Unrealistic Models in Mathematics

William D'Alessandro

Forthcoming in *Philosophers' Imprint*

From particle physics to climatology to macroeconomics, scientists confront phenomena which they'd like to better understand, but which are too complex to feasibly study in realistic detail. In such situations, researchers often turn to models: surrogate systems that are simpler or more tractable than the target phenomenon, but similar enough to it to offer insight.

Scientists' modeling practices raise a number of philosophical questions. Some especially challenging and interesting ones pertain to *unrealistic*, *fictional* or *essentially idealized* models and their role in science. Such cases have been a major theme of recent work in philosophy of modeling, which has seen debates around questions like these:

- Can we gain genuine scientific understanding from unrealistic models?
- Can unrealistic models explain anything about their target systems? If so, is this the source of their ability to generate understanding?
- Are unrealistic models useful primarily because they give us counterfactual knowledge about their target systems?

My goal is to bring pure mathematics into these conversations. Doing so is appropriate because mathematicians are modelers too: models of all sorts, and unrealistic models in particular, are used in similar ways and for similar reasons across pure math and the empirical sciences.

This fact is worth advertising in its own right. Although idealized models are an indispensable part of the toolkit in many areas of mathematics, their existence is rarely noted either by philosophers of mathematical practice or by modeling theorists. Indeed, as far as I'm aware, no single instance of modeling in this sense has ever been examined in detail—a stark contrast with the myriad case studies from across the sciences.

Toward a correction of this omission, my first objective is look carefully at two examples from contemporary number theory. The first is *Cramér's random model* of the primes, the second the *function field model* of the integers. Both models are important and widely used research tools whose existence philosophers ought to be better informed about.

Perhaps more importantly yet, paying attention to modeling in mathematics can help settle some of the contentious questions mentioned above. With the help of the two case studies, I plan to argue that:

- Mathematicians make extensive use of unrealistic models and derive understanding from them.
- It's not always the case that this understanding is mediated by explanation. An unrealistic model can help us understand a phenomenon even when it offers no explanation of the phenomenon.

- It's not always the case that unrealistic models contribute to understanding (or otherwise make themselves useful) by imparting counterfactual knowledge.

The latter two claims especially constitute challenges to popular views in philosophy of science. Taking cases from mathematics seriously, then, can help move debates about modeling forward.

§1 below makes some preliminary clarifications, about what I take models to be and what makes a model unrealistic. §2 and §3 discuss Cramér's model and the function field model, respectively. §4 answers the questions posed above about understanding, explanation and counterfactual knowledge, and closes with a plea for greater contact between pure mathematics and philosophy of science.

1 Unrealistic models in mathematics

Let me begin with some comments about the scope of my study and its rationale.

First, the subject matter of this paper—mathematics and models—might bring to mind the branch of mathematical logic known as model theory. For the model theorist, a model is a structure that satisfies a given set of sentences in a specified formal language under an interpretation. I'm not primarily interested in this special sense of 'model', but rather in the broader scientific meaning of the term. In this sense, I take it, a model is any object M that's used to represent some other phenomenon, system, or body of information P. There are no a priori restrictions here on the nature of M or its relationship to P: in particular, there's no requirement that M satisfy some set of sentences associated with P.

Another clarification. It's in the nature of a model to take some liberties with its target phenomenon—"all models are wrong", as the old saying goes[1]—but different models do so in different ways and to different degrees. Some abstract away from irrelevant details but are basically realistic on the whole: their elements represent only real features of the target phenomenon, all the most important features are represented, and these representations are more or less accurate. Moreover, such models can often be "de-idealized" even further without fundamentally changing their character (by adding in missing details or relaxing simplifying assumptions, for example).

In other cases, however, the relationship between surrogate and reality is less tidy. Many models explicitly and essentially misrepresent key aspects of their target phenomena, and hence are nonveridical in a deeper sense. In these cases, no simple de-idealization procedure is available: the models are what they are, and function as they do, precisely on account of the distortions they contain. This latter sort of case is what I mean by an *unrealistic model*.[2]

To get a clearer sense for the distinction, consider on the one hand a schematic street map, or a simple lunar model of the tides. Both models omit some features of their targets: the map may not depict the relative widths of the streets, or the locations of alleys and unpaved drives, while the model of the tides neglects the gravitational influence of the Sun and the effects of Earth's rotation. Nevertheless, both accurately represent the most important features of their target systems without introducing major ontological or ideological distortions.

[1] The saying is usually attributed to the statistician George Box; the Wikipedia entry "All models are wrong" has an extensive discussion of its history (https://en.wikipedia.org/wiki/All_models_are_wrong).

[2] Other names in the literature for roughly this type of model include "fictional model", "essentially idealized model", "pervasively distorted model", and so on.

Compare, on the other hand, Bohr's model of the atom or Schelling's model of housing segregation. It's essential to Bohr's model that it portrays electrons as moving in well-defined orbits around their nuclei, when in fact they do no such thing. The electron orbitals, as Alisa Bokulich puts it, are *fictions*, which "[cannot] be properly thought of as an 'idealization' of the true quantum dynamics" ([Bokulich 2011], 43). Meanwhile, the Schelling model represents an agent's housing choices as completely determined by two factors: their preference to be surrounded by a certain percentage of neighbors from their own group, and the current composition of their immediate neighborhood. Cost considerations and other factors of obvious real-world importance are absent from the model, which is therefore usually viewed as a *toy* model: a "strongly idealized" and "extremely simple" representation that omits most of the factors on which the target phenomenon depends (cf. [Reutlinger et al. 2018]). Unrealistic models raise some especially interesting questions, and their appearances in mathematics will be my focus below.

2 The Cramér random model

One phenomenon that's often studied via models is the distribution of the prime numbers. In this section I'll describe one of the most important and widely used of these: Cramér's random model of the primes. Since Cramér introduced the model in 1932, a large class of refinements, spinoffs and other variants have emerged. Some of these are more accurate or useful than the original model for certain purposes. I focus mostly on the original below, since it's the simplest and it remains in frequent use.

Let me start with some background. Famously, and perhaps to a greater degree any other branch of mathematics, number theory is rife with simple and natural questions that have proven very hard to answer. Among the most well-known examples are the four Landau problems[3]:

1. **Goldbach's conjecture:** Is every even integer greater than 2 the sum of two primes?
2. **Twin Primes conjecture:** Are there infinitely many pairs of prime numbers of the form p, $p+2$?
3. **Legendre's conjecture:** Is there a prime between n^2 and $(n+1)^2$ for every positive integer n?
4. **Fourth Landau conjecture:** Are there infinitely many primes of the form n^2+1?

Settling these conjectures requires understanding how the primes are arranged among the natural numbers. This is no easy task, since "the series of prime numbers exhibits great irregularities of detail" ([Ingham 1932], 1) and "do[es] not follow any apparent pattern" ([Koukoulopoulous 2019], 1).[4] Thus the first three questions have each remained open for 170 years or more.[5] Although existing technology still doesn't seem up to the challenge of solving the Landau problems, we've learned enough to have made some headway.

The very first relevant discovery was Euclid's theorem that there are infinitely many prime numbers. This is a precondition for the conjectures' possible truth, but not helpful for their resolution, since it gives no distributional information about the primes. For this we need the much more recent prime number theorem

[3]Named for Edmund Landau's 1912 address to the International Congress of Mathematics, which characterized them as "unattackable" by the methods of contemporary number theory.

[4]Of course, this isn't to say that the prime sequence is completely random (in the sense that there's no deterministic procedure for generating its terms), or that the sequence has no meaningful structure at all. Neither of these things are true. The point is that the sequence's structure is in certain respects elusive and hard to study.

[5]Goldbach's conjecture dates to a letter from Goldbach to Euler in 1742 and is one of the oldest unsolved problems in mathematics. The Twin Primes conjecture is first known to have been explicitly stated by de Polignac in 1849, but the idea was plausibly considered much earlier. Legendre's conjecture is from his *Essai sur la Théorie des Nombres*, published in 1797-8.

(1896, proved independently by Hadamard and de la Vallée Poussin). Where $\log x$ is the natural logarithm and $\pi(x)$ is the prime-counting function (giving the number of primes less than or equal to x), the PNT says that

$$\pi(x) \sim \frac{x}{\log x},$$

i.e., that the number of primes up to x approaches $\frac{x}{\log x}$ as x goes to infinity. This means the primes steadily thin out among the natural numbers, but they do so at a relatively slow rate. A consequence of the PNT is that, for sufficiently large n, the probability that n is prime is about $\frac{1}{\log n}$—a fact we'll return to below.[6]

Unfortunately, this still isn't enough information about the distribution of the primes to settle Landau's problems. (A proof of the Riemann Hypothesis would help—the RH can be viewed as an improvement on the PNT, bounding how far off $\pi(x)$ can be from $\frac{x}{\log x}$—but that seems unlikely to be forthcoming any time soon.)

