Annals of Mathematics and Philosophy

$M \times \Phi$

$M \times \Phi$ vol. 1.2 © 2023

Volume I, number 2, 2023

Spartacus IDH



Annals of Mathematics and Philosophy

$M \times \Phi$

Volume I, number 2

© 2023

Spartacus Graduate Journals

Annals of Mathematics and Philosophy

$M \times \Phi$

Volume I, number 2

2023

PUBLICATION DIRECTOR

Victor Rabiet, École Normale Supérieure

EDITORS-IN-CHIEF

Jean-Pierre Marquis, Université de Montréal
Frédéric Patras, CNRS – Université Côte d'Azur

EDITORS

Paola Cantù, CNRS – Aix-Marseille Université
Jessica Carter, Aarhus University
Carlo Cellucci, Università Roma 1
Gabriela Crocco, Aix-Marseille Université
Moritz Epple, Goethe-Universität Frankfurt am Main
José Ferreiros, Universidad de Sevilla
Fausto Fraisopi, Albert-Ludwigs-Universität Freiburg
Gila Hanna, University of Toronto
Frédéric Jaëck, Aix-Marseille Université
Brendan Larvor, University of Hertfordshire
Giuseppe Longo, CNRS – École Normale Supérieure
Mario Piazza, Scuola Normale Superiore di Pisa
Alessandro Sarti, CNRS – École des hautes études en sciences sociales
Andrea Sereni, Scuola Universitaria Superiore di Pavia
Hourya Sinaceur, CNRS – Université de Paris
Jean-Jacques Szczeciniarz, Université de Paris
Fernando Zalamea, Universidad Nacional de Colombia

Spartacus IDH

© 2023

$M \times \Phi$ *vol. I.2* © 2023

ISBN : 978-2-36693-132-7
© Spartacus-idh, Paris 2023

Contents

Articles

Clint Enns — Much Badiou about Nothing. Productive Misreadings of Mathematical Ideas and Isiah Medina's *88:88* **1**

Andrew Granville — Accepted proofs: objective truth, or culturally robust? **25**

Joel David Hamkins — Fregean abstraction in Zermelo-Fraenkel set theory: a deflationary account **91**

Michael Harris — Virtues of Priority **127**

Carlos Lobo — Rota's variation on the eidetic identity **155**

Tribute

Gonzalo E. Reyes — Homage to Francis William Lawvere (1937–2023) **189**

Review Essay

Palle Yourgrau — Does Reason Have Limits? **201**

M×Φ vol. I.2 © 2023

Annals of Mathematics and Philosophy
***vol.* I, *no.* 2**, 2023

$M \times \Phi$

Much Badiou about Nothing:

Productive Misreadings of Mathematical Ideas and Isiah Medina's *88:88*

Clint Enns

© 2023

vol. I.2

$M \times \Phi$

Abstract. Kepler's model for the solar system was constructed following the formal beauty of the Platonic Solids. Although not scientifically accurate, this model prompted many scientific advancements given that Kepler attempted to justify it using empirical evidence and, until that point in history, scientific research was mainly descriptive. In this article, I will argue that *productive misreadings* of mathematical ideas have the potential to lead to original concepts, and are not necessarily detrimental to the social sciences as physicist Alan Sokal (1998) and others contend. In particular, I will argue that French philosopher Alain Badiou's claim that "mathematics is ontology" is based on a mistake analogous to Kepler's — namely, that Badiou based the underlying structure for his ontological claims on set theory due to its perceived beauty. In spite of this, it will be shown that Badiou's conceptualization allows for novel ontological insight.

By conceptualizing film editing through mathematical conceptions of the infinite, experimental filmmaker Isiah Medina uses mathematics as one of the inspirations for his art. Medina treats philosophy, mathematics, and cinema as a form of free association, a methodology of philosophic freestyle, or academic hip hop resulting in numerous philosophical and cinematic insights. The influence of Badiou's ontological framework on Medina cannot be overstated where even the title of his debut feature *88:88* (2015) functions as metaphor for Badiou's

conception of the *void*. In addition to providing a detail analysis of Badiou's ontological framework, I will also examine Medina's appropriation of Badiou's philosophy and will discuss some of the other ways he applies mathematical concepts in his film *88:88*.

Acknowledgements. The research for this article came while writing my PhD dissertation titled "The Poetry of Logical Ideas: Towards a Mathematical Genealogy of Media Art," for the Department of Cinema and Media Studies at York University under the supervision of Michael Zryd. Thank you to my readers, Janine Marchessault and Mark-David Hosale, and to my examination committee members, Bruce Jenkins, Mike Zabrocki and Leslie Korrick, for their constructive criticisms, comments, and feedback.

I would like to acknowledge that this dissertation would not have been possible without the generous support of the Joseph-Armand Bombardier Canadian Graduate Scholarship through the Social Sciences and Humanities Research Council of Canada (SSHRC), and the Susan Mann Dissertation Scholarship.

$M\times\Phi$ **vol. I.2** © 2023

§ — Beautiful Monsters: Productive Misreadings of Mathematics.

I use the term *productive misreading* to describe a misinterpretation of a concept or an idea in a way that generates new forms of knowledge. While misreadings have the potential to expand knowledge, ideas from the sciences have often been appropriated by those in the humanities to justify their theories. A productive misreading is not a form of justification, it is a way of generating new insight. Scientific concepts should only be used to support new theories in the humanities if the author can justify that their claim satisfies the criteria of the original concept. For instance, the Axiom of Choice should not be used to justify a position regarding freedom of choice, an argument Sokal satirically presents in his faux paper "Transgressing the Boundaries." The axioms of set theory, which include the Axiom of Choice, were obviously not constructed to deal with issues concerning pregnancy; however, they may provide new ways of thinking through these types of social issues.

Given that all communication is, at least to some degree, imperfect some have argued that every interpretation is a form of misinterpretation. For instance, philosopher Slavoj Žižek reinforces this idea in his book *Organs Without Bodies*, claiming that the entire history of philosophy is based on productive misreadings. He writes,

> As Alain Badiou put it, philosophy is inherently *axiomatic*, the consequent deploying of a fundamental insight. Hence, all great 'dialogues' in the history of philosophy were so many cases of misunderstanding: Aristotle misunderstood Plato, Thomas Aquinas misunderstood Aristotle, Hegel misunderstood Kant and Schelling, Marx misunderstood Hegel, Nietzsche misunderstood Christ, Heidegger misunderstood Hegel [...] Precisely when one philosopher exerted a key influence upon another, this influence was without exception grounded in a productive misreading — did not the entirety of analytic philosophy emerge from misreading the early Wittgenstein? (Žižek, 2004, p. ix)

In other words, Žižek argues that "strong" philosophers develop out of productive misreadings of prior philosophers. For Žižek, misreadings are due to the fact that dialogue (as opposed to logical reasoning deduced from axioms) is imperfect. For instance, he suggests many philosophers hold different positions and cannot come to agreement precisely due to the fact that they misinterpret each other and often are "speaking different, totally incompatible, languages, with no shared ground between them" (Žižek, 2004, p. 47).

