluster-concept account of mathematical explanation, as long as one keeps in mind that the relevant cluster is located a couple of analytical steps away.)

The following table, inspired by Rice and Rohwer's similar chart (1037), displays some of the features shared by various explanatory objectives:

	Abstract	Deep	Transparent	Unifying	Impure	Counterfactual	Mechanistic	Salience-based (Lange)	Deformability-based (Steiner)
Rationalizes striking feature of a theorem								✓	✓
Cognitively compelling and tractable		✓				✓			
Links diverse subject areas		✓		✓					
Identifies dependence relations	✓				✓	✓			✓
Introduces higher level of abstraction	✓	✓		✓					
Aims at broad context and generalizability		✓		✓	✓				✓
Seeks a proof with specific formal features			✓				✓	✓	
Seeks a proof with specific contentual features	✓	✓		✓					

Table 1: Some types of explanatory objectives and their characteristic features.

As the table suggests, there's significant overlap between the features possessed by different objectives, but no single feature common to all of them. We can also discern several families of objectives sharing broadly similar explanatory goals. ✓

Impure, deep and unifying explanations are similar, for instance, in seeking links between theories and a bird's-eye perspective. We might call this family of objectives Seeing the Big Picture. (Steinerian explanation belongs here too, to a lesser degree, as it also requires a limited kind of generalizability at the level of proof.) Transparent and mechanistic explanations look for vivid, compelling reasons presented in a cognitively accessible way. Call this family Getting a Handle. Abstract, counterfactual, mechanistic and Steinerian explanations are all interested in dependence relations of various kinds, and thus aim to identify grounds or difference-making features. This family might be called Putting Things in Order. Finally, both Steiner's and Lange's styles of explanation focus on rationalizing striking properties of a theorem with the help of an appropriately matched proof. Call this family Accounting for Surprises. (Incidentally, this taxonomy helps explain why Steiner's theory has provoked both widespread interest and widespread criticism—the view is a peculiar multi-family hybrid incorporating objectives of several kinds.)

Each of these families represents what's plausibly a natural and important sort of explanatory understanding. Seeing the Big Picture reflects our interest in comprehending patterns and placing facts in context. Getting a Handle captures the drive to grasp phenomena with intuitive clarity. We understand what determines what by Putting Things in Order. And Accounting for Surprises eliminates the appearance of coincidence by showing that the observed facts were to be expected. As our view predicts, then, we gain explanatory understanding by meeting explanatory objectives.

This account also sheds light on why many proofs (and other pieces of mathematics) fail to explain. Supposing that the above taxonomy is reasonably complete, if a proof meets nothing like any of the above objectives and produces none of the associated forms of explanatory understanding, it's likely that the proof won't succeed as an explanation.

Consider again the Appel-Haken proof of the four color theorem, for instance. The proof uses only tools internal to graph theory and offers no broad or unifying perspective, so it doesn't help with Seeing the Big Picture. Being a massive brute-force case analysis, it's far from the clarity and simplicity associated with Getting a Handle. The proof offers no positive ground, reason or source for the truth of the theorem, but only shows that various potential counterexamples don't occur. So it brings us no closer to Putting Things in Order. Finally, there's no salient feature of the theorem whose Surprisingness the proof Accounts for—as evidenced by mathematicians' claims that the sufficiency of four colors remains mysterious.

No matter what one's particular explanatory aims might be, then, the Appel-Haken proof is likely to disappoint. This fact also accounts for mathematicians' tendency to judge the proof unexplanatory without reference to any specific objective. As we predicted above, unqualified judgments about (un)explanatoriness typically signal the (non)existence of at least one objective which the prospective explanans successfully meets. Incidentally, since many proofs by exhaustion share the relevant features of the Appel-Haken proof, this analysis also shows why such proofs are typically unexplanatory (cf. [Baker 2009], [Colyvan 2012]).

3.2.3 Objective value and the value of objectives

On our view, mathematical explanations are pieces of mathematics that meet explanatory objectives. There will, of course, typically be many possible objectives associated with a given explanandum, and we're free to choose which of these to adopt according to our knowledge, goals and interests.