In view of these difficulties, the Swedish mathematician Harald Cramér proposed a new way of approaching the distribution problem. Rather than directly studying the primes themselves, he constructed a more tractable surrogate, now known as *Cramér's model* or the *random model* of the primes. The idea, set out in [Cramér 1936], is to build a subset of the natural numbers by independently choosing to include each $n > 2$ with probability $\frac{1}{\log n}$. The resulting sequence might look something like this[7]:

$$3, 4, 6, 11, 12, 25, 26, 28, 32, 34, 35, 36, 43, 57, 66, 68, 80, 83, 87, 93, ...$$

$$100005, 100006, 100008, 100018, 100045, 100055, 100074, 100094, 100096, 100106, 100119, \ldots$$

Think of this set as a model of the real prime sequence. By the consequence of the prime number theorem mentioned above, the "primes" in Cramér's model will have the same asymptotic density in the natural numbers as do the real primes (with probability 1)—one can observe, for instance, that the larger terms in the sample sequence are spaced out somewhat more than the smaller ones. So it's reasonable to hope that other statistical and distributional properties of the real prime sequence will also resemble those of the model. (Note, by the way, that talk of 'the model' refers to an arbitrary sequence generated by Cramér's procedure, not to any definite sequence in particular. Correspondingly, claims of the form 'P is true in the model' mean that P holds for an arbitrary such sequence with probability 1, perhaps with finitely many exceptions.[8])

Given that the the "Cramér primes" and the real primes are similarly distributed in $\mathbb{N}$, what's the benefit of working with the former instead of the latter? As it turns out, it's much easier to study the statistics of distributions with strong joint independence properties like those of the random model.[9] Consequently, we know a lot about the behavior of the surrogate primes.

Some of these statements were independently known, or strongly believed, to be true of the actual primes.

[6] To see this, note that there are about $\frac{n}{\log n}$ prime numbers among the first n integers. So the probability that any given number between 1 and n is prime is $\left(\frac{n}{\log n}\right)/n = \frac{1}{\log n}$.

[7] These are excerpts from a sequence generated by Mathematica code written by Glenn Harris. Many thanks to him for the code.

[8] A bit more precisely and in terms of a concrete example: the claim that Goldbach's conjecture holds in Cramér's model means that, in an arbitrary sequence of Cramér primes, with probability 1, the number of ways to express an even integer n as a sum of two primes grows large as $n \to \infty$. (We can't yet even prove that this limit is bigger than zero in the case of the actual primes.)

[9] Recall that each choice of a Cramér prime is made independently of all the other choices. Things obviously don't work this way in the real world; if p is an odd prime, for instance, then $p+1$ and $2p$ are necessarily composite.

Others are considered to have gained support from the fact that they hold in the model (e.g. the Riemann Hypothesis and the Landau conjectures). Yet other hypotheses were originally motivated by observations about the model itself. Among these is the important Cramér conjecture on the sizes of gaps between primes,[10] introduced in Cramér's original paper on the model, which even in recent years "does not seem to be attackable by other methods" ([Granville 1995b], 391).

I want to make one observation and two more substantive claims about this case. The observation is that the Cramér model is manifestly not a "model of the theory of the primes" in the model-theoretic sense. There are many sentences true of the real primes that are false of the Cramér primes (for instance, "exactly one even number is prime").[11] So model theory isn't the proper framework for thinking about this situation, as per the remarks in §1 above.

The first substantive claim I want to defend is that Cramér's model (and similar random models) have significantly improved our understanding of the distribution of the primes. The model's epistemic contributions are of several kinds.

To start with, mathematicians take the model seriously because it correctly predicts many known facts about the primes. Kannan Soundararajan notes, for instance, that "the Cramér model makes accurate predictions for the distribution of primes in [very] short intervals" ([Soundararajan 2007a], 64). (Note that this isn't just a trivial consequence of the fact that the model gets the asymptotic density right: two sequences can have similar long-run behavior without looking alike at small scales.[12]) More generally, Andrew Granville writes that "the probabilistic model usually gives one a strong indication of the truth" ([Granville 1995b], 391). In Tao's words, "[we] have a number of extremely convincing and well supported models for the primes... the most accurate [of these] in practice are *random* models" ([Tao 2015]).

Since Cramér-type models have proven generally reliable in regimes where their predictions can be verified, number theorists view them as useful guides to unknown territory. Their contributions along these lines fall into at least three categories: (1) increasing or decreasing our confidence in conjectures derived from independent sources; (2) motivating entirely new conjectures; and (3) suggesting novel methods of proof.

Start with (1). As mentioned above, a number of fundamental statements in number theory are known to hold in the Cramér model, though not yet known to hold for the actual primes. These include the Riemann Hypothesis and the four Landau conjectures. Each of these hypotheses was suspected to be true before the advent of the Cramér model. But the results from the model served to increase mathematicians' confidence, in some cases significantly. For instance, van der Poorten claims that "the most compelling" evidence in favor of the Riemann Hypothesis is the fact that RH holds in a related random model of the primes, the Hawkins model ([van der Poorten 1996], 147).[13] Similarly, according to Patterson's textbook on the zeta function, the validation of RH by random models "represents one of the more reassuring reasons for expecting the Riemann Hypothesis to be true" ([Patterson 1988], 75).

The credence calibration provided by Cramér-style models extends much further than RH. Tao elaborates on this point. In the setting of these models, he writes, "many difficult conjectures on the primes reduce to relatively simple calculations... Indeed, the models are so effective at this task that analytic number theory

[10] Cramér's conjecture is the statement that, for p_n the nth prime, the difference $p_{n+1} - p_n$ is asymptotically bounded by $(\log p_n)^2$. Hence the gaps between consecutive primes are consistently small in the long run.

[11] I'm assuming sentences like these are given their obvious interpretations in the model.

[12] As Colin McLarty put the point in correspondence: "A similar fallacy, fed by motivated thinking, is important today when climate change deniers say things like 'They can't even be sure if it will rain next Sunday! How can they make predictions about 20 years from now?'" In general, the moral is that the ability to predict large-scale trends over long intervals with a high degree of accuracy doesn't imply the ability to do the same with fine-grained details over short intervals.

[13] The Hawkins model generates a set of surrogate primes by a random sieve technique. For an accessible introduction to the Hawkins model, including a comparison with the Cramér model, see [Lorch & Ökten 2007].

is in the curious position of being able to confidently predict the answer to a large proportion of the open problems in the subject, whilst not possessing a clear way forward to rigorously confirm these answers!" ([Tao 2015])

Next, (2). In addition to bolstering confidence in independently motivated hypotheses, "the probabilistic heuristic, in which independence is assumed, provides a useful means of constructing conjectures" ([Montgomery & Vaughan 2007], 57).[14] The most famous of these is the Cramér conjecture on prime gaps mentioned above, which Cramér arrived at by way of the model. Random models have also led to progress in other parts of number theory. One example is the theory of "lucky numbers" (the sequence $1, 3, 7, 9, 13, 15, 21, 25, ...$, generated by a certain sieve process).[15] On the basis of his random model of the primes, Hawkins conjectured (in [Hawkins 1957]) and was later able to prove (in [Hawkins & Briggs 1957]) a PNT-type result for lucky numbers, to the effect that their asymptotic density in the natural numbers is also $\frac{1}{\log n}$. Random models continue to be deployed on the front lines of research, sometimes in novel ways. The recent [Lozano-Robledo 2020] "propose[s] a new probabilistic model for the distribution of ranks of elliptic curves... in the spirit of Cramér's model for the prime numbers" (2), which is used to generate predictions about the number of elliptic curves of a given rank. In general, then, Cramér-type models "[give] a clearer indication of what results one expects to be true, thus guiding one to fruitful conjectures" ([Tao 2015]).

Finally, (3). As we've seen, models of the primes are generally used for heuristic purposes rather than as tools for proving theorems. Nevertheless, in the view of the number theorist János Pintz, "probabilistic models can help or could have helped not only to conjecture but also prove results about primes" ([Pintz 2007], 362). Pintz goes on to show how a particular result—Maier's theorem about the number of primes in small intervals—could have been established much earlier on the basis of a modified Cramér model.

In addition to these three applications, Tao mentions several other uses of random models: "providing a quick way to scan for possible errors in a mathematical claim (e.g. by finding that the main term is off from what a model predicts...); gauging the relative strength of various assertions (e.g. classifying some results as 'unsurprising' [and] others as 'potential breakthroughs'...); or setting up heuristic barriers... that one has to resolve before resolving certain key problems" ([Tao 2015]). In view of this list of uses, benefits and insights, I conclude that number theorists have gained significant understanding from Cramér-type models.

One could try to push back against this claim by noting the lack of philosophical consensus around the notion of understanding. In the absence of a widely accepted explicit theory, which criteria are being used to judge cases like this? And why should we think those criteria are appropriate?