Žižek's idea for the productive misreading is borrowed from Deleuze, who believed that the history of philosophy is not actually about dialogue, nor a search for truth, but a series of productive misreadings that have led to the production of "monstrous" new ideas. Deleuze uses the act of buggery as a metaphor for productive misinterpretation:

> I suppose the main way I coped with it at the time [an aversion to the academic history of philosophy, which Deleuze saw as repressive] was to see the history of philosophy as a sort of buggery or (it comes to the same thing) immaculate conception. I saw myself as taking an author from behind and giving him a child that would

$M \times \Phi$ vol. I.2 © 2023

> be his own offspring, yet monstrous. It was really important for it to be his own child, because the author had to actually say all I had him saying. But the child was bound to be monstrous too, because it resulted from all sorts of shifting, slipping, dislocations, and hidden emissions that I really enjoyed. (Deleuze, 1995, p. 6)

In other words, Deleuze is suggesting productive misinterpretation gives rise to ideas that are mutated and deformed versions of ideas held by previous philosophers, born out of the impossibility of remaining true to another author in spite of the reader's best intentions. By re-imagining the history of philosophy as a form of buggery (combined with immaculate conception), Deleuze is queering it while being blasphemous; an impressive manoeuvre despite the fact that it inherently reinforces a form of misogyny in which one male genius inseminates other.

If we attempt to accurately interpret Deleuze, a gesture which he might argue is impossible, he argues that the monstrous offspring of ideas mutated occur by immaculate conception, implying that these ideas arrive without any real fixed explanation. In spite of the best intentions of the reader to accurately understand the original author, new deformed ideas miraculously arrive. In contrast, it is possible to view these misinterpretations not simply as monstrous reproductions or inaccurate interpretations, but as informed interventions that constitute an attempt to push beyond the original text or context. For instance, mathematical and scientific ideas are often appropriated by artists and theorists, and forced into a new (and often unintended) context allowing for the generation of new ideas outside of their original sanctioned environment. Of course, this form of experimentation is not always successful. However, when it is, it potentially has the power to generate novel ways of thinking about the universe — simply recall Kepler's model of the solar system based on the Platonic solids.

§ — Zermelo-Fraenkel Mysticism: Rationalizing Ontology.

Badiou's magnum opus, *L'Être et Evénement* [*Being and Event*], was originally published in 1988. While Badiou has engaged with a wide range of mathematical concepts and ideas throughout his career, *Being and Event* — arguably the foundational text of his

$M\times\Phi$ **vol. I.2** © 2023

œuvre — presents Zermelo-Fraenkel set theory (ZFC) as more than merely the foundations of modern mathematics. The book presents a discourse on the science of being *qua* being — the discourse of ontology — where Badiou defines an ontology as "a presentation of presentation" (Badiou, 1988/2005, p. 28). The core principle of Badiou's philosophical framework is fundamentally the equivalence between mathematics and philosophy, a proposition succinctly summarized by Badiou in his most well-known slogan: "mathematics *is* ontology" (Badiou, 1988/2005, p. 4). Philosopher Alex Ling reinforces the mathematical foundations of Badiou's entire philosophical project:

> Fundamental to Badiou's later philosophy is his declaration in *Being and Event* that 'mathematics is ontology.' [...] One could even argue — such is the rigour with which Badiou constructs *Being and Event* — that those who reject Badiou's core philosophy do so foremost because they reject his initial thesis on the equivalence of mathematics and ontology. For if this thesis is unfounded, so too is Badiou's entire philosophy.
>
> (Ling, 2010, p. 48)

$M\times\Phi$ **vol. I.2** © 2023

While Ling succinctly describes Badiou's core philosophy, I reject his claim that "if this thesis is unfounded, so too is Badiou's entire philosophy." Despite arguing that Badiou's methodology is cause for concern, I am not entirely convinced that his philosophical arguments *necessarily* need a mathematical basis in order to be considered valid, nor do they need to directly correspond to the mathematics following the rigorous presentation in *Being and Event*. As suggested by philosopher A. J. Bartlett, some philosophers have taken Badiou's declaration that mathematics is ontology "too literally" (Bartlett, 2014, p. 307).[(1)]

The framework of *Being and Event* is not entirely novel, and its basis can be found in a form of mathematical reductionism, a branch of mathematics that argues numbers are sets, or that set theory is the ontological basis of mathematics. For Badiou, mathematics has been a sub-genre of philosophy for too long, reduced to the use of mathematical logic or the area of specialization known

[(1)] In particular, Bartlett names philosophers Bruno Bosteels and Peter Hallward. He argues that Badiou's claim that mathematics is ontology "is taken too literally by Bosteels (and Hallward) in the sense that they suppose that a philosophy exists of Badiou ultimately untouched by this ontology, the 'science of being qua being.' "

as philosophy *of* mathematics. In contrast, Badiou argues philosophy is, in actuality, a sub-genre of mathematics: "*philosophy must enter into logic via mathematics, not into mathematics via logic*" (Badiou, 2004, p. 15). Specifically, "mathematics is the science of being *qua* being" (Badiou, 2004, p. 15); or, mathematics *is* ontology.

Badiou's claim that ontology is mathematical does not imply that being is mathematical in nature. According to Ling, Badiou argues mathematics "figure[s] the *discourse* on being" (Badiou, 2010, p. 48); or perhaps more specifically, according to Badiou, "mathematics through the entirety of its historical becoming, pronounces what is expressible of being *qua* being" (Badiou, 1988/2005, p. 8). It is only at this time that we can have such knowledge since the development of Zermelo-Fraenkel set theory occurred in the early twentieth century. Zermelo-Fraenkel set theory, named after mathematicians Ernst Zermelo and Abraham Fraenkel, is an axiomatic system that, when combined with first-order logic (or predicate logic), provides a satisfactory and generally accepted formalism for almost all current mathematics. At this point in time, ZFC is generally considered by most working mathematicians as the foundations for mathematics, but other axiomatic systems do exist.

The foundational claim of Badiou's ontology is found in the statement: *the one is not* (Badiou, 1988/2005, p. 23). This is the very kernel of Badiou's ontology, and an original rethinking of traditional ontological claims where *what is* is one and *what is there* is multiple. Badiou explains:

> We find ourselves on the brink of the decision, a decision to break with the arcana of the one and the multiple in which philosophy is born and buried, phoenix of its own sophistical consumption. This decision can take no other form than the following: *the one is not*.
>
> (Badiou, 1988/2005, p. 23)

To Badiou, there is no unity or consistency to being, making his ontology ultimately anti-theological. In other words, being qua being is an inconsistent multiplicity making ontology the science of the pure multiple. Every moment is a tiny fragment selected from the multiples of multiples.

In order to unpack this, let us consider three key terms in Badiou's text, namely "situation," "count," and "void." First, what "is there" is the presentation of the multiplicity, that is, a rendering consistent of the inconsistent. This unified presentation is, according to Badiou,

$M\times\Phi$ vol. I.2 © 2023

a *situation*. In Ling's words, "a situation is the constitution of inconsistent multiplicity" (Badiou, 2010, p. 50); or to Badiou, "the place of taking-place" (Badiou, 1998/2005, p. 24). Second, the way in which pure multiplicity is *situated* or unified is, according to Badiou, the *count-as-one* or the *count*. For instance, in cinematic terms, the editing of a film leads to pure multiplicity at every cut, but the editor must count-as-one, situating the film. On the one hand, this can seem like a rather pretentious way of saying an editor makes choices in order to complete a film. On the other hand, experimental filmmakers often think about the pure multiplicity at every cut, an act that sometimes allows them to overcome predetermined conventions by expanding the lexicon. Finally, according to Badiou, the *void* is the space between the situation and its underlying being or more precisely, "every structured presentation unpresents 'its' void, in the mode of this non-one which is merely the subjective face of the count" (Badiou, 1988/2005, p. 25). Ling uses this idea to further argue, "first, according to the situation, the void is the proper name of being; and second, that everything that *is* is woven from the void" (Badiou, 2010, p. 51).