Given this picture, it's tempting to infer that all objectives are created equal. But we don't subscribe to this brand of egalitarianism. On the contrary, there are many dimensions of value along which one explanatory objective can be better or worse than another.

For instance, some objectives are simply unsatisfiable in some contexts, in the sense that there is no explanation answering to the relevant demands. An objective that asks for an explanation of a false proposition falls into this category. So does an objective that requests a nonexistent type of explanation for a true proposition (say, a purely algebraic proof of the fundamental theorem of algebra), or any explanation at all for an unexplainable truth (say, the fact that $3 + 5 = 8$).

Among objectives that can actually be fulfilled, some are more natural, fruitful, or consequential than others. A successful geometric explanation of the Riemann hypothesis would throw open the floodgates of mathematical knowledge, leading to profound and far-ranging insights about the relationship between number theory and geometry. Explaining why calculator numbers are divisible by 37 is entertaining but comparatively pointless. Both objectives are legitimate, but their epistemic demands and benefits are completely unalike.

The virtues associated with different types of objective may trade off against one another. For instance, transparent explanations are clear, vivid, and cognitively tractable, but often "shallow", in that they use elementary methods which don't generalize easily. On the other hand, deep explanations are intellectually rich but often "opaque", in that they require complex and unintuitive theoretical machinery. So the goodness of an objective isn't a simple linear matter. An explanatory objective (or successful explanation) that's outstanding in one respect may be subpar in others.

Still, it may make sense to ask whether an objective A is more worthwhile than another objective B, all things considered—that is, whether A's package of merits and demerits is superior to B's, with the elements of each package appropriately weighted by importance. Mathematicians often make such judgments about overall pursuitworthiness, and reasonably so. It's also conceivable that some types of explanatory objective are more valuable than others in general. For instance, [Lehet 2021a] and [Ryan 2023] argue that contemporary mathematics has more use for methodologically impure explanations than for pure ones.

The fact that one can choose which objectives to pursue, then, doesn't mean that all objectives are on equal footing. Mathematical progress depends crucially on asking the right questions: simply accumulating answers isn't enough. The evaluation and careful selection of objectives is therefore an important part of explanatory practice.

3.2.4 Explanatory objectives and why-questions

Many explanatory objectives can be expressed as why-questions: "Why is every planar map four-colorable?" "Why is every calculator number divisible by 37?" Often enough, we're in a position to formulate these questions in advance of inquiry, and we know what kind of information would count as a satisfying answer. So it's no surprise that why-question theories like Lange's seem quite appealing.

Recall that, on Lange's view, a proof is explanatory when it "exploits a certain kind of feature in the problem: the same kind of feature that is outstanding in the result being explained" ([Lange 2014], 489). For instance, explaining a theorem which exhibits a striking symmetry calls for a proof that makes use of a corresponding symmetry in the problem setup. Thus it only makes sense to talk about explanatory proof when a why-question has been prompted by some specific noteworthy quality of a theorem. As Lange says, his account "predicts that if [a] 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).

We think Lange has identified a common and important type of explanatory objective: to answer questions of the form "Why does theorem T exhibit surprising feature F?" But this account isn't the whole story about mathematical explanation, or even about explanatory proof. We discuss several ways in which explanatory objectives can diverge from Lange's picture.

First, not all successful explanations start with a sharply posed question about a specific feature of a result or phenomenon. In some cases, our initial explanatory objective is relatively vague: we've discerned something mysterious, obscure, or suggestive about the observed facts, and we seek better understanding without knowing quite what form it will take. In other situations, an objective is revealed only during or after inquiry, as for example when a proof reveals an illuminating but unanticipated aspect of a phenomenon.