It's true that philosophers disagree about understanding. For instance, some equate understanding a phenomenon with having an explanation of it ([Strevens 2013]). Others link understanding with the possession of certain abilities ([Delarivière & Van Kerkhove 2021]) or suitably structured knowledge ([Kelp 2015]), or with the disposition to generate new knowledge from a minimal core ([Wilkenfeld 2019]). I won't choose a side in this debate here (though I do argue against the explanation account in §4 below). My approach is different, and it has two components. First, I claim that the Cramér model ought to count as a source of understanding on any reasonable view, and similarly for the function field model discussed in the next section. Second, I offer the appraisals of mathematicians themselves, which I take to count at least as strongly

[14] "The probabilistic heuristic, in which independence is assumed" refers, of course, to the method of constructing random models by independently choosing surrogate primes.

[15] The lucky numbers are generated in the following way. First list all the natural numbers starting with 1. Then cross out every second number, leaving the sequence $1, 3, 5, 7, 9, 11, \ldots$. Next cross out every third number, leaving $1, 3, 7, 9, 13, 15, \ldots$. At every successive step, cross out every nth number, where n is the first surviving number on the list such that every nth number hasn't yet been crossed out.

as philosophical arguments in this context.[16]

The Cramér model, as just shown, has strengthened number theorists' confidence in some important hypotheses and has played a key role in generating some others. It's led to a clearer overall picture of the phenomena. It helps mathematicians organize, justify and check their reasoning. It would be a tendentious and implausible theory that regarded these achievements as insufficient for improving understanding. In particular, the model evidently confers abilities associated with understanding, lends valuable structure to number theorists' knowledge, and allows lots of novel information to be spun out from a compact representational core. So theories in the spirit of the last three mentioned above will count the model as a source of understanding. And this seems correct.

What's more, the same conclusion has been reached by number theorists who are intimately familiar with the model and its uses. Granville, for instance, refers to "Cramér's probabilistic approach [to] understanding the distribution of prime numbers, which underpins most of the heuristic reasoning still used in the subject today" ([Granville 1995a], 15). Absent compelling reasons to do otherwise, good methodology recommends taking such judgments at face value.[17]

I conclude from these considerations that the Cramér model is a source of understanding. (To be precise, its contribution is to understanding the distribution of the prime numbers. I take this to be a case of understanding a phenomenon, as opposed to, say, a case of understanding-why.)

The second main claim of this section is that Cramér's model is quite *unrealistic*, in the sense discussed in §1 above. That is, rather than a mild idealization which merely abstracts away from inessential details, the model involves an explicit and extensive misrepresentation of its subject matter.

One major distortion is that the Cramér primes are chosen probabilistically, but the actual primes aren't in any sense random. Rather, as George Pólya says, whether a number is prime "can be decided by the 'definite rules' of arithmetic—where and how could chance enter the picture?" ([Pólya 1959], 376). Although the assumption of randomness is unrealistic, however, it's essential to all Cramér-type models. There's no prospect of de-idealizing to remove this assumption without discarding the model framework entirely.

Cramér's model also fails to capture the important multiplicative structure of the actual primes—for instance, the fact that if p is prime then $n \cdot p$ can't be. (Recall that the Cramér primes are chosen independently, so the selection of one number has no effect on the probability of choosing any other number.) Hence the model generates infinitely many even primes, pairs of consecutive primes and other absurdities—for example, in the run of the Cramér algorithm given above, 34, 35 and 36 are all chosen. Some modifications of the simple Cramér model reintroduce basic aspects of multiplicative structure, e.g. by forbidding even primes greater than 2. But going much further in the direction of realism would again be counterproductive, since the joint independence of the surrogate primes is exactly the feature that makes the models more tractable than the real primes.

Thus, "[d]espite its predictive power, Cramér's model is a vast oversimplifcation" ([Klarreich 2018a],

[16] This isn't to suggest that philosophers should mechanically rubber-stamp any opinion a mathematician expresses in print. Experts in every field make mistakes and throwaway comments; taking mathematical practice seriously also means exercising discretion in choosing, reading and interpreting potential sources of evidence. But it's nevertheless true that the relevant specialists are better positioned than most philosophers to judge what qualifies as a source of mathematical understanding. When the best-informed and most thoughtful experts make such judgments deliberately, repeatedly and for coherent reasons, taking their word for it is the appropriate default.

[17] One such reason might be that the experts disagree among themselves. In that case, philosophy can play a useful role by investigating the source and nature of the disagreement. For a mathematical case study, see [D'Alessandro 2020]. See also the previous footnote for further elaboration of this epistemological stance.

25)—indeed, a distortion—of its target, the prime number sequence. This is interesting for a number of reasons, most obviously because it shows that mathematicians, like empirical scientists, make serious use of unrealistic models and rely on them to gain understanding. I defer discussion of further philosophical consequences to §4 below, after introducing my second example in the next section.

3 The function field model of the integers

This section discusses so-called *dyadic models* of linear structures, and in particular the model of the integers as polynomials over a finite field. As will become clear, this is a further example of an unrealistic model in widespread use as a source of mathematical understanding. This second case also bears consequentially on the questions about modeling mentioned at the start of the paper.

Dyadic models in mathematics take on a variety of forms depending on the settings in which they're deployed, which range from differential equations and harmonic analysis to combinatorics and number theory. But a common motivation of models in the family is to avoid the "spillover" between scaled exhibited by the integers, real numbers, cyclic groups and other linearly structured sets.

Here I'll focus on the integers. One manifestation of the spillover phenomenon in this domain is the need to carry digits when adding numbers together. In the sum $28 + 75 = 103$, for example, the addition of 8 and 5 in the units place spills over to affect the values in the tens and hundreds places. This kind of interaction between fine and coarse scales can be inconvenient. For instance, when adding many integers together, an accumulation of tiny ("fine-scale") errors can significantly distort the final ("coarse-scale") result.

The most common dyadic model of the integers is the ring of polynomials $F[t]$ over a finite field F. This is known as the *function field model* of the integers.[18] The elements of $F[t]$ are polynomials with coefficients from F. The role of positive integers in the model is played by monic polynomials, i.e. polynomials of the form $t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$ with leading coefficient 1.

(Some relevant definitions: a *field* is an algebraic structure with commutative addition and multiplication operations in which every nonzero element has both an additive inverse and a multiplicative inverse. The real numbers are a familiar example. A *finite field* is a field with finitely many elements. In fact, a finite field always has p^n elements, with p prime and $n \geq 1$. The simplest examples are the fields $\mathbb{F}_p$, whose elements are $\{0, 1, 2, \ldots, p-1\}$. Addition in $\mathbb{F}_p$ works like addition modulo p. For example, in the field with seven elements $\mathbb{F}_7$, we have $5 + 6 = 11 \pmod 7 = 4$.)

In the function field model, arithmetical operations on fine-scale terms don't affect the values of coarse-scale terms, and vice versa. Compare adding $\left(t^2 + 2t + 5\right) + (t + 6)$ in $\mathbb{F}_7[t]$, for example, with the analogous $125 + 16$ in $\mathbb{Z}$.[19] The latter exhibits spillover, since the units-place sum $(5 + 6 = 11)$ contributes 1 to the tens-place result. But not so in the model. In $\mathbb{F}_7[t]$, as noted above, $5 + 6$ equals 4, and so the sum $\left(t^2 + 2t + 5\right) + (t + 6) = t^2 + 3t + 4$ is spillover-free.

Another important feature of the function field model is that its "integers", being polynomials, have

[18] The terminology here is standard but a bit confusing. The ring of polynomials $F[t]$ is not itself a function field, as it isn't a field at all. The model's namesake is rather the rational function field $F(t)$, consisting of polynomials with coefficients in F as well as their multiplicative inverses. Since it contains fractions, the function field $F(t)$ is most naturally viewed as a model for the rational numbers $\mathbb{Q}$. But the name "function field model" (or "function field analogy") is generally applied to all models in this family, including $F[t]$ as a model for $\mathbb{Z}$.

[19] This example shouldn't be taken too literally. The function field model generally deals with an arbitrary finite field F rather than a specific one like $\mathbb{F}_7$, and it generally doesn't assign specific polynomials to serve as the representatives of specific integers. The point is just to compare spillover in $\mathbb{Z}$ with its absence in function fields.

nontrivial derivatives. (The derivative of an ordinary integer is of course always 0.) I'll say more about why this is useful below.

$F[t]$ might seem at first glance like a strange model for the integers. Why would representing numbers as polynomials seem appropriate, and how might it be useful? As Michael Rosen writes (in his textbook on the subject, *Number Theory in Function Fields*):

> Early on... it was noticed that $\mathbb{Z}$ has many properties in common [with $F[t]$]... Both rings are principal ideal domains... both rings have infinitely many prime elements, and both rings have finitely many units. Thus, one is led to suspect that many results which hold for $\mathbb{Z}$ have analogues [in $F[t]$]. This is indeed the case. ([Rosen 2002], vii)

In other words, the two structures have importantly similar algebraic properties. Hence "number theory in $F[t]$" should (and does) resemble ordinary number theory to a significant degree. The study of function fields as a source of number-theoretic insight goes back at least to Dedekind and Weber's 1882 paper "Theory of Algebraic Functions of One Variable" ([Dedekind & Weber 2012]). (As the translator John Stillwell notes, "the paper revealed the deep analogy between number fields and function fields—an analogy that continues to benefit both number theory and geometry today" (vii).)