In *Being and Event*, Badiou is attempting to demonstrate that ZFC correlates to these ontological claims, justifying his controversial claim that mathematics is ontology. As Ling argues,

> Badiou's position on the multiple leads him to conclude that mathematics is ontology. After all, his two major ontological doctrines — that the science of being *qua* being (ontology) can only be the theory of pure or inconsistent multiplicity, and that all that *is* is woven from the void — are precisely what mathematics — or more precisely axiomatic or Zermelo-Fraenkel set theory [with the Axiom of Choice] (ZFC) — thinks.
>
> (Ling, 2010, pp. 51-2)

In other words, Badiou sees a direct correlation between his ontological arguments and ZFC. The first major connection is between the void in Badiou's ontology and what is referred to as the *empty set*, *void set*, or *null set* in ZFC, traditionally denoted $\varnothing$. One of the axioms of ZFC is the Axiom of Existence — which Badiou refers to as the Axiom of the Void Set — which asserts there exists a set with no elements; however, it is significant that this axiom can be deduced from two other axioms, namely, the Axiom of Power Set and the Axiom of Infinity, which implies that asserting it as an

$M\times\Phi$ **vol. I.2** © 2023

axiom is redundant (Hrbáček & Jech, 1999). Moreover, it is possible to deduce that the empty set is unique through the Axiom of Extensionality. Badiou argues,

> In its technical formulation — the most suitable for conceptual exposition — the Axiom of the Void Set [Axiom of Existence] will begin with an existential quantifier (thereby declaring that being invests the Ideas), and continue with a negation of existence (thereby un-presenting being), which will bear on belonging (thereby unpresenting being as multiple since the idea of the multiple is $\in$). Hence the following (negation is written $\sim$):
>
> $$(\exists\beta)[\sim(\exists\alpha)(\alpha\in\beta)]$$
>
> This reads: there exists β, such that there does not exist any α which belongs to it.

He continues:

> The mathematicians say in general, quite light-handedly [or more accurately, logically], that the void-set is unique 'after the Axiom of Extensionality'. Yet this is to proceed as if 'two' voids can be identified like two 'something's', which is to say two multiples of multiples, whilst the law of difference is conceptually, if not formally, inadequate to them. The truth is rather this: the unicity of the void-set is immediate because nothing differentiates it, not because its difference can be attested. An irremediable unicity based on in-difference is herein substituted for unicity based on difference.
>
> (Badiou, 1988/2005, p. 68)

Finally, he concludes, "it is because *the one is not* that the void is unique" (Badiou, 1988/2005, p. 69).

Given that this is one of the ways in which Badiou connects his ontological framework to ZFC, it is worth pointing out one major flaw, namely, that Badiou seems to have privileged knowledge about void-ness. For instance, Badiou simply asserts that the unicity of the void-set is *immediate* because nothing differentiates it; but how does Badiou have such knowledge? The *only way* that we should have any knowledge about the empty set is through ZFC

© 2023 $M\times\Phi$ vol. I.2

— the concept of the empty set does not have any implicit properties, and its uniqueness is obtained directly through the Axiom of Extensionality. There even exists a set theory where it is impossible to determine if the empty set is unique, namely, the set theory that consists of only the Axiom of Existence. In this ontological system, all that 'is' is void. In other words, its uniqueness is contingent on the set of axioms, not a fundamental truth, and unless we have some implicit or privileged knowledge of void-ness, we cannot determine if it directly corresponds to ZFC. Badiou is also correct in asserting that the *law of difference* is inadequate to mathematicians, but this is because it cannot be deduced from the axioms and is therefore irrelevant within this formal system. Does it just happen that Badiou's ontology corresponds to ZFC, or does the mathematics dictate his ontological claims?

The set-theoretic axioms weren't formulated in order to provide ontological arguments, but were introduced in order to produce a theory of sets that avoided paradoxes such as Russell's paradox. In "Matheme and Mathematics," mathematician Maciej Malicki further suggests that Badiou does not sufficiently or explicitly explain the connection between mathematical axioms and physical reality:

> One also needs to answer the question about the role played by axioms in the structure of *historical* situations. If in the domains of specific languages (of politics, science, art or love) the effects of event are not visible, the content of *Being and Event* is an empty exercise in abstraction: even science — perhaps excluding some entirely formalized areas of theoretical physics — let alone art or love — cannot for obvious reasons be exhaustively described solely in terms of the relation of belonging.
>
> (Malicki, 2015, p. 440)

In one specific instance, Badiou attempts to explain the French Revolution in terms of his conception of the event. His explanation is consistent within his own terminology; however, he does not explicitly show how this could be reduced solely to relation of belonging and the axioms of set theory.

In a review of Badiou's *Numbers and Numbers*, philosopher John Kadvany defends Badiou against a Sokal-esque claim that Badiou is simply misusing scientific and mathematical vocabulary as a form of pretension:

$M\times\Phi$ vol. I.2 © 2023

> Badiou, probably not noticed by Sokal, is in this way a *conservative* French philosopher, accepting modernist heterogeneity, but believing it to be mere appearance. The structure of Being, for Badiou, enables us to cognize it in excessively, possibly disastrously, manifold ways, exploring paths of innumerable options, scenarios, frames, and templates, with the whole an inconsistent multiplicity made up of Being's constituent elements. Our world, any world, is a tiny fragment selected from Being's 'multiples of multiples.' It's not that being is mathematical, but that mathematical discourse 'pronounces' what is expressible of 'being qua being.' Theories of anything, but mostly the natural and social world as described using numeric methods, are re-presentations of this ontology. Consistent with the primacy of natural science, numbers and numeric structure have to be 'immanent,' and especially, not 'constructed' via syntax, grammar or other inductive procedures, of which Badiou is completely disdainful: 'if it is true that mathematics, the highest expression of pure thought, in the final analysis consists of nothing but syntactical apparatuses, grammars of signs, then a fortiori all thought falls under the constitutive rule of language.' No Sapir-Whorf hypothesis for this fellow. The need is for an immanent structure of number and numbers, overwhelmingly efflorescent in its structure, persistently unbounded, always already beyond completion in every detail. (Kadvany, 2008)

Beyond giving a beautifully succinct overview of Badiou's overarching project (complete with Badiou's slightly idiosyncratic use of language), Kadvany is arguing that Badiou is not attempting to abuse mathematical ideas to further his own philosophical agenda, but rather attempting to create a unifying, mathematically based, philosophical framework. Moreover, by doing this, Badiou, like Sokal, is attempting "to undo the disastrous consequences of philosophy's 'linguistic turn' " (Badiou, 1988/2005, p. 16). Badiou's conviction in the highest expression of pure thought, mathematics, and his mathematical comprehension are beyond question, unlike many other postmodern philosophers who use and abuse mathematics, as demonstrated by Sokal and Bricmont.