Both types of case are common. Gauss's quadratic reciprocity theorem is an example of the first (cf. [D'Alessandro 2021]). In the early days of the theorem's conjecture and proof, mathematicians recognized it as a striking statement that hinted at a mysterious relationship between pairs of prime numbers. Explaining the theorem was a primary goal of Gauss and his successors: Gauss called it his "golden theorem" and returned to it repeatedly, producing eight different proofs during his lifetime. But these early explorers could hardly have known just what they were looking for. Number theory was still in its infancy, and mathematicians lacked both the basic facts surrounding reciprocity and the tools needed to conceptualize and investigate them. Truly enlightening proofs would come only after more of this machinery had been developed. Yet Gauss was capable of pursuing an explanatory proof even without a precise idea of what such an explanation would amount to.[12] He recognized the theorem as profound, intuition-defying, and indicative of major gaps in understanding, and this was enough to arm him with an actionable explanatory objective.[13]

For an example of the second type, consider the unexpected change in perspective brought about by abstract algebra with respect to familiar number systems. Pre-modern mathematicians had never thought to ask, for instance, why Euclid's lemma is true—that is, why prime numbers p have the property that p divides a product ab only if p divides at least one of a or b alone. This seems at first glance like a basic property of the primes that requires no explanation. But the modern turn in algebra led to a new understanding of primality. According to the revised picture, it's in fact Euclid's lemma that captures

[12]It's also clear that, contra Lange, the reciprocity theorem exhibited no particular formal feature (such as symmetry or simplicity) that an explanatory proof was expected to reduplicate ([D'Alessandro 2021], 8659).

[13]See [Sandborg 1998] for a similar criticism of van Fraassen's why-question theory ([van Fraassen 1980]).

the correct general definition ($p \mid ab \Longrightarrow p \mid a$ or $p \mid b$), while the familiar condition ($n \mid p \Longrightarrow n = 1$ or $n = p$) represents an accidental feature called *irreducibility* that prime elements possess in some rings but not others. From this perspective, Euclid's lemma is a non-obvious fact which we might well hope to explain: Why do primality and irreducibility in $\mathbb{N}$ happen to coincide, when they come apart elsewhere? The answer is that the integers are a unique factorization domain, and Euclid's lemma holds in all such rings. And this is a perfectly good explanation, even if nobody thought to pose the why-question until after the answer was available.

Lange's account seems to rule out the possibility of both types of case, since it claims that one can only seek or give an explanatory proof after one has formulated a why-question about some particular salient feature of the theorem to be explained. This prediction is incorrect. By contrast, our account allows for such cases. Some explanatory objectives are broad and open-ended. And we sometimes add a given objective to our list only after seeing the associated explanation.

This last point deserves further comment. Although the term 'objective' may suggest a goal to be set aside once achieved, this isn't our usage here. Rather, an explanatory objective may (and typically does) remain in force even after we've found what we sought, in the sense that the question and its answer continue to interest us. Of course, our objectives do sometimes change, if we decide that some are misconceived, unfulfillable or no longer relevant. But the mere fact of finding a satisfying explanation doesn't force us to abandon the associated objective.[14]

For the same reason, it's perfectly possible to take up an objective only after discovering an explanation whose existence we didn't suspect. Doing so just means endorsing the explanatory question as worthwhile.

There's a final respect in which we depart from views like Lange's. Some explanatory objectives, it seems, don't take the form of why-questions at all. Rather, they're more naturally understood as "what's-going-on-here" questions, whose goal is to gain clarity about or familiarity with a set of phenomena. [Lehet 2021b] gives several examples. Consider, for instance, mathematicians' puzzlement about Weierstrass's discovery of continuous but nowhere differentiable functions. The surprising existence of such functions showed that analysts understood the foundations of their subject less clearly than they'd realized, and their desire to get to the bottom of things was in part an explanatory objective. But, as Lehet argues, "there does not seem to be a corresponding why-question [in this case]. Perhaps we could ask 'why are there examples of continuous nowhere differentiable functions?', but this doesn't seem to be what we are really interested in. We are interested in getting a clearer picture of the real numbers, one that highlights which properties give rise to such examples" (559).