Indeed, the function field model has proven fruitful in many ways. Several of its uses resemble those of Cramér's model of the primes: increasing our confidence in independent hypotheses, suggesting new conjectures and offering novel methods of proof. The function field model is also unique in at least one important way: it suggests to many mathematicians that there ought to exist a novel kind of object, the "field with one element", to complete certain aspects of the correspondence between $F[t]$ and $\mathbb{Z}$. If efforts to make sense of this notion prove successful, momentous developments in algebra, number theory and geometry are expected to ensue.

Let me fill in some details, starting with the first items mentioned above. A famous open problem in number theory is the *abc* conjecture of Oesterlé and Masser. The conjecture is roughly as follows: Let a, b, c be relatively prime integers such that $a + b = c$, and let D denote the product of the distinct prime factors of abc. Then c is significantly bigger than D in only finitely many cases. A large number of other major conjectures are known to be true conditional on *abc*; it would also yield a simple proof of Fermat's Last Theorem. So *abc* is of great interest to number theorists.[20]

It's therefore significant that the counterpart of the *abc* conjecture, known as the Mason-Stothers theorem, is known to hold in the function field model. (The Mason-Stothers theorem concerns relatively prime polynomials $a(t), b(t), c(t)$, not all constant and with $a + b = c$. Where D is the degree of the product of the distinct irreducible factors of a, b and c, the theorem asserts that D is significantly bigger than the maximum among the degrees of a, b and c.) The Mason-Stothers theorem has an elementary proof in $F[t]$ based on taking derivatives of a, b and c ([Snyder 2000])—a trick unavailable for proving *abc* in the integers. This is an example of the aforementioned usefulness of derivatives in the function field model.

The *abc* conjecture is just one hypothesis on which the function field model sheds light. As Rudnick notes, a variety of classic problems "which are currently viewed as intractable over the integers, have recently been addressed in the function field context... and the resulting theorems can be used to check existing

[20] Notoriously, the Japanese mathematician Shinichi Mochizuki has claimed to have proven *abc* since 2012, but the consensus among number theorists is that the proof is unconvincing and the conjecture remains open. See [Dutilh Novaes 2013] for a philosophical analysis and [Klarreich 2018b] for an account of recent developments, including Peter Scholze and Jakob Stix's engagement with Mochizuki and their conclusion in 2018 that his proof contains an unfixable gap.

conjectures over the integers, and to generate new ones" ([Rudnick 2014], 443). (Rudnick's paper discusses five such problems.)

In some cases, results in the function field model can be used to directly prove the corresponding statements in ordinary number theory. One example is the Ax-Kochen theorem, an important result about the zeroes of certain polynomials over the p-adic numbers. In the standard proof of Ax-Kochen, the first step is to show that the analogous claim holds in the function field model. Using the "transfer principle" technique from model theory, it's then possible to import the function field statement back to the original p-adic context, thus proving the theorem.

A final way in which the function field model has advanced number theory is by motivating research around the notional "field with one element" $\mathbb{F}_1$. In standard algebra, a field with one element isn't possible, since fields by definition have an additive identity 0 and a multiplicative identity 1 such that $0 \neq 1$. So the quest for $\mathbb{F}_1$ can be seen as an exercise in conceptual engineering: the task is to build a coherent theory in which an $\mathbb{F}_1$-like object exists and has certain desirable properties.[21]

Impetus for this quest comes from several sources, a major one being the success and promise of the function field model. Work on $\mathbb{F}_1$ is technical, and describing it in adequate detail would be unduly lengthy, so I'll confine myself to a brief sketch.

The starting point is André Weil's 1948 proof of the Riemann Hypothesis for function fields ([Weil 1948]); as Oliver Lorscheid writes, "[t]he analogies between number fields and function fields led to the hope that one can mimic these methods for $\mathbb{Q}$ and approach the [standard] Riemann hypothesis" ([Lorscheid 2018], 94). (Here $\mathbb{Q}$ denotes the set of rational numbers.)

Weil's proof starts with a "global" function field F, that is, a finite field extension of the rational function field $\mathbb{F}_p(t)$. (See the first footnote of this section for more on $\mathbb{F}_p(t)$.) It turns out that F can be interpreted as the function field of a *curve* C over the base field $\mathbb{F}_p$. One can then define a zeta function ζ_C for this curve and use the tools of algebraic geometry to prove the analogue of the Riemann Hypothesis for ζ_C. In particular, Weil's proof counts the number of intersection points of C with a "twisted" version of itself inside the fiber product $C \times_{\mathrm{Spec}\mathbb{F}_p} C$.[22]

Many mathematicians have hoped, as per Lorscheid's remarks above, to get a proof of the standard RH by translating Weil's proof from the function field model to the integers. The first step in this process is to identify the curve C such that ζ_C is the ordinary Riemann zeta function featuring in RH. This curve turns out to be the spectrum of the integers, $\mathrm{Spec}\mathbb{Z}$. The remaining task is to specify the base field over which $\mathrm{Spec}\mathbb{Z}$ is to be viewed as a function field. It's at this point that the field with one element enters the scene:

> The analogy between number fields and function fields finds a basic limitation with the lack of a ground field. One says that $\mathrm{Spec}\mathbb{Z}$ is... like a (complete) curve; but over which field? In particular, one would dream of having an object like
>
> $$\mathrm{Spec}\mathbb{Z} \times_{\mathrm{Spec}\mathbb{F}_1} \mathrm{Spec}\mathbb{Z},$$

The field with one element $\mathbb{F}_1$, then, should be an object over which it makes sense to view $\mathrm{Spec}\mathbb{Z}$ as a curve.

[21] See [Tanswell 2018] for discussion of conceptual engineering in mathematics.

[22] $\mathrm{Spec}R$, the spectrum of a commutative ring R, is the set of all prime ideals of R, often equipped with the Zariski topology.

Using $\mathbb{F}_1$ to prove the Riemann Hypothesis is just one motivation for its study. Broadly speaking, the idea of doing geometry with the integers over $\mathbb{F}_1$ "emerged from certain heuristics in combinatorics, number theory and homotopy theory that could not be explained in the framework of Grothendieck's scheme theory" [Lorscheid 2018], 83). Since scheme theory has served as the foundation for algebraic geometry for over half a century, a fully realized theory of the field with one element would necessitate a major rethinking of a large body of mathematics.

Although we still lack a definition of $\mathbb{F}_1$ that seems likely to yield a proof of RH, much exploratory theory-building has been done by an impressive collection of mathematicians: Jacques Tits (credited with first suggesting $\mathbb{F}_1$ in [Tits 1957]), Alain Connes, Caterina Consani, Yuri Manin and Christophe Soulé, to name a few particularly influential contributors. [Thas 2016] and [Lorscheid 2018] are a recent essay collection and survey paper, respectively. Even if RH remains elusive, work on $\mathbb{F}_1$ goes on, and has already produced a richer picture of the relationship between number theory and geometry.

My final claims in this section will come as no surprise: I want to argue that mathematicians have gained significant understanding from the function field model in spite of its unrealistic character. In support of the claim about understanding, the considerations from the last section apply here also. In particular, as with the Cramér model, the experts best positioned to assess the merits of the function field model describe it as a source of understanding. Here for instance is Tao. (Recall that the *dyadic models* are a family that includes the function field model, as noted at the beginning of this section.)

> In some areas [dyadic constructions] are an oversimplified and overly easy toy model; in other areas they get at the heart of the matter by providing a model in which all irrelevant technicalities are stripped away; and in yet other areas they are a crucial component in the analysis of the non-dyadic case. In all of these cases, though, it seems that the contribution that dyadic models provide in helping us understand the non-dyadic world is immense. ([Tao 2008], 68)

Finally, the function field model is an unrealistic representation of the integers. Most obviously and importantly, there's no natural notion of order on the elements of $F[t]$, so the function field model completely lacks the linear structure of $\mathbb{Z}$. This difference has far-reaching consequences. For instance, mathematical induction doesn't make sense in $F[t]$, whereas the availability of induction is often taken to be a characteristic property of the whole numbers. (Cf. Stewart and Tall's *Foundations of Mathematics*: "What is a number? ...The first step [in finding the answer] was to characterise natural numbers. It turned out that their most important defining feature wasn't counting, or arithmetic: it was the possibility of proving theorems using mathematical induction" ([Stewart & Tall 2015], 159).)

Additionally, the metric on $F[t]$ is non-Archimedean, meaning that the familiar triangle inequality $d(x,z) \leq d(x,y) + d(y,z)$ takes the stronger "ultrametric" form $d(x,z) \leq \max\{d(x,y), d(y,z)\}$. This implies, for instance, that all triangles in $F[t]^n$ are isosceles, that every point on the interior of a ball is its center, and that for any two intersecting balls one is contained inside the other. These properties are of course very much unlike those of $\mathbb{Z}^n$ equipped with the usual metric.