$M\times\Phi$ vol. I.2 © 2023

One of the major points of contention in the infamous *Critical Inquiry* debate between philosophers Bartlett and Clemens, and father and son duo of mathematician Ricardo Nirenberg and historian David Nirenberg was "the Matheme of the Event." According to Malicki, matheme to Badiou is "understood as a philosophical idea subjected to rigours of deduction, and opposed to the pre-platonic poem" (Malicki, 2015, p. 434). Badiou defined the matheme of the event as:

$$e_X = \{x \in X, e_X\}$$

where $X \in S$ [that is, X belongs to S, or, using Badiou's terminology, X is *presented* by S] and S is a situation (Badiou, 1988/2005, p. 179). Badiou refers to X as the "evental site," so e_X is the event of the site X. The event is the set made up of all the elements of X and the event itself. To Badiou,

> [The *evental site* is] an entirely abnormal multiple; that is, a multiple such that none of its elements are presented in the situation. The site, itself, is presented, but 'beneath' it nothing from which it is composed is presented. As such, the site is not a part of the situation. I will also say of such a multiple that it is *on the edge of the void*, or *foundational*.
>
> (Badiou, 1988/2005, p. 175)

He later clarifies, "I term event of the site X a multiple such that it is composed of on the one hand, elements of the site, and on the other hand, itself" (Badiou, 1998/2005, p. 179).

The definition of *event* is another of the major points of contention for the Nirenbergs. They argue that e_X is not a set in conventional set theory:

> His 'set' e_X contains 'an inventory,' or 'the historical approach' — namely, $x \in X$ — but also, as we can see, something else: it contains itself. Rather than being defined in terms of objects previously defined, e_X is here defined in terms of itself; you must already have it in order to define it. Set theorists call this a *not-well-founded* set. This kind of set never appears in mathematics — not least because it produces an unmathematical *mise-en-abîme*: if we replace e_X inside the bracket by its expression as a bracket, we can go on doing this forever — and so can hardly be called 'a matheme'.
>
> (Nirenberg & Nirenberg, 2011, pp. 598-9)

$M \times \Phi$ vol. I.2 © 2023

More precisely, it can be shown that Badiou's formulation of an event is not a set in ZFC due to the Axiom of Foundation, which states that for every non-empty set there is an element of the set that shares no member with the set. In essence, this axiom exists to prevent Russell's Paradox.

In a review of *Being and Event*, philosopher Paul Livingston is a bit more generous, suggesting that this is not a mathematical mistake at all, but that Badiou intended for the event to be outside of ZFC. According to Livingston, "Badiou terms the 'event,' that which (as he argues) escapes any possible ontological reckoning, but is nevertheless at the core of history and the basis of any possible intervention in it" (Livingston, 2008, p. 218). As Badiou argues, "the event belongs to that-which-is-not-being-qua-being" (Badiou, 1988/2005, p. 189). As if directly responding to the Nirenbergs' objection Badiou states,

> The Axiom of Foundation de-limits being by the prohibition of the event. It thus brings forth that-which-is-not-being-qua-being as a point of impossibility of the discourse on being-qua-being, and it exhibits its signifying emblem: the multiple such as it presents itself, in the brilliance, in which being is abolished, of the mark-of-one. (Badiou, 1988/2005, p. 190)

In other words, Badiou suggests that his definition of event lies beyond the scope of ZFC precisely due to the Axiom of Foundation. Livingston reinforces this reading:

> As further set-theoretical reflection has shown, however, the Axiom of Foundation, though the most direct way to avoid Russell's paradox, is not strictly necessary for the logical coherence of an axiomatization of the nature of sets; various versions of 'non-well founded' set theory take up the consequences of its suspension. [...] For Badiou, however, this is not the basis of a rejection of the axiom itself as a fundamental claim of ontology, but rather an index of the event's capability to go beyond ontology in introducing happening into the intrinsically non-evental order of being. (Livingston, 2008, p. 225)

As Livingston observes, an event is purposefully outside of ZFC in what is called a non-well founded set theory, a branch of mathematics originally initiated by Dmitry Mirimanoff between 1917 and

$M\times\Phi$ vol. I.2 © 2023

1920. In fact, Badiou explicitly states, "sets which belong to themselves were baptized *extraordinary* sets by the logician Mirimanoff," and continues, "we could thus say the following: an event is ontologically formalized by an extraordinary set" (Badiou, 1988/2005, p. 190). Unfortunately, Badiou is not correct when he asserts that "event is prohibited" by being since, as Malicki observes, "the Axiom of Foundation may be consistently replaced with its negation — for example, the Aczel Anti-Foundation Axiom is consistent with the remaining axioms of set theory" (Malicki, 2015, p. 446).

One of the more substantial critiques put forth by the Nirenbergs is that Badiou's "mathematical ontology disguises the contingent in robes of necessity" (Nirenberg & Nirenberg, 2011, p. 612). Even the axioms Badiou has chosen are contingent. Badiou argues,

> We definitely have the entire material for an ontology here [namely, ZFC]. Save that none of these inaugural statements in which the law of Ideas is given has yet decided the question: 'Is there something rather than nothing?' [...] The solution to the problem is quite striking: maintain the position that nothing is delivered by the law of the Ideas [namely, ZFC; specifically the Axiom of the Existence], but make this nothing be through the assumption of a proper name. In other words: verify, via the excedentary choice of a proper name, the unpresentable alone as existent; on its basis the Ideas will subsequently cause all admissible forms of presentation to proceed. (Badiou, 1988/2005, pp. 66-7)

From the Axiom of the Existence, there is something rather than nothing and "out of *nothing* (which Badiou interprets the set $\varnothing$ to be) the whole cosmos, he will show us, will be created or rather deduced" (Nirenberg & Nirenberg, 2011, p. 590). Of course, this is an overstatement since Badiou is not attempting to define the whole cosmos; nevertheless, as previously observed, the Axiom of Existence isn't a necessary axiom since it can be deduced from the Axiom of the Power Set and the Axiom of Infinity. Moreover, as the Nirenbergs argue, it is also possible to obtain the empty set from a Weaker version of the Existence Axiom which guarantees the existence of some set and the Axiom of Separation. In other words, from this weaker version of the Existence Axiom, we begin with something beyond the void, prompting a totally new Badiou-ean

$M\times\Phi$ vol. I.2 © 2023

interpretation that can be deduced from the law of Ideas (ZFC), namely, that *separation produces the void*.[(2)]

Another problem with Badiou's ontological reductionism is analogous to one of the major objections to the reduction of numbers to sets; namely, there are many ways to reduce arithmetic to set theory. Here are two of the most standard interpretations presented by von Neumann and Zermelo respectively. Take 0 to be $\varnothing$. Next, there are two natural ways to define the successor of x, either as $x \cup \{x\}$ or as simply $\{x\}$. In other words, depending on which system you subscribe to, the number two is interpreted as $\{\varnothing, \{\varnothing\}\}$ or $\{\{\varnothing\}\}$. As philosopher Paul Benacerraf argues, "any feature of an account that identifies a number with a set is a superfluous feature of the account (i.e. not one that is grounded in our concept of number)" (Benacerraf, 1965, p. 52). Therefore, Benacerraf concludes numbers cannot be sets. Philosopher Alexander Paseau addresses some of these concerns, ultimately arguing that the arbitrary nature of the interpretation is irrelevant and that "reductionism is not damaged by the availability of incompatible reductions" (Paseau, 2009, p. 51). While Paseau's claim may be true for numbers, the problem is much more significant for Badiou given the availability of totally incompatible interpretations, ultimately suggesting that Badiou's ontological claims are contingent on their interpretation. In other words, Badiou's interpretations of mathematical results seem to determine his ontological claims.