Let us offer another example. During the development of his famous conjecture, Poincaré confronted and eventually answered a significant what's-going-on-here-question. The formulation of the Poincaré

[14]One reason for keeping an objective active after having met it is that, a priori, it might be possible to meet the objective again in a different or better way. You might have wanted, and found, a *unifying* explanation of theorem T, for instance. Even so, it may be wise to keep an eye out for a stronger and further-reaching unification, or one that subsumes T under combinatorial principles rather than algebraic ones (or whatever).

conjecture as we now know it began with Poincaré's 1895 publication of "Analysis Situs", the founding work of algebraic topology which introduced the concepts of homotopy and homology. "Analysis Situs" was a work in progress periodically supplemented by Poincaré over the next decade. It culminated in the fifth supplement, which presents the Poincaré conjecture about the topological characterization of the 3-sphere.

This project included several revisions to the notion of homology and resulted in multiple drafts of the conjecture. Throughout this work, it's clear that Poincaré is working toward a more adequate grasp of homology, homotopy and the relationship between the two. This process was in part a search for explanatory understanding: Poincaré wasn't merely trying to accumulate facts, but to clarify a set of profound and puzzling phenomena at the heart of algebraic topology.

Poincaré worked toward this objective throughout the process of writing "Analysis Situs" and its supplements, and we can see evidence of him progressing towards his aim. The first piece of evidence appears in the second supplement, where Poincaré is motivated by criticisms of his original publication to revise his notion of homology. In particular, objections raised by Poul Heegaard led him to incorporate a torsion coefficient into his conception of homology. This consideration was Poincaré's first step toward refining the distinction between homology and homotopy.

A second piece of evidence comes from the fifth supplement. Here Poincaré constructs a space, the Poincaré homology sphere, which serves as a counterexample to an early draft of the Poincaré conjecture proposed in the second supplement. This early version of the conjecture speculated that any homologically simple space would also be homotopically simple. The Poincaré homology sphere is an example of a space that meets the first condition but not the second.[15] Poincaré's discovery of this counterexample—which incorporated his new notion of a space's *fundamental group*, now recognized as a key invariant in algebraic topology—illustrates the progress he'd made toward conceptual clarity. And indeed, it's immediately after presenting this counterexample that Poincaré proposes the canonical version of his conjecture.

It seems undeniable that explanation was among Poincaré's epistemic goals. But the sort of explanatory understanding he sought is hard to capture in the form of a why-question: we can't point to any single fact, or any single outstanding feature of a fact, whose truth Poincaré was bent on accounting for. A what's-going-on-here formulation better captures Poincaré's objective: "What's going on in the foundations of topology such that homology and homotopy are closely related yet apparently somehow different, and both important for characterizing the 3-sphere?"

3.2.5 Mathematical and scientific explanation

A fourth important question about our account: To what extent does this picture generalize to scientific explanation more broadly? Is our brand of noeticism viable outside pure mathematics?

[15] Note that the converse claim is true: every homotopically simple space is indeed homologically simple. So this early conjecture proposed (incorrectly) that the two properties are equivalent.

Broadly speaking, we think the answer is yes, although the two domains differ in some interesting ways. As indicated above, we're sympathetic to [Rice & Rohwer 2021]'s cluster-concept view of scientific explanation, although we prefer to maintain monism about explanation itself and instead embrace a cluster view at the level of explanatory objectives. Our core noeticist proposal generalizes quite straightforwardly in this sense. Indeed, many of the same kinds of objectives figure prominently in scientific and mathematical practice alike—inquirers in both realms are broadly interested in Seeing the Big Picture, Getting a Handle, Putting Things in Order and Accounting for Surprises.

That said, the precise objectives in play and their relative importance differ substantially between the two domains. Most obviously, *causal* explanation is widespread in natural science but seemingly absent in mathematics. And while many scientific explanations are noncausal ([Reutlinger & Saatsi 2018]), a large share of these appear to involve ontic relations of dependence or determination which at least formally resemble causation. Explanations which identify dependence relations also occur in pure mathematics, but we take them to be relatively rare and peripheral there. Counterfactual information seems to exhibit the same asymmetry: a common explanatory target in natural science but a minor one in mathematics.[16]

There's a good epistemic reason for this divergence in objectives. Empirical scientists justifiably care more about control, manipulation, mechanisms and difference-makers than mathematicians because only the former are in a position to perform interventions and observe the results. Mathematicians enjoy no such interactive relationship with their objects of study. And mathematics, in any case, lacks much of the temporal, causal and metaphysical dependence structure of the natural world.