In spite of sharing some algebraic features, then, the function field model differs from the integers in fundamental ways. As with the Cramér model, these differences amount to more than elisions of small or unimportant detail. Nor is it possible to recover the key features of the integers by any straightforward process of de-idealization. $F[t]$ is an unrealistic model.

4 Morals: Understanding, Explanation, Counterfactuals

At the outset of the paper, I listed three pressing questions about unrealistic scientific models. The previous sections have shown that mathematicians engage in modeling, and that some widely used models in pure mathematics are unrealistic. So these cases are relevant to the questions at issue. What can we learn from them?

I've already argued that the Cramér model of the primes and the function field model of the integers are sources of understanding. So the first of the three questions—Can we gain genuine understanding from unrealistic models?—has an affirmative answer.

The use of unrealistic models in mathematics has gone unappreciated by philosophers, and cases like the ones I've described are noteworthy for that reason. But this answer isn't otherwise surprising. Various kinds of epistemically salutary unrealistic models have been well studied in recent years, and the mere fact of their existence no longer seems especially controversial. See for instance [Batterman & Rice 2014], [Bokulich 2011], [de Regt 2015], [Hindriks 2013], [Mäki 2009], [Morrison 2015], [Rice 2016], and the papers in *Synthese*'s recent collection "What to Make of Highly Unrealistic Models": [Boesch 2019], [Knuuttila & Koskinen 2020], [Papayannopoulos 2020], [van Eck & Wright 2020].

The remaining questions, though, are very much under debate. The mathematical cases I've described can help resolve both.

4.1 Explanation and understanding

First: Can unrealistic models explain, and is this how they generate understanding? I take no position here on the first part of the question, but the answer to the second part is "in general, no". Some unrealistic models help us understand phenomena for which they offer no explanation.

Cramér's model is a case in point. As I've argued, the model has improved number theorists' understanding of the distribution of the primes. But the model doesn't explain the primes' distribution. There are circumstantial, theoretical and commonsensical reasons to think this is so.

The circumstantial reason is the lack of evidence from mathematical practice. Mathematicians are often quite interested in the explanatory value of theorems, proofs, heuristics and other tools, and they tend to make their positive appraisals known.[23] This is especially true of widely used and frequently discussed pieces of mathematics like the Cramér model. If the model were explanatory, this fact would be of interest to the community of number theorists who have studied, worked with and instructed their students about it for decades. After consulting what must be a large percentage of the published literature on the model (as well as many less formal online discussions), however, I've encountered no such appraisals, either explicit or oblique.[24] If mathematicians consider the Cramér model explanatory, they've been uncharacteristically quiet about it for almost a hundred years.

The theoretical reason is that, insofar as we have anything like a general understanding of explanation (in mathematics or elsewhere), Cramér's model doesn't seem to fit the bill. A standard idea is that explanations require *dependence relations* of some sort, either ontic or counterfactual. On the former view, an explanans has to cause, ground, or otherwise metaphysically undergird its explanandum. On the latter view, what's required is counterfactual dependence—if the explanans had been different, the explanandum would have been too. (For defenses of these two views, see for instance [Ruben 1990] and [Reutlinger 2016] respectively.)

[23] For overviews of the role of explanation in mathematics, see [D'Alessandro 2019] or [Mancosu 2018].

[24] A reasonably large sample of this literature is cited in §2 above.

The distribution of the prime numbers evidently doesn't depend on the facts about the Cramér model, in either sense of "depend". Indeed, it's absurd to suggest that a randomly generated subset of the natural numbers might produce or give rise to any properties of the actual primes. (We regrettably lack a well-developed theory of the metaphysics of mathematical objects, but this ought to be beyond doubt if anything is.)

It's also highly implausible that the distribution of the primes counterfactually depends on the properties of the model. Again, philosophy has yet to reach a consensus about how to deal with mathematical counterpossibles (see [Baron et al. 2020] for a start). But the prevailing idea, following the Lewis-Stalnaker semantics for ordinary counterfactuals, is to somehow identify the (impossible) worlds closest to actuality where the antecedent is true and check whether the consequent also holds in those worlds.

In the case at issue, we're supposed to imagine that the Cramér model is different in some way—say, that the Goldbach conjecture is false in the model instead of true. At this world, some even natural number greater than 2 is no longer the sum of two Cramér primes. Is it also the case here that some even $n > 2$ isn't the sum of two ordinary primes? I see no reason to think so. The closest worlds at which Goldbach fails in the Cramér model are worlds at which it fails just barely—say, where exactly one even $n > 2$ isn't the sum of two Cramér primes. And the fact that the model falsifies Goldbach by the slimmest of margins seems to entail nothing at all about whether Goldbach holds in $\mathbb{N}$. To whatever extent (if any) the properties of the natural numbers counterfactually depend on the properties of the Cramér primes, the dependence surely isn't so extraordinarily sensitive.

There are, of course, other proposals about the nature of explanation. A final one worth mentioning is Kitcher's unifcationist theory ([Kitcher 1989]), which was explicitly intended to apply to explanations in pure mathematics. On Kitcher's approach, a proof counts as explanatory just in case it instantiates an argument pattern from the "explanatory store", that is, the set of argument patterns that most efficiently systematizes our knowledge in a given domain. Kitcher tends to argue that a given proof $\mathscr{P}$ is explanatory by comparing it to another proof of the same result—one which generalizes less readily or less widely than $\mathscr{P}$, or one that's more mired in the details of a special case than $\mathscr{P}$. (See §3.2 of [Kitcher 1989] for these mathematical examples.)

In a straightforward sense, the models I've discussed aren't even eligible to count as explanatory on Kitcher's view, because they usually don't let us directly prove things about their target systems. The kind of assurance they provide is heuristic and analogical rather than deductive.[25] Even setting this issue aside, the model-based approach is often decidedly closer to the purpose-built, single-use end of the inferential spectrum—much like the forms of reasoning Kitcher dismisses as unexplanatory.

For instance, the Cramér model is good at providing insights about the distribution of the primes. But it isn't derived from any grand general theory, and it suggests no unifying perspective that's expected to help with other kinds of problems. By contrast, if and when we manage to prove claims like the Riemann Hypothesis, these proofs are expected to flow from a highly fruitful theory with consequences for many parts of mathematics. The argument patterns obtained from this theory will be far stronger candidates for the explanatory store than any inferences associated with Cramér-style models. So unificationism provides no reason to judge the models explanatory either.

I lack space to review other accounts of explanation. But I believe the situation seems much the same

[25] While Kitcher is a self-avowed "deductive chauvinist", he has a story to tell about how seemingly statistical or probabilistic explanations can be accommodated within his framework. (See [Kitcher 1989], 448-459.) But inferences from the properties of models to the expected properties of their target systems don't appear to be statistical or probabilistic arguments, and it seems unlikely that Kitcher would consider such inferences potentially explanatory. (Thanks to an anonymous referee for prompting me to discuss Kitcher's view.)

on any plausible, mathematically applicable theory.[26]

There's perhaps a second theoretical reason to deny that the Cramér model is explanatory. As noted in §2, number theorists use a large family of random models to study the distribution of the primes, Cramér's being the original. Many of these models make incompatible assumptions. Some allow infinitely many even primes, while others allow none besides 2. Some generate the surrogate prime sequence by a completely different random procedure than Cramér's. And so on. The models in this menagerie validate many of the same basic claims (the Landau conjectures and the Riemann hypothesis, for instance), but they have little else in common apart from their use of various random methods. Which of these models are explanatory, if any are? Singling out one in particular is indefensible: no individual model is uniquely worthy of the title. But declaring that all are explanatory is equally problematic. Since the models have so little in common, in virtue of what shared feature do they count as giving the same explanation?

Finally, there's a commonsensical reason to deny that Cramér's model explains. To explain a phenomenon is to give the reason why it occurs or obtains. And Cramér's model doesn't do this. In response to the question "Why is it the case that Goldbach's conjecture holds?", for instance, one wouldn't accept as the reason "Because it holds in random models of the primes". The fact about the model might be (and probably is) a good reason to *believe* that Goldbach's conjecture is true, but it doesn't tell us *why* the conjecture is true. Someone who knew the relevant facts about the model wouldn't be confused or misguided for continuing to seek an explanation elsewhere. Such at least is my intuition.

Many of the same remarks apply to the function field model, and indeed we have further evidence from mathematical practice in this case. Here is Lorscheid:

> For a not yet systematically understood reason, many arithmetic laws have (conjectural) analogues for function fields and number fields. While in the function field case, these laws often have a conceptual explanation by means of a geometric interpretation, methods from algebraic geometry break down in the number field case. The mathematical area of $\mathbb{F}_1$-geometry can be understood as a program to develop a geometric language that allows us to transfer the geometric methods from function fields to number fields. ([Lorscheid 2014], 408-9)

Here Lorscheid is contrasting the situation in the function field model—where results like the Riemann hypothesis are not only known, but have been successfully explained via algebraic geometry—with the situation in ordinary number theory, where these results are believed to hold but where no corresponding explanation is yet available. Number theorists would like to find a geometric explanation for RH. Hence their interest in studying the field with one element. But Lorscheid is clear that the function field model, at least as it currently stands, doesn't do this explanatory work. The model helps us understand features of the natural numbers without explaining those features.