Badiou's claim that "mathematics is ontology" is also heavily compromised in light of the fact that there are many different ways to axiomatize mathematics — some of which, as previously demonstrated, lead to conflicting interpretations. Badiou is fundamentally making a much weaker (and less quotable) claim, namely, ZFC is an ontology. In other words, Badiou's ontological framework is relative to the axiomatic system chosen. Malicki further provides a mathematical argument to demonstrate that Badiou's ontological claims are true only if one neglects some of the results of ZFC. Using the logic of ZFC, Malicki demonstrates Badiou's "generic theory of truth and his philosophy of event can coexist only at a price of selective and instrumental interpretation of the mathematical component" (Malicki, 2015, p. 446). In other words, Badiou is fundamentally making an even weaker claim, namely, that some selective portion of ZFC is an ontology. From this,

[(2)] Just in case it isn't obvious, this formulation is intended to be tongue-in-cheek.

$M\times\Phi$ **vol. I.2** © 2023

Malicki concludes, "*Being and Event* provides no grounding for a deep ontological structure behind the realms of science, art, love and politics, and that the mathematical formulation of the theory of event has no positive content" (Malicki, 2015, p. 446).

Being and Event may not provide an ontological structure for science, art, love and politics; however, Badiou's attempted endeavour is quite admirable and not, like Malicki asserts, without positive content. As Livingston argues,

> From these results of set theory [Badiou] draws a host of provocative conclusions about being, knowledge, language, and truth, the paradoxical 'event' that interrupts them, and the structure of a reconceived subjectivity whose essence is 'fidelity' to its consequences. In deriving this wide-ranging philosophical discourse, Badiou treats the axioms and theorems of set theory (on one of its various possible formulations) as if they were something like a revelatory text in which one can directly read the contours of being itself, as well as their inherent limitations. [...] This identification, like other decisive claims throughout the text, is not the result of any deductive or inductive argument, but rather of a basic and free decision, which Badiou likens to the mathematician's decision to adopt or refuse a particular axiom in the course of speculative mathematical thinking.
>
> (Livingston, 2008, p. 219)

Once again, the "free decision" Badiou is making is precisely what weakens his claim for an ontological basis; however, this does not make his insights less illuminating. As observed by Kadvany, "Badiou's ontological narrative is allusive, poetic, and deeply metaphorically *inspired* by his understanding of modern set theory" (Kadvany, 2008). This, in essence, can be read as the heart and soul of the Nirenbergs' actual criticism:

> In deducing philosophical and political consequences from his set-theoretical arguments, Badiou confuses contingent attributes of informal models with necessary consequences of the axioms (we will call this type of confusion a *Pythagoric snare*). The politico-philosophical claims that result have no grounding in the set theory that is deployed to justify them.
>
> (Nirenberg & Nirenberg, 2011, p. 590)

$M \times \Phi$ vol. I.2 © 2023

To this, Bartlett and Clemens have responded, "that mathematics is ontology means precisely and decisively that philosophy does not do ontology and that mathematics does not determine philosophy. Nor does mathematics constitute truth. Nor does being subject on Badiou's terms mean, as Nini [Nirenberg & Nirenberg] brazenly claims, that 'our only choice lies with the axioms of set theory' " (Bartlett & Clemens, 2012, p. 368). Despite the fact that it is incredibly bad form to argue something *precisely and decisively* in the negative, it is worthwhile to heed their criticisms in the light of Kadvany and Livingston's idea of creating an ontological narrative based on set theoretical claims.

In spite of its major shortcomings, *Being and Event* is an example of an incredibly productive misinterpretation of set theory. Using ZFC as a basis, Badiou produces a consistent ontology, one that doesn't necessarily follow from its mathematical inspirations, but is nevertheless a highly creative misinterpretation of the mathematics that it attempts to mirror. As Kadvany argues,

> I think Badiou has the roles of informal mathematical narrative and proof exactly reversed. He believes, like set theorists of old, in mathematical *realism*. But that's not what counts in mathematics, Gödel's platonism notwithstanding. Believe what you want. What matters are new systems, logics, heuristics, conjectures, counterexamples, theorems, proofs. However you explain these is fine, but don't take mathematical metaphors too seriously, even as these are essential to understanding, communication, and teaching. In particular, the idea that ZF, or other set theories, provide 'foundations' is itself a metaphor, true in part, but today far from having the ultimate status envisioned by Frege, Russell, or Gödel. (Kadvany, 2008)

The ontology created by Badiou is one that deviates through a productive misinterpretation. The ideas it presents lay the foundations for an ontology that moves beyond theological or mystical ideas (which have no place in mathematics nor ontology), and attempts to provide an ontological framework outside of language and hermeneutics. There are also some connections between Badiou's ontological framework and the world. For instance, philosopher Christopher Norris (2009) suggests that Badiou's concept of the *count-as-one* can be applied to the ways in which political systems exclude and disenfranchise those who do

$M\times\Phi$ vol. I.2 © 2023

not *count-as-one* within the system, like unrecognized immigrants. Once again, this is highly speculative, since the rules of set theory were constructed to provide foundations of mathematics without considering potential legal interpretations. Despite providing an underlying framework, studying set theory to better understand a system in which being accounted for determines a person's political status seems akin to studying the retinal system to better understand those who exist on the *periphery* of a legal system.

At this point, I feel safe in asserting that Badiou's claim that *mathematics is ontology* is fairly unsatisfactory. Like Kepler's model in *Mysterium Cosmographicum*, Badiou has put too much faith in the underlying mathematics and mathematical metaphor, or perhaps he does not intend for us to take him "too literally." From Kepler's model, it is possible to see one of the ways to transform an 'error' into an advancement of knowledge. By observing the ways in which the empirical evidence deviated from the model, it was possible to build upon and transform Kepler's model into a more accurate model, one that reflects the real world. Returning to Badiou's model, this is precisely the way in which it can be used. Badiou is basing his ontological arguments on what he perceives to be a perfect model, namely ZFC, an idealistic pursuit. Badiou's model and methodology are flawed; however, if this model is to be useful, its utility will be found in observing the ways in which deviates from empirical evidence. Like Kepler's model, Badiou's philosophical system should not be seen as a final model but as setting the foundations for a new approach.

$M\times\Phi$ vol. I.2 © 2023

§ — The Logic of *88:88*: Appropriating Badiou.

Medina's film *88:88* blends diaristic filmmaking, formal experimentation, and staged scenarios, presenting brief glimpses and fragmented moments. The film also experiments with multiple images superimposed on the screen, frames within frames, and video feedback. The sound design is often jarring, with text and music cut abruptly, and with much of the fragmented texts whispered to the audience. The texts themselves are cut up and layered, and come from numerous sources including readings of philosophy and poetry, hip hop and personal conversations between friends. Through brief glimpses and snippets of conversations, the film creates an intimate portrait of life in the West End of Winnipeg, Manitoba documenting Medina's friends and the poverty and social

injustice many of them face. Reading the film in the diaristic tradition, the images and recorded conversations can be seen as documenting Isiah's external life, his interactions with his friends and daily experiences, while the spoken texts can be seen as framing Isiah's internal life, as materials that informs his worldviews and perspectives.