So it's no surprise that mathematicians' explanatory goals tend to prioritize other kinds of insight. While unifying explanations are important in both domains, abstraction, generalization and intertheoretic connections are more prominent epistemic goals in mathematics. A more unique feature is the special role of proof in mathematical practice. The ability of an argument to display reasons clearly and efficaciously is thus a distinctively mathematical concern.

Much more could be said about the relationship between scientific and mathematical explanation. Although we can't do the task full justice here, we believe a uniform noeticist account along the lines we've sketched has the resources to make sense of both the similarities and differences between the two domains.

3.2.6 Empirical work on mathematicians' judgments

Let us finally connect our proposal to some relevant empirical work.[17] The best study to date of mathematicians' judgments about explanation is [Mejía Ramos et al. 2021]. Using a pairwise-comparison setup, Mejía Ramos and colleagues obtained 760 judgments from 38 mathematicians on the relative explanatory values of nine proofs (of a simple statement about the roots of a cubic polynomial). Their

[16]For compelling criticism of monist counterfactual theories of mathematical explanation, see [Kasirzadeh 2021] and [Lange 2022].

[17]Thanks to an anonymous referee for prompting the discussion in this section.

analysis turned up several noteworthy results. Attempting to account for these phenomena provides a nice test for us (and, of course, for competing theories).

The mathematicians studied strongly agreed on average about which proofs were more and less explanatory, their ratings exhibiting a split-half inter-rater reliability coefficient of .947 (with 1 indicating perfect correlation) (587). This is good news for views which expect experts' judgments to track objective explanatory success, and bad for those which cast explanatory preferences as arbitrary, unstable or context-bound in ways that make such agreement unlikely. We regard ourselves as belonging to the former camp. Still, one might think the existence of such a strong consensus is somewhat mysterious on our view. Why, after all, wouldn't a large sample of mathematicians exhibit considerable diversity in explanatory objectives, and hence substantial disagreement about which proofs best explain?

We hypothesize that most reasoners, when instructed to evaluate explanations for an arbitrary fact regarding which they have no special prior context or personal motivation, tend to default to a general-purpose explanation-seeking mode especially attuned to readily identifiable and non-domain-specific types of explanatory reasoning (e.g., to transparent proofs). This hypothesis predicts the observed agreement, since the highest-rated proofs were simple and direct, while the least preferred proofs were idiosyncratic, cumbersome and fussy (e.g. an approximation argument including a large data table from a computer algebra system, or a laborious cube-slicing proof consisting of 15 annotated images).

Our hypothesis also predicts a second striking finding. Among the nine proofs shown to participants, two presented essentially the same reasoning in different textual forms: the "Elementary" proof gave a short algebraic argument in a standard prose style, while the "Two-column" proof displayed each step of this argument in an unwieldy list. In spite of their logical similarity, the Elementary proof was the most preferred overall in the study, while the Two-column proof was judged much less explanatory. This result further confirms our suggestion. We'd expect readers in this contrived setting to be particularly attuned to general-purpose explanatory virtues like transparency, and the Elementary proof rates much more highly on this score. We see no compensating virtue in the Two-column proof.

By contrast, ontic and counterfactual accounts of explanation can't easily account for these data. As Mejía Ramos and colleagues point out, such accounts seem to imply that the Elementary and Two-column proofs should stand or fall together. It's also unclear why, for instance, the cube-slicing Visual proof would be judged unexplanatory on these accounts, since it's plausible that the geometric interpretation of polynomial expressions involves an appropriate kind of intertheoretic dependence relation.[18]

We hope to see further high-quality empirical work in this vein. It would be useful, for instance, to probe the reasons why mathematicians exhibit this particular pattern of explanatory preferences; data of this sort could help test hypotheses like the one we've offered here, and more generally help adjudicate the case between noeticism and its rivals.