This conclusion is significant because it challenges a popular view in philosophy of science, according to which understanding requires (or perhaps just is) possession of an explanation. Defenses of such a view include [de Regt 2009], [Hannon 2019], [Khalifa 2012], [Strevens 2013] and [Trout 2007]. Strevens provides a clear statement of exactly the claim I deny: "An individual has scientific understanding of a phenomenon just in case they grasp a correct scientific explanation of that phenomenon" ([Strevens 2013], 510).

[26] An account that it might seem strange not to mention is Marc Lange's theory of mathematical explanation, defended in [Lange 2014]. Lange's theory, like Kitcher's, is about explanatory proofs, which the models under discussion generally don't provide. I find it even less clear whether or how Lange's view might apply to model-based inference.

What's the right way to think about understanding—and, in particular, the sort of understanding gained from unrealistic models—if not in terms of explanation? I'll have more to say about this below.

4.2 Counterfactual knowledge and understanding

Second: Are unrealistic models useful primarily because they impart counterfactual knowledge about their target systems? Is this how such models contribute to understanding? Again, the answer is "in general, no". The epistemic benefits gained from some unrealistic models have little to do with counterfactual knowledge.

Before I argue against this view, I want to briefly explain why it's seemed plausible to many authors, since it may not be as intuitive as the purported link between explanation and understanding. The Schelling model of housing segregation mentioned in §1 provides a good illustration. Recall that the model represents a city as a square grid, with each grid cell depicting a housing unit. A housing unit can either be empty or occupied by a single agent. Agents are split into two disjoint groups, say red and blue. Each agent prefers that a certain ratio R of the adjacent occupied squares are occupied by members of their own group; if the ratio falls below R at any time, the agent will then move to an empty unit where their preferences are satisfied. Schelling showed that the red and blue populations eventually segregate themselves even for relatively small values of R (roughly $R \geq 1/3$ if the two groups are equally sized).

It's obvious that the Schelling model doesn't (and isn't meant to) realistically represent the factors actually responsible for segregation. The model considers only one variable that's potentially relevant to housing choice, and everything about its treatment of that variable is highly idealized. Yet the Schelling model is often taken to have improved our understanding of segregation. How so? Perhaps by imparting *counterfactual knowledge*. While the model teaches us little about the causes of segregation in real cities, it plausibly does show how things would (and wouldn't) change if the world were different. For example, one might infer from the model that segregation is a robust phenomenon, likely to persist even in the absence of social and economic inequalities. And one could use this counterfactual knowledge to evaluate proposed interventions. A diversity course that convinced people to be comfortable living in at least 50% same-race neighborhoods, for example, wouldn't be a promising remedy for segregation.

This looks like a reasonable diagnosis of the usefulness of the Schelling model. Is it possible that all unrealistic models serve our epistemic purposes in the same way? Many philosophers have thought so. On this view—defended, for instance, in [Bokulich 2011], [Grimm 2011], [Hindriks 2013], [Lipton 2009], [Rice 2016], [Levy 2020], [Saatsi forthcoming]—unrealistic models contribute to understanding mainly by providing us with counterfactual knowledge about their target systems. Indeed, some of these authors identify understanding in general with counterfactual knowledge. Levy, for example, writes that "understanding something is having a representation of it that allows one to draw inferences about its counterfactual behavior. ...[O]ne understands [a target phenomenon] T when one can use one's representation of T to say what would happen to the target if this or that change were made to it" ([Levy 2020], 281-2).

Unfortunately, this appealing view doesn't fit the facts. The function field model of the integers, as we've seen, is an unrealistic model that confers understanding. But it doesn't do so by imparting counterfactual knowledge. When number theorists work with the function field model, they aren't seeking information about a scenario in which the properties of the integers have been somehow altered. Their interest isn't in what mathematics would be like if $\mathbb{Z}$ lacked its linear structure, numbers had nontrivial derivatives, and additive spillover didn't occur. Rather, what they seek (and have gained) is a suite of evidence and heuristics about the expected properties of $\mathbb{Z}$, ideas about how the relevant claims might or might not be proved, and

clues about the geometric structure undergirding the $\mathbb{Z}/F[t]$ analogy. These are insights about the actual integers that contain no hint of counterfactual content.

Let me be clear about the claims I'm not making. First, it's not my view that mathematical counterpossibles are inherently defective in some way. I see no problem with admitting that some such statements are meaningful and have substantive truth conditions—for example, "If 6 were prime, then it wouldn't be divisible by 3" seems true, while "If 6 were prime, then it wouldn't be divisible by 6" seems false. Second, it's not my view that counterpossibles are never of interest to mathematicians. Claims like "If the traveling salesman problem were solvable in polynomial time, then the clique problem would be too" are common and perfectly reasonable (cf. [Jenny 2018]).[27] Third, it's not my view that the function field model yields no counterfactual knowledge whatsoever. I suppose one can infer from the model, say, that if the integers had well-behaved derivatives, then the *abc* conjecture would be easy to prove. But this sort of fact is just an uninteresting instance of an obvious general principle: for any two things A and B, if A had some of B's properties, then some B-ish things would be true of A. Truths of this form aren't enlightening unless we have some reason to care about and take seriously the counterfactual scenario in question. And we don't in this case: mathematicians simply don't entertain the prospect of the integers acquiring the properties of $F[t]$.

In summary, then, I don't reject the intelligibility or potential usefulness of mathematical counterpossibles in general.[28] My claim is just that the function field model doesn't improve understanding by delivering knowledge of this sort. Its primary epistemic contributions take the form of information about the integers' actual properties.

One might think that this conclusion leaves us with a puzzle. If unrealistic models in science often seem to improve understanding by conferring counterfactual knowledge—as is plausibly the case with Schelling's model, for example—why is this not generally true of unrealistic models in mathematics? To pose the question another way, why have philosophers mistakenly identified one possible element or symptom of understanding with a general rule about gaining understanding from unrealistic models?

One reason, I think, is that philosophers of science in our Woodwardian era have focused overmuch on control, manipulation, interventions and difference-makers—factors closely associated with counterfactuals, and often analyzed within a counterfactual framework. These factors are important in empirical science, and gaining knowledge about them can indeed contribute to understanding. But they aren't the only game in town. Mathematicians, for instance, care a great deal about understanding, but have little use for manipulationist machinery (since they aren't in a position to perform interventions and observe the results[29]). Instead they favor models that offer other kinds of goods—confirmation, predictions, heuristics, analogies, proof ideas, plausibility checks, hints at deeper structure. These sources of understanding exist in empirical science too, of course. And they're no less valuable there, even if philosophers are prone to neglect them. So we don't need to accept a disjunctive picture, according to which unrealistic models in science and mathematics contribute to understanding in fundamentally different ways. Rather, they do so in mostly similar ways, except that counterfactual knowledge associated with control and manipulability plays a larger role in (some parts of) empirical science.

Having considered and rejected two accounts of model-based understanding, it's natural to ask what positive picture suggests itself in their stead. This isn't the place to mount a defense of a novel theory

[27] These are counterfactuals assuming that $P \neq NP$.

[28] For more on the uses of counterpossibles in science, see [Tan 2019] and [McLoone 2020], as well as the paper of Jenny cited above.

[29] And more broadly, of course, because causal reasoning is generally inapplicable in pure mathematics.

(or to campaign at length on behalf of an existing one). Broadly speaking, however, accounts that link understanding to cognitive systematization look more promising than those focused on possession of a specific type of knowledge or ability. The understanding gained from unrealistic models often has the character of a broad-spectrum improvement to a variety of epistemic states (belief, credence, expectation, attention, inquiry) and cognitive functions (reasoning, intuition, similarity-detection, problem-solving). Trying to single out any one of these contributions as necessary or sufficient for understanding strikes me as an unpromising project. But a theory that takes the whole package as primary ought to do better. I think, for example, that the account of [Kelp 2015] is in the right ballpark, although its exclusive focus on knowledge and explanation/justification relations may be a weakness. I lack the space to fully engage with Kelp's or other views here, however.

4.3 Mathematics as a special science

One final moral suggested by this discussion concerns the relationship between pure mathematics and philosophy of science. That relationship is rather tenuous at present. Even if most philosophers of science accept some sort of Quinean continuity thesis in principle, those who pay serious attention to the content and practice of contemporary mathematics are a rare breed in practice. Why is this? Probably at least in part because few philosophers of science are aware of these developments. And why aren't they aware? Even if they'd hesitate to say so, I suspect a widespread sense persists that pure math and empirical science are fundamentally dissimilar enterprises—concerned with different goals, about different kinds of things, making use of different methods of inquiry. On this picture, there's just not much reason for the two disciplines to intersect (except, every so often, in the context of wondering about unreasonable effectiveness and indispensability).