The title *88:88* is a reference to the graphical display that is left on a digital clock when there is no power, or the graphical default that blinks on and off on a digital clock to indicate a power outage. Medina provides a political reading, stating "88:88 (or --:--) appears if you cannot afford to pay your bills, demonstrating that people who live in poverty live in suspended time" (Enns, 2015, p. 9). Film critic Benjamin Crais extends this reading, suggesting its connection to the void and Badiou's concept of the pure multiplicity:

> With *88:88*, Isiah Medina gives a name to nothing. 88:88 signifies no money, no electricity, but also no time: a digital clock reads a time that does not exist, that has never and will never come. Yet, in giving it a name, Medina grants nothing a positive existence. It is not only a lack, but what the philosopher Alain Badiou calls a pure multiplicity or inconsistency: the 'stuff' that is ordered and structured in the presentation of being. 88:88 — an image that contains every possible readout a digital appliance can present (11:35, 02:50, 12:00, etc.). Nothing — 88:88 — thus becomes the ground from which all articulations can emerge, a pure potential from which individual existences are cut. (Crais, 2016)

In other words, Crais sees Medina's use of 88:88 as the void and as a metaphor for pure multiplicity. Medina further reinforces this observation, suggesting cinema is a *no-thing to see* (or in Badiou's terminology *inconsistent* or a *pure multiplicity*) before it is realized (made *consistent,* or *situated,* or *count-as-one*). Isiah argues "there will always have been no-thing to see, but this inconsistent no-thing, the interval, must be given structure, must be made consistent" (Coldiron, 2016).

Almost following directly from Badiou's conception of the void, Medina defines *88:88* similarly:

> There is no given, and even if when in poverty you can say 'I have nothing,' to be completely clear, this nothing is itself not given. So poor, even nothing itself is not given.

$M\times\Phi$ vol. I.2 © 2023

> So we need a new name of nothing. Our own name, to be able to begin. And that name for us was 88:88.
>
> (Coldiron, 2016)[3]

To express this idea in the film, Medina samples a line from the American rapper Big L: "I wasn't poor, I was po! I couldn't afford the – 'or.' " This lyric can be interpreted in a few ways: too poor to be able to afford the education to correctly pronounce poor; too poor to be able to afford the operating room (O.R.); too poor to be able to afford gold, given "or" means "gold" in French. In the case of Medina, there is also "so poor, even nothing itself is not given," a new conceptualization of the void, the space of total poverty. To live in this space means to embody pure multiplicity and to not count-as-one.

The concept of infinity also plays a role in *88:88*. Graphically, ∞ is simply the number 8 turned on its side. As observed by Crais, "it is only a matter of orientation to conceive nothing (88:88) as infinity $(\infty \times 4)$" (Crais, 2016). In the film, while one of Medina's friends discusses hearing voices, he mentions losing his trust in infinity. Medina responds "Wait?! So you lost trust in infinity?" Medina connects his friend's schizophrenic episode to his temporary loss of faith in the infinite (among other things), an idea that seems as problematic as equating schizophrenia to a loss of faith in God. In another scene, a woman dressed as a revolutionary wears a red armband with a black ω on it. In set theory, ω usually denotes the first infinite ordinal, a concept introduced by Georg Cantor in the late nineteenth century to conceptualize infinite sequences. By putting ω on the armband of a revolutionary, Medina implies the revolutionary potential of the infinite.

In an interview about *88:88*, Medina associates ω with the cut, hence asserting the revolutionary potential of montage in his filmmaking. Medina explains:

> The shot inevitably comes to an end, and let us call ω the end of this repetitive model of succession; ω ends the repetition of $n + 1$. A historical interruption to the tendency of naturalization in a tracking shot. It is not considered in frame, nor does it succeed it — as a point it surpasses the potential 'tracking shot' not by adding to it, but by

[3] Isiah explicitly references the symbol $\varnothing$. Moreover, many lines from *Being and Event* are read throughout the film and a well thumbed copy of the book makes an appearance.

$M\times\Phi$ **vol. I.2** © 2023

> being the horizon of its succession. A cut, ω, retroactively totalizes the potentially infinite shot, and becomes its *limit*. We can succeed by applying the same operation, $\omega + 1$, reopening succession. But ω is not a successor to the first succession; ω was itself a support for the prior potentially infinite succession. The consequence is that there is more than one form of the *intervallic*. The space between frames is not the same one, because, if so, then we are, despite appearances, still within a tracking shot, within the one, within the same form of succession that is $n + 1$, and no cut has taken place. (Coldiron, 2015)

Given that cinema is a *finite* sequence of shots, Medina sees the cut as a way of realizing cinema's infinite potential. Medina reinforces this reading by arguing that "Cantor would claim there are infinite paradises, infinitely new forms of the cut (the mark of infinity, the end of repetition), in the finite images of our world" (Enns, 2015, p. 9). In other words, at every cut a new realm of possibilities is opened up, or "without the cut, we remain in finitude" (Coldiron, 2015). By viewing the cut in these terms, Medina is attempting to use mathematics to expand and rethink the concept of the cut beyond its traditional uses.

To Medina, it is through the cut that cinema begins to represent the filmmaker's thoughts. As he suggests, "there is a mental image and a material image, but these are held together by the cut" (Coldiron, 2015). It is at the place of the cut that we begin to see the filmmaker's perspective, remembering "that there must be a cut before the shot even begins and a cut for the shot to end" (Coldiron, 2015). That is, in order to construct a shot, the filmmaker must cut it from reality. Through the cut, Medina believes that it is possible to free the footage from one subjective perspective. He argues,

> Without the subjective action of cutting creating new exclusions, true lines of division, we will only have one interval. There will only have been one cut, and the cut will be an objective law, rather than the infinite, subjective production of new truths. (Coldiron, 2015)

Although the cut may have the potential for "the infinite, subjective production of new truths," in the end there will be only one film. Doesn't this, using Medina's logic, make it an objective law? In contrast, even if the filmmaker chooses only one shot, *only one*

$M\times\Phi$ vol. I.2 © 2023

interval without a cut, they have chosen that shot, as Medina has previously argued, from all of the infinite shots available potentially allowing for the infinite, subjective production of new truths.

Finally, the graphical representation of infinite, ∞, resemble handcuffs, which also play a role in Medina's film. With his hands cuffed behind his back, Medina is shown walking through downtown Toronto. Chained by the infinite, Medina, shot from below, walks past massive skyscrapers and through a field of shadowy 8s. As Crais suggests, "Medina equates 88:88 with imprisonment, with numerous 8s forming the material out of which a chain-link fence is constructed" (Crais, 2016).

88888888
88888888
88888888

The revolutionary potential of the infinite is juxtaposed with the reality of those who conform to societal norms. Medina presents this argument in a slightly different form in the film while comforting a friend. He suggests, "when people think they're crazy, they just assume a certain way of being in the world is correct."