[18]See [D'Alessandro 2020] for further discussion of ontic accounts of mathematical explanation.

3.2.7 Summary

Let us summarize the theory of mathematical explanation we've begun to develop here. (Incidentally, if our view needs a name to distinguish it from other noeticist accounts, "telic noeticism" might do.) The view is that we possess explanatory understanding when we've successfully met an explanatory objective (and correctly recognized that we've done so). *Explanatory objective* is a cluster concept: such objectives fall into a few broad and overlapping categories without sharing any essential feature in common. Objectives sometimes take the form of specific, explicit why-questions formulated in advance of inquiry, but what's-going-on-here questions and other sorts of objectives also play important roles in mathematical practice. The value of an explanatory objective is multifaceted, but mostly independent of particular agents' beliefs and preferences. Finally, we endorse a noeticist theory of scientific explanation that largely overlaps with our mathematical account, though the landscape of objectives differs somewhat in composition and emphasis between the two domains.

In his groundbreaking [Friedman 1974], Michael Friedman proposed three conditions on an adequate theory of explanation. First, such a theory should be *sufficiently general*, in the sense that it applies to and makes sense of the diversity of explanations throughout the sciences. Second, it should be *objective*, in the sense that it doesn't deem facts about explanatoriness to hold in virtue of arbitrary non-rational predilections. Third, it should explicitly *connect explanation and understanding*, in the sense that it tells us "what kind of understanding scientific explanations provide and how they provide it" (14). Friedman's constraints have, we think, aged well, though few accounts of explanation satisfy them even today. We note that ours does so.

So far as we know, the view nearest our own in the literature is that of [Frans & Van Kerkhove 2023]. Frans and Van Kerkhove's account reflects a noeticist orientation toward explanation[19], and it highlights the link between achieving understanding and satisfying specific "epistemic aims". Though our projects are in alignment in these respects, we differ on other important points.

For one, Frans and Van Kerkhove accept an ability account of understanding. We don't share this assumption. Also, like IMR, Frans and Van Kerkhove appear sympathetic to the idea that explanatory understanding reduces to objectual understanding (30). We've argued against this notion at length above. Finally, Frans and Van Kerkhove seem to be primarily concerned with explanations which are *regarded as* good by a given individual or group, while our account focuses on objective explanatory success. As a result, they make much of acceptance conditions (such as an inquirer's background knowledge and historical context) which we largely gloss over. On the other hand, they have little to say about what makes some epistemic aims distinctively explanatory, which we take to be the central task of a noeticist account. In spite of sharing certain motivating commitments, then, we view the details of our projects as

[19]Frans and Van Kerkhove speak of "epistemic approaches to explanation" where we use the language of noeticism.

largely either orthogonal or at odds for reasons previously canvassed.

4 Conclusion

The view we've defended, "telic noeticism", has many appealing features. Like all versions of noeticism, it possesses the virtues described in §1—it accounts straightforwardly for the link between explanation and understanding, it makes sense of and can incorporate the successes of non-noetic theories, and it can handle diverse types of mathematical explanantia. By observing the important distinction between objectual and explanatory understanding, our view also avoids the problems faced by Inglis and Mejía Ramos's noeticist account. Last but not least, the details of our view answer a number of questions faced by noeticists and non-noeticists alike: about the relationship between explanations and why-questions, the proper response to the apparent monism-pluralism dilemma, how to reconcile the agent-based aspects of explanation with an objective conception of explanatory value, what to say about the unity of scientific and mathematical explanation, and how to square theory with empirical data on experts' explanatory judgments. In doing so, the view answers Friedman's half-century-old demand for a general and objective theory which sheds light on the link between explanation and understanding. These virtues should help establish noeticism as a serious option for those fascinated and vexed (as we've long been) by explanation in mathematics.[20]

William D'Alessandro, William & Mary, USA

Ellen Lehet, independent, USA[21]

References

[Appel & Haken 1976] Appel, K. and W. Haken. 1976. "Every planar map is four colorable." *Bulletin of the American Mathematical Society* 82, 711-712.