Perhaps this is beginning to change, ever so slightly. The recent surge of interest in noncausal explanation has led more philosophers to recognize the importance of explanation in pure mathematics, and to propose, partly on the basis of mathematical examples, theories of explanation that apply to math and empirical science alike. (See for example [Lange 2014], or the duo [Pincock 2015a] and [Pincock 2015b].)

This is an encouraging development, but the two disciplines have more to say to one another. Mathematics and empirical science are much more alike than is often supposed. Researchers in both domains share the same basic epistemic desiderata (knowledge, understanding, explanation, evidence acquisition, theory construction), they pursue these goals using many of the same techniques (including modeling and other non-deductive strategies), and in doing so they have to weigh similar values and confront similar problems. The differences between mathematics and empirical science, on the other hand, are real—but not obviously more so than the differences between psychology and theoretical physics, say, or economics and geology. The natural sciences already admit a wide variety of inferential methods, degrees of certainty and apriority, and entities of more or less exotic metaphysical status. While mathematics occupies a distinctive place on some of these scales, there's little reason to view it as completely sui generis.

Indeed, I believe it's both appropriate and enlightening to view pure mathematics as a special science on part with the rest. Doing so can raise questions, present problems and suggest solutions that wouldn't have been obvious otherwise.

5 Conclusion

This paper has argued for three main claims. First: that unrealistic models have important uses in pure mathematics, and their epistemic benefits include improving our understanding of their target phenomena.

Second: that the understanding gained from these models (and hence from unrealistic models in general) need not flow from explanations of the target phenomena. Third: that it need not flow from counterfactual knowledge either.

Future work on the philosophy of modeling can benefit from further examination of mathematical cases. Consider the metaphysics of models. One popular view holds that models are artifacts ([Thomasson 2020], [Thomson-Jones 2020]); a related view holds that models in general, or perhaps unrealistic models in particular, are fictional entities of some sort ([Bokulich 2011], [Frigg 2010], [Salis 2021]). It would seem to follow from such views that the polynomial ring $F[t]$ (say) is an artifact or a fiction. But $F[t]$ is also a piece of ordinary mathematics, whose ontological status is presumably the same as that of other mathematical objects. Artifactualism about models therefore seems to imply artifactualism about mathematics in general. Depending on one's metaphysical commitments, this may be either a welcome consequence or a convincing reductio of artifactualist views.

Some other important questions are broadly epistemological. For instance, there's a tradition of viewing (some) model-based inference as a kind of analogical reasoning ([Bartha 2009], [Hesse 1963]), and mathematical models may offer some unique data. Mathematics should also join the conversation about thought experiments and imagination in science ([Brown 2010], [Murphy 2022]), and the relationship between these activities and modeling practices ([Arfini 2006]).

These are just a few of the ways in which philosophy stands to benefit from taking mathematics seriously. The kingdom will prosper when the sequestered queen of the sciences is allowed to return to court.[30]

References

[Arfini 2006] Arfini, Selene. 2006. "Thought experiments as model-based abductions." In Lorenzo Magnani and Claudia Casadio (eds.), *Model Based Reasoning in Science and Technology: Logical, Epistemological, and Cognitive Issues*, New York: Springer-Verlag.

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

[Bartha 2009] Bartha, Paul. 2009. *By Parallel Reasoning: The Construction and Evaluation of Analogical Arguments*. New York: Oxford University Press.

[Batterman & Rice 2014] Batterman, Robert and Collin Rice. 2014. "Minimal model explanations." *Philosophy of Science* 81, 349-376.

[Boesch 2019] Boesch, Brandon. 2019. "Scientific representation and dissimilarity." *Synthese*, DOI: 10.1007/s11229-019-02417-0.

[Bokulich 2011] Bokulich, Alisa. 2011. "How scientific models can explain." *Synthese* 180, 33-45.

[Brown 2010] Brown, James Robert. 2010. *The Laboratory of the Mind: Thought Experiments in the Natural Sciences* (2nd edition). New York: Routledge.

[30] I'm grateful to all the people who read or heard versions of this paper. Many thanks to Alan Baker, Nicola Bonatti, Kenny Easwaran, Joachim Frans, Silvia Jonas, Colin McLarty, Alex Paseau, Fenner Tanswell, Karim Zahidi and Benjamin Zayton for their comments and encouragement. All of you are lovely. Likewise to audiences at the Munich Center for Mathematical Philosophy, Ghent University, the University of Turin, the 4th Scientific Understanding and Representation meeting at Fordham University, the 2022 German Society for Philosophy of Science conference, and the participants in John Burgess and Silvia De Toffoli's seminar on mathematical rigor at Princeton University in Fall 2021. Each one was a pleasure. Two anonymous referees for this journal provided much-appreciated suggestions. Finally, special thanks to Lauren Woomer for patiently listening to me give this talk on Zoom many times while we were stuck inside during the pandemic.

[Button & Walsh 2018] Button, Tim and Sean Walsh. 2018. *Philosophy and Model Theory*. New York: Oxford University Press.

[Cramér 1936] Cramér, Harald. 1936. "On the order of magnitude of the difference between consecutive prime numbers." *Acta Arithmetica* 2, 23-46.

[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. "Proving quadratic reciprocity: Explanation, disagreement, transparency and depth." *Synthese*, DOI: 0.1007/s11229-020-02591-6.

[Dedekind & Weber 2012] Dedekind, Richard and Heinrich Weber. 2012. *Theory of Algebraic Functions of One Variable*. Trans. by John Stillwell. Providence, Rhode Island: American Mathematical Society.

[Delarivière & Van Kerkhove 2021] Delarivière, Sven and Bart Van Kerkhove. 2021. "The mark of understanding: In defense of an ability account."

[de Regt 2009] de Regt, Henk. 2009. "The epistemic value of understanding." *Philosophy of Science* 76, 585-597.

[de Regt 2015] de Regt, Henk. 2015. "Scientific understanding: Truth or dare?" *Synthese* 192, 3781-3797.

[Dutilh Novaes 2013] Dutilh Novaes, Catarina. 2013. May 14. "What's wrong with Mochizuki's 'proof' of the ABC conjecture?" *M-Phi*. Online at <https://m-phi.blogspot.com/2013/05/whats-wrong-with-mochizukis-proof-of.html>. 172, 251-268.

[Frigg 2010] Frigg, Roman. 2010. "Models and fiction." *Synthese*

[Granville 1995a] Granville, Andrew. 1995. "Harald Cramér and the distribution of prime numbers." *Scandinavian Actuarial Journal* 1, 12-28.

[Granville 1995b] Granville, Andrew. 1995. "Unexpected irregularities in the distribution of prime numbers." In S.D. Chatterjee (ed.), *Proceedings of the International Congress of Mathematicians*. Basel: Birkhauser.

[Green 2005] Green, Ben. 2005. "Finite field models in additive combinatorics." In Bridget Webb (ed.), *Surveys in Combinatorics 2005*, Cambridge: Cambridge University Press.

[Grimm 2011] Grimm, Stephen. 2011. "Understanding." In Duncan Pritchard and Sven Berneker (eds.), *The Routledge Companion to Epistemology*, New York: Routledge.

[Hannon 2019] Hannon, Michael. 2019. *What's the Point of Knowledge? A Function-First Epistemology*. New York: Oxford University Press.

[Hawkins 1957] Hawkins, David. 1957. "The random sieve." *Mathematics Magazine* 31, 1-3.

[Hawkins & Briggs 1957] Hawkins, David and W. E. Briggs. 1957. "The lucky number theorem." *Mathematics Magazine* 31, 81-84.

[Hesse 1963] Hesse, Mary. 1963. *Models and Analogies in Science*. South Bend, Indiana: University of Notre Dame Press.

[Hindriks 2013] Hindriks, Frank. 2013. "Explanation, understanding, and unrealistic models." *Studies in History and Philosophy of Science* 44, 523-531.

[Ingham 1932] Ingham, A.E. 1932. *The Distribution of Prime Numbers*. Cambridge: Cambridge University Press.

[Jenny 2018] Jenny, Matthias. 2018. "Counterpossibles in science: The case of relative computability." *Noûs* 52, 530-560.

[Kelp 2015] Kelp, Christoph. 2015. "Understanding phenomena." *Synthese* 192, 3799-3816.

[Khalifa 2012] Khalifa, Kareem. 2012. "Inaugurating understanding or repackaging explanation?" *Philosophy of Science* 79, 15-37.

[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.

[Klarreich 2018a] Klarreich, Erica. 2018. "Mathematicians discover prime conspiracy." In Thomas Lin (ed.), *The Prime Number Conspiracy*, Cambridge: MIT Press.

[Klarreich 2018b] Klarreich, Erica. 2018, September 20. "Titans of mathematics clash over epic proof of ABC conjecture." *Quanta Magazine*. Online at <https://www.quantamagazine.org/titans-of-mathematics-clash-over-epic-proof-of-abc-conjecture-20180920/>.

[Knuuttila & Koskinen 2020] Knuuttila, Tarja and Rami Koskinen. 2020. "Synthetic fictions: Turning imagined biological systems into concrete ones." *Synthese*, DOI: s11229-020-02567-6.