Where Badiou uses mathematical concepts to develop an ontological framework, Medina uses mathematics to develop a cinematic framework. Through the use of mathematics, Medina is attempting to see cinema in new ways. Like Badiou, he is also applying the same mathematical systems in an attempt to understand social, legal and psychological systems to explore problems beyond the underlying logic of ZFC — an axiomatic system not conceived with these concerns in mind. Nevertheless, by evoking the infinite, Medina is pushing the poetic potential of cinema by suggesting the cut as having the potential to create new possibilities. In thinking through the infinite, Medina is attempting to conceptualize the infinite possibilities of cinema, a cinema beyond that which has already been shown to us.

§ — Bibliography.

Badiou, A. (2004). *Theoretical Writings* (R. Brassier, A. Toscano, & A. Toscano, Eds.; R. Brassier, Trans.). London: Continuum.

— (1988/2005). *Being and Event*. (O. Feltham, Trans.). London: Continuum.

— (2008). *Number and Numbers*. (R. Mackay, Trans.). New York: Polity.

© 2023 $M\times\Phi$ vol. I.2

Bartlett, A. J. (2014). The Singular Invariant: On the dialectic which is not in Alain Badiou. *Crisis and Crititque*, 1(3), pp. 288–335.

Bartlett, A. J. and Clemens, J. (2012). Neither Nor. *Critical Inquiry*, 38(2), pp. 365–380.

Benacerraf, P. (1965). What Numbers Could Not Be. *Philosophical Review*, 74(1), pp. 47–73.

Coldiron, P. (2015). Necessary Means: Isiah Medina on *88:88*. *Cinema Scope*, 64. http://cinema-scope.com/cinema-scope-magazine/necessary-means-isiah-medina-on-8888/

Crais, B. (2016). *Program notes for 88:88*. http://kenotype.tumblr.com/post/149485926871/program-notes-from-the-8888-screening

Deleuze, G. (1995). *Negotiations, 1972-1990*. (M. Joughin, Trans.). New York: Columbia University Press.

Enns, C. (2015). Towards Infinite Light: An Interview with Isiah Media. *BlackFlash*, 32(1), pp. 4–9.

Hrbáček, K., & Jech, T. (1999). *Introduction to Set Theory* (Third Edition). New York: Marcel Dekker, Inc.

Kadvany, J. (2008). Review of *Number and Numbers* by Alain Badiou. *Notre Dame Philosophical Reviews*. https://ndpr.nd.edu/news/number-and-numbers/

Ling, A. (2010). Ontology. In A. J. Bartlett & J. Clemens (Eds.), *Alain Badiou: Key Concepts* (pp. 48–58). Durham: Acumen.

Livingston, P. (2008). Review of Being and Event. *Inquiry*, 51(2), pp. 217–238.

Malicki, M. (2015). Matheme and Mathematics: On the Main Concepts of the Philosophy of Alain Badiou. *Logique et Analyse*, 58(231), pp. 433–446.

Medina, I. (2016). Isiah Medina Introduces *88:88*. *Mubi Notebook*. https://mubi.com/notebook/posts/director-s-statement-isiah-medina-introduces-88-88

Nirenberg, R. L. and Nirenberg, D. (2011). Badiou's Number: A Critique of Mathematics as Ontology. *Critical Inquiry*, 37(4), pp. 583–614.

$M\times\Phi$ *vol. I.2* © 2023

Norris, C. (2009). *Badiou's Being and event: A reader's guide*. Continuum.

Paseau, A. (2009). Reducing Arithmetic to Set Theory. In O. Bueno & Ø. Linnebo (Eds.), *New Waves in Philosophy of Mathematics* (pp. 35–55). London: Palgrave Macmillan.

Sokal, A. (1996). Transgressing the Boundaries: Towards a Transformative Hermeneutics of Quantum Gravity. *Social Text*, 46/47, pp. 217–252.

Sokal, A. and Bricmont, J. (1998). *Fashionable Nonsense: Postmodern Intellectuals' Abuse of Science*. New York: Picador.

Žižek, S. (2004). *Organs Without Bodies: Deleuze and Consequences*. New York: Routledge.

$M\times\Phi$ *vol. I.2* © 2023

Biography. Clint Enns is a writer and visual artist living in Tiohtià:ke/Montréal. He currently holds a SSHRC Postdoctoral Fellowship in the English Department at the University of Winnipeg. He has a Master's degree in mathematics from the University of Manitoba, and PhD in cinema and media from York University. His writings have appeared in journals including *Leonardo*, *Millennium Film Journal*, *Incite! Journal of Media*, *Found Footage Magazine* and in books including *Process Cinema: Handmade Film in the Digital Age* (McGill-Queen's University Press) edited by Scott Mackenzie and Janine Marchessault and *Models, Logics, and Higher-Dimensional Categories: A Tribute to the Work of Mihàly Makkai* (American Mathematical Society and Centre de Recherches Mathématiques).

Clint Enns (clint.enns@gmail.com)
SSHRC Postdoctoral Fellowship,
University of Winnipeg

⋆
⋆ ⋆

Annals of Mathematics and Philosophy
***vol.* I, *no.* 2**, 2023

$M \times \Phi$

Accepted proofs: Objective truth, or culturally robust?

Andrew Granville

Vanity of vanities! Vanity of mathematics!
– Frederick II of Prussia (*1778*)

Abstract. How does the mathematical community accept that a given proof is correct? Is objective verification based on explicit axioms feasible, or must the reviewer's experiences and prejudices necessarily come into play? Can automated provers avoid mistakes (as well as experiences and prejudices) to provide objective verification? And can an automated prover's claims be provably verified?

We will follow examples of proofs that were found to be flawed, but then corrected (as the proof plan was sufficiently robust), as well as accepted "proofs" that turned out to be fundamentally wrong. What does this imply about the desirability of the current community standard for proofs?

We will discuss whether mathematical culture is unavoidably part of the acceptance of a proof, no matter how much we try to develop foolproof, objective "proof systems".

© 2023

The quotation in the dedication is extracted from a letter [17] from Frederick II to Voltaire, complaining about the "expert advice" given to him by one Leonhard Euler. A special thanks to Michael Hallett for explanations of the history of the development of axiomatic systems, help with formulating my viewpoints and pointing out some remarkable flows of ideas, Rodrigo Ochigame for developing similar ideas from a very different and thought-provoking perspective, as well as Peter Scholze for several comments and corrections. I would also like to thank Dave Benson, Kevin Buzzard, Dan Goldston, Deirdre Haskell, Anthony Kennedy, Heather Macbeth, Ján Mináč, Mel Nathanson, Peter Sarnak, and Carlos Simpson for helpful discussions, responding to emails and useful references, as well as the referee for a careful reading and good suggestions. The author is partially supported by grants from NSERC (Canada).

Contents

1. *Proof – why and how*: Hilbert's axiomatic approach to mathematical theory.
2. *Living with, and ignoring, the Gödel crisis*
3. *Formal proof vs culturally appropriate, intuitive explanation.*
Including a discussion of approaches to proof and the use of language.
4. *What is an accepted proof in pure mathematics?* The realities of the refereeing process, and what is "interesting".
5. *Mistakes*. Famous recent mistakes and how the community responded to them.
6. *Rethinking axioms and language*. Investing in new theories has its pitfalls.
7. *Computers and proofs*. Establishing, assisting with, and generating, proofs.
8. *Uses of computers in major theorems*. Inspiring examples and computers in their proofs.
9. *Computer error*. The obvious pitfalls and whether we can elude them.
10. *Protocols for automated theorem checkers/provers*. What can we hope for?
11. *The future of proof*. Computer proofs and a modified Turing test.
12. *The Lean Theorem prover*. Extraordinary recent developments with Peter Scholze.
13. *Myths of objectivity*. Growing myths about the correctness of computer-proofs fit with well-known examples of refuted objectivity.
14. *Will machines change accepted proof*? Some conservative speculation about the future.