[Baker 2009] Baker, Alan. 2009. "Mathematical accidents and the end of explanation." In Otávio Bueno and Øystein Linnebo (eds.), *New Waves in Philosophy of Mathematics*, New York: Palgrave Macmillan, 137-159.

[20] We're happy to have many people to thank for their advice and encouragement during this paper's journeys through the philososphere. We're grateful to generous audiences at the 2022 Philosophy of Science Association meeting in Pittsburgh, the 2022 Midwest Philosophy of Mathematics Workshop at Notre Dame, the 2022 Understanding Progress and Progress in Understanding conference at the University of Iceland, the 2023 Understanding Mathematical Explanation conference at Rutgers University, the Czech Academy of Sciences, the Munich Center for Mathematical Philosophy and the University of Oxford. We especially thank Silvia De Toffoli, Dan Isaacson, Marc Lange, Pablo Mejía Ramos, Alex Paseau and Chris Pincock, along with two anonymous referees for this journal whose comments led to many improvements.

This paper is dedicated to Joachim Frans, who died suddenly and tragically in late 2023. Joachim was a valued colleague, friend and pillar of the European philosophy of math community. We regret that he can't carry on the debate we've joined, but we hope to honor his work by carrying forward our shared views about explanation and understanding.

[21] Both authors contributed equally to developing the paper's ideas; the majority of the writing was done by WD.

[Baron et al. 2020] Baron, Sam, Mark Colyvan and David Ripley. 2020. "A counterfactual approach to explanation in mathematics." *Philosophia Mathematica* 28, 1-34.

[Bokulich 2018] Bokulich, Alisa. 2018. "Representing and explaining: The eikonic conception of scientific explanation." *Philosophy of Science* 85, 793-805.

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

[D'Alessandro 2017] D'Alessandro, William. 2017. "Mathematical explanation beyond explanatory proof." *British Journal for the Philosophy of Science* 71, 581-603.

[D'Alessandro 2019] D'Alessandro, William. 2019. "Explanation in mathematics: Proofs and practice." *Philosophy Compass.* DOI: 10.1111/phc3.12629.

[D'Alessandro 2020] D'Alessandro, William. 2020. "Viewing-as explanations and ontic dependence." *Philosophical Studies* 177, 769–792.

[D'Alessandro 2021] D'Alessandro, William. 2021. "Proving quadratic reciprocity: Explanation, disagreement, transparency and depth." *Synthese* 198, 8621-8664.

[Dellsén 2021] Dellsén, Finnur. 2021. "Understanding scientific progress: The noetic account." *Synthese* 199, 11249-11278.

[Detlefsen & Arana 2011] Detlefsen, Michael and Andrew Arana. 2011. "Purity of methods." *Philosophers' Imprint* 11, 1-20.

[Frans & Van Kerkhove 2023] Frans, Joachim and Bart Van Kerkhove. 2023. "Mathematical explanation: Epistemic aims and diverging assessments." *Global Philosophy* 33, 1-26.

[Frans & Weber 2014] Frans, Joachim and Erik Weber. 2014. "Mechanistic explanation and explanatory proofs in mathematics." *Philosophia Mathematica* 22, 231-248.

[Friedman 1974] Friedman, Michael. 1974. "Explanation and scientific understanding." *Journal of Philosophy* 71, 5-19.

[Inglis & Mejía Ramos 2021] Inglis, Matthew and Juan Pablo Mejía Ramos. 2021. "Functional explanation in mathematics." *Synthese* 198 (Suppl. 26), S6369-S6392.

[Jacquette 2003] Jacquette, Dale. 2003. "Introduction: Psychologism the philosophical shibboleth." In Dale Jacquette (ed.), *Philosophy, Psychology, and Psychologism: Critical and Historical Readings on the Psychological Turn in Philosophy*, Dordrecht: Kluwer, 1-20.

[Kasirzadeh 2021] Kasirzadeh, Atoosa. 2021. "Counter countermathematical explanations." *Erkenntnis* 88, 2537-2560.