[Koukoulopoulous 2019] Koukoulopoulous, Dimitris. 2019. *The Distribution of Prime Numbers*. Providence, Rhode Island: American Mathematical Society.

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

[Levy 2020] Levy, Arnon. 2020. "Metaphor and scientific explanation." In Arnon Levy and Peter Godfrey-Smith (eds.), *The Scientific Imagination: Philosophical and Psychological Perspectives*, New York: Oxford University Press.

[Lipton 2009] Lipton, Peter. 2009. "Understanding without explanation." In Henk W. de Regt, Sabina Leonelli, and Kai Eigner (eds.), *Scientific Understanding: Philosophical Perspectives*, Pittsburgh: University of Pittsburgh Press.

[Lorch & Ökten 2007] Lorch, John and Giray Ökten. 2007. "Primes and probability: The Hawkins random sieve." *Mathematics Magazine* 80, 112-119.

[Lorscheid 2014] Lorscheid, Oliver. 2014. "Blueprints—toward absolute arithmetic?" *Journal of Number Theory* 144, 408-421.

[Lorscheid 2018] Lorscheid, Oliver. 2018. "$\mathbb{F}_1$ for everyone." *Jahresbericht der Deutschen Mathematiker-Vereinigung* 120, 83-116.

[Lozano-Robledo 2020] Lozano-Robledo, Álvaro. 2020. "A probablistic model for the distribution of ranks of elliptic curves over $\mathbb{Q}$." *Journal of Number Theory*, DOI: 10.1016/j.jnt.2020.05.022.

[Mäki 2009] Mäki, Uskali. 2009. "Realistic realism about unrealistic models." In Harold Kincaid and Don Ross (eds.), *The Oxford Handbook of Philosophy of Economics*, New York: Oxford University Press.

[McLoone 2020] McLoone, Brian. 2020. "Calculus and counterpossibles in science." *Synthese*, DOI: 10.1007/s11229-020-02855-1.

[Mancosu 2018] Mancosu, Paolo. 2018. "Explanation in mathematics." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2018 edition), online at <https://plato.stanford.edu/archives/sum2018/entries/mathematics-explanation/>.

[Montgomery & Vaughan 2007] Montgomery, Hugh L. and Robert C. Vaughan. 2007. *Multiplicative Number Theory I: Classical Theory*. New York: Cambridge University Press.

[Morrison 2015] Morrison, Margaret. 2015. *Reconstructing Reality: Models, Mathematics, and Simulations*. New York: Oxford University Press.

[Murphy 2022] Murphy, Alice. 2022. "Imagination in science." *Philosophy Compass* 17, 1-12, DOI: 10.1111/phc3.12836.

[Papayannopoulos 2020] Papayannopoulos, Philippos. 2020. "Unrealistic models for realistic computations: how idealisations help represent mathematical structures and found scientific computing." *Synthese*, DOI: 10.1007/s11229-020-02654-8.

[Patterson 1988] Patterson, S.J. 1988. *An Introduction to the Theory of the Riemann Zeta-function*. Cambridge: Cambridge University Press.

[Pincock 2015a] Pincock, Christopher. 2015. "Abstract explanations in science." *British Journal for the Philosophy of Science* 66, 857-882.

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

[Pintz 2007] Pintz, János. 2007. "Cramér vs. Cramér: On Cramér's probabilistic model for primes." *Functiones et Approximatio* 37, 361-376.

[Pólya 1959] , George. 1959. "Heuristic reasoning in the theory of numbers." *American Mathematical Monthly* 66, 375-384.

[Potochnik 2017] Potochnik, Angela. 2017. *Idealization and the Aims of Science*. Chicago: University of Chicago Press.

[Reutlinger 2016] Reutlinger, Alexander. 2016. "Is there a monist theory of causal and non-causal explanations? The counterfactual theory of scientific explanation." *Philosophy of Science* 83, 733-745.

[Reutlinger et al. 2018] Reutlinger, Alexander, Dominik Hangleiter and Stephan Hartmann. 2018. "Understanding (with) toy models." *British Journal for the Philosophy of Science* 69, 1069-1099.

[Rice 2016] Rice, Collin C. 2016. "Factive scientific understanding without accurate representation." *Biology and Philosophy* 31, 81-102.

[Rosen 2002] Rosen, Michael. 2002. *Number Theory in Function Fields*. New York: Springer-Verlag.

[Ruben 1990] Ruben, David-Hillel. 1990. *Explaining Explanation*. New York: Routledge.

[Rudnick 2014] Rudnick, Zeev. 2014. "Some problems in analytic number theory for polynomials over a finite field." In Sun Young Jang, Young Rock Kim, Dae-Woong Lee and Ikkwon Yie (eds.), *Proceedings of the International Congress of Mathematicians, Seoul 2014, Volume II: Invited Lectures*, Seoul: Kyung Moon SA.

[Saatsi forthcoming] Saatsi, Juha. Forthcoming. "Realism and explanatory perspectivism." In Michela Massimi and C.D. McCoy (eds.), *Understanding Perspectivism: Scientific Challenges and Methodological Prospects*, New York: Routledge.

[Salis 2021] Salis, Fiora. 2021. "The new fiction view of models." *British Journal for the Philosophy of Science* 72, 717-742.

[Snyder 2000] Snyder, Noah. 2000. "An alternate proof of Mason's theorem." *Elemente der Mathematik* 55, 93-94.

[Soulé 1999] Soulé, Christoph. 1999. "On the field with one element." Lecture notes from the Arbeitstagung 1999 of the Max Planck Institute for Mathematics, available online at <https://www.ihes.fr/~soule/f1-soule.pdf>.

[Soundararajan 2007a] Soundararajan, Kannan. 2007. "The distribution of prime numbers." In Andrew Granville and Zeév Rudnick (eds.), *Equidistribution in Number Theory: An Introduction*, Dordrecht: Springer.

[Soundararajan 2007b] Soundararajan, Kannan. 2007. "Small gaps between prime numbers: The work of Goldston-Pintz-Yıldırım." *Bulletin (New Series) of the American Mathematical Society* 44, 1-18.

[Stewart & Tall 2015] Stewart, Ian and David Tall. 2015. *The Foundations of Mathematics*, second edition. New York: Oxford University Press.

[Strevens 2013] Strevens, Michael. 2013. "No understanding without explanation." *Studies in History and Philosophy of Science* 44, 510-515.

[Tan 2019] Tan, Peter. 2019. "Counterpossible non-vacuity in scientific practice." *Journal of Philosophy* 116, 32-60.

[Tanswell 2018] Tanswell, Fenner Stanley. 2018. "Conceptual engineering for mathematical concepts." *Inquiry* 61, 881-913.

[Tao 2008] Tao, Terence. 2008. *Structure and Randomness*. Providence, RI: American Mathematical Society.

[Tao 2015] Tao, Terence. 2015, January 4. "Probabilistic models and heuristics for the primes." *What's New*. Online at <https://terrytao.wordpress.com/2015/01/04/254a-supplement-4-probabilistic-models-and-heuristics-for-the-primes-optional/>

[Tenenbaum & France 2000] Tenenbaum, Gérald and Michel Mendès France. 2000. *Prime Numbers and Their Distribution*. Trans. by Philip G. Spain. Providence, RI: American Mathematical Society.

[Thas 2016] Thas, Koen (ed.). 2016. *Absolute Arithmetic and* $\mathbb{F}_1$*-Geometry*. Zurich: European Mathematical Society.

[Thomasson 2020] Thomasson, Amie. 2020. "If models were fictions, then what would they be?" In In Arnon Levy and Peter Godfrey-Smith (eds.), *The Scientific Imagination: Philosophical and Psychological Perspectives*, New York: Oxford University Press.

[Thomson-Jones 2020] Thomson-Jones, Martin. 2020. "Realism about missing systems." In Arnon Levy and Peter Godfrey-Smith (eds.), *The Scientific Imagination: Philosophical and Psychological Perspectives*, New York: Oxford University Press.

[Tits 1957] Tits, Jacques. 1957. "Sur les analogues algébriques des groupes semi-simples complexes." *Colloque d'algèbre supérieure, tenu à Bruxelles du 19 au 22 décembre 1956, Centre Belge de Recherches Mathématiques Établissements Ceuterick, Louvain.* Paris: Librairie Gauthier-Villars.

[Trout 2007] Trout, J.D. 2007. "The psychology of explanation." *Philosophy Compass* 2, 564-596.

[van der Poorten 1996] van der Poorten, Alf. 1996. *Notes on Fermat's Last Theorem.* New York: Wiley.

[van Eck & Wright 2020] van Eck, Dingmar and Cory Wright. 2020. "Mechanist idealisation in systems biology." *Synthese*, DOI: 10.1007/s11229-020-02816-8.

[Weil 1948] Weil, André. 1948. *Sur les Courbes Algébriques et les Variétés qui s'en Déduisent.* Paris: Hermann.

[Wilkenfeld 2019] Wilkenfeld, Daniel. 2019. "Understanding as compression." *Philosophical Studies* 176, 2807-2831.