[Kitcher 1989] Kitcher, Phillip. 1989. "Explanatory unification and the causal structure of the world." In Phillip Kitcher and Wesley Salmon (eds.), *Scientific Explanation* (Minnesota Studies in the Philosophy of Science, Volume XIII), Minneapolis: University of Minnesota Press, 410-505.

[Lange 2014] Lange, Marc. 2014. "Aspects of mathematical explanation: Symmetry, unity, and salience." *Philosophical Review* 123, 485-531.

[Lange 2016] Lange, Marc. 2016. *Because Without Cause: Non-Causal Explanations in Science and Mathematics.* New York: Oxford University Press.

[Lange 2022] Lange, Marc. 2022. "Challenges facing counterfactual accounts of explanation in mathematics." *Philosophia Mathematica* 30, 32-58.

[Lehet 2021a] Lehet, Ellen. 2021. "Impurity in contemporary mathematics." *Notre Dame Journal of Formal Logic* 62, 67-82.

[Lehet 2021b] Lehet, Ellen. 2021. "Mathematical explanation in practice." *Axiomathes* 31, 553-574.

[Mancosu 2008] Mancosu, Paolo. 2008. "Mathematical explanation: Why it matters." In Paolo Mancosu (ed.), *The Philosophy of Mathematical Practice*, New York: Oxford University Press, 134-150.

[Mejía Ramos et al. 2021] Mejía Ramos, Juan Pablo, Tanya Evans, Colin Rittberg and Matthew Inglis. 2021. "Mathematicians' assessments of the explanatory value of proofs." *Axiomathes* 31, 575–599.

[Moore & Mertens 2011] Moore, Cristopher and Stephan Mertens. 2011. *The Nature of Computation.* New York: Oxford University Press.

[Pincock 2015] Pincock, Christopher. 2015. "The unsolvability of the quintic: A case study in abstract mathematical explanation." *Philosophers' Imprint* 15, 1-19.

[Pincock 2023] Pincock, Christopher. 2023. *Mathematics and Explanation.* Cambridge: Cambridge University Press.

[Nummela 1987] Nummela, Eric. 1987. "No coincidence." *Mathematical Gazette* 71, 147.

[Reutlinger & Saatsi 2018] Reutlinger, Alexander and Juha Saatsi (eds.). 2018. *Explanation Beyond Causation: Philosophical Perspectives on Non-Casual Explanations.* New York: Oxford University Press.

[Rice & Rohwer 2021] Rice, Collin and Yasha Rohwer. 2021. "How to reconcile a unified account of explanation with explanatory diversity." *Foundations of Science* 26, 1025-1047.

[Ryan 2023] Ryan, Patrick. 2023. "Szemerédi's theorem: An exploration of impurity, explanation, and content." *Review of Symbolic Logic* 16, 700-739.

[Sandborg 1998] Sandborg, David. 1998. "Mathematical explanation and the theory of why-questions." *British Journal for the Philosophy of Science* 49, 603-624.

[Seymour 2016] Seymour, Paul. 2016. "Hadwiger's conjecture." In John Forbes Nash and Michael Rassias (eds.), *Open Problems in Mathematics*, Berlin: Springer, 417-437.

[Sheredos 2016] Sheredos, Benjamin. 2016. "Re-reconciling the epistemic and ontic views of explanation." *Erkenntnis* 81, 919-949.

[Steiner 1978] Steiner, Mark. 1978. "Mathematical explanation." *Philosophical Studies* 34, 135-151.

[Trout 2002] Trout, J.D. 2002. "Scientific explanation and the sense of understanding." *Philosophy of Science* 69, 212-233.

[van Fraassen 1980] van Fraassen, Bas. 1980. *The Scientific Image.* Oxford: Clarendon Press.

[Wilkenfeld 2014] Wilkenfeld, Daniel. 2014. "Functional explaining: A new approach to the philosophy of explanation." *Synthese* 191, 3367-3391.

[Wright 2015] Wright, Cory. 2015. "The ontic conception of scientific explanation." *Studies in History and Philosophy of Science* 54, 20-30.