§ 1. — Proof – why and how.

To begin we discuss why proofs are desirable, what is the generally accepted approach to proof, and what aspects are theoretically problematic.

The purpose of proof. Aristotle wrote

> If ... understanding is as we posited, it is necessary for demonstrative understanding ... to depend on things which are true and primitive and immediate and more familiar than and prior to and explanatory of the conclusion.

More generally we can prove new concepts by reducing them to those which have already been accepted; in particular one does not need to always deduce the latest assertion directly from the axioms, as beautifully explained by Nathanson [49]:

$M\times\Phi$ vol. I.2 © 2023

> We mathematicians have a naive belief in truth. We prove theorems. Theorems are deductions from axioms. Each line in a proof is a simple consequence of the previous lines of the proof, or of previously proved theorems. *Our conclusions are true, unconditionally and eternally.*

That is, modern mathematicians dream that all theorems should be provable from appropriate axioms, and we just have to find the proofs. However, even back in the late 19th and early 20th century, it was found to be difficult to decide what precisely is meant by these statements.

Hilbert's vision. In early discussion of the foundations of mathematics, the consensus was to build on the already-acknowledged correct "Aristotlian primitives", rather than question whether one could (and should) start with any set of coherent, consistent axioms. For example, Hallett [30] writes:

> Like Dedekind, Cantor argued that progress in mathematics depends on conceptual innovation, the central constraints being the 'integration' of the new concepts with already accepted concepts and the condition that the concepts be consistent.

Although this suggests that there is unlikely to ever be an eternally fixed set of axioms, there were two important attempts to formulate a single sufficient axiomatic system: the first, due to Frege, was flawed because it was inconsistent, giving rise to Bertrand Russell's *antinomies*, that is, pairs of reasonable-sounding statements which are contradictory (leading to logical paradoxes[(1)]; though Gödel [20] points out that these antimonies "do not appear within mathematics" itself but rather within its "interpretation"). The second, developed in Russell and Whitehead's *Principia Mathematica*, was also flawed since it contained three axioms that they could not justify (and indeed seemed ad hoc).

Cantor's work on infinities was partly developed under the assumption[(2)] that every set can be well-ordered. Better understanding this assumption formed part of Hilbert's First Problem,

[(1)] For example, the wonderfully relevant paradoxical assertion, "This statement is not true".

[(2)] Which Cantor called "a law of thought" in 1883, what might be called "self-evident" today.

$M\times\Phi$ vol. I.2 © 2023

resolved in 1904 by Zermelo's *Axiom of choice*. These issues led Hilbert in the 1920s to develop his foundational program, for the first time explicitly beginning with axioms to establish a framework to study the foundations of mathematics, where abstract, ideal mathematics forms the very subject matter, without earlier prejudices that a particular route is desirable. The axioms create the theory, not vice-versa.

Hilbert wanted axioms that are developed consistently through clear rules, and do not lead to antinomies and paradoxes.[3] Hilbert therefore did not pretend, like Frege and Russell, to select axioms because they are 'truths', but rather because they are reasonably straightforward assertions which allow the subject matter to develop consistently. In particular an axiomatic system does not need to start with 'basic incontrovertible truths', but rather starting points that make sense and are consistent.

> Hilbert did not share Russell's interest in 'getting the foundations right', in isolating the 'right' set of primitives. For Hilbert, there is no 'right' set of primitives; some might be better than others for certain purposes, but there is never really any final word, even for relatively simple theories [4] ... Hilbert stressed that any theory is only a 'schema of concepts'; it is 'up to us' how to fill it with content. — Hallett [30]

Indeed one prefers to supplement the basic theory, not so much by using new axioms, but rather through different conceptual frameworks, embracing a much wider range of concepts than just the numbers and their properties. Axioms cannot be proved, and are not in need of proof.

Hilbert also worried about language and interpretation, demanding that all interpretations of the theory should be isomorphic, and all deductive techniques should be invariant under

[3] Hallett [30] remarks that part of the inspiration for this came from the proofs of the consistency of the alternative, non-Euclidean, geometry of Gauss, Bolyai and Lobatchevsky.

[4] Other leading logicians, like Frege, felt that a line is a line, the physical entity in the common vernacular, and if you produce a theory yielding something different with an obscure, though consistent set of axioms, then you should discard that theory; Hilbert disagreed.

$M\times\Phi$ vol. I.2 © 2023

different interpretations.[5] Hilbert's program lies at the base of modern mathematics so let's clarify the details of some of the ingredients in these theories.

Constructing a "formal reasoning system". Language is often imprecise, and people's interpretations and recollections can differ. We want to create a language that promotes accuracy and a lack of ambiguity, and so we focus on its rules:

We begin with an alphabet of symbols and variables; anything you like as long as it is countable and then sentences are only allowed to be finite combinations of those symbols. Then we have grammatical rules of how to write meaningful sentences; again these can be whatever you like as long as they seem to be consistent and flexible. The only substantial idea is that there should be "theorems", formulas without free variables (also called "closed formulas"). The first theorems are called axioms and can be enunciated without proof; then all others must be deduced using (only) our deduction rules from the axioms. Moreover one should be able to decide in a finite time (defined appropriately) whether a proof is correct. Consistency is key: we should not be able to determine a theorem and its negation from our axioms.

First order logic keeps the special symbols down to a sensible minimum, allowing the quantifiers $\forall, \exists$, the formula connectives $\vee, \wedge$ and $\implies$ (with which we can express A implies B by $A \implies B$), and the negation $\neg$. This is enough of a language to work with most intuitive mathematics, and the hope was that it, or it supplemented by one or two necessary refinements, could cover all mathematical truths.[6] But we must always remember that

> Mathematics is writing. For all the quantification it makes possible and all the technological and scientific discoveries it has helped to produce, it is ultimately words upon words. There is a bedrock of definitions ... crosscut by axioms ... whose only restrictions are that they not contradict one another. From this starting material we derive the terse assertions of consequences that are known as theorems and lemmas and corollaries. ...

[5] Not only should all models of the same theory be isomorphic, but no matter how one expresses an idea, and no matter how one reads it, the ideas must all boil down to the same thing.

[6] Codifying the means allowed in proof has led to modern logic, based on a small list of logical primitives, and such simple rules of inference.

$M\times\Phi$ vol. I.2 © 2023

> The arguments are the key. These are the "proofs" ... without which the assertions are just so much blather. *The proofs actually conjure mathematics into existence.*
>
> — Dan Rockmore [54], 2022.

Plurality. The foundations of mathematics starts with a set of axioms, but which set? Hilbert's proposals leave open the possibility of working within different axiomatic systems, and perhaps those different axiomatic systems will lead to contradictory conclusions to the same simple questions. Then how do we decide which axiomatic system is the correct one to use? This has not really been an issue for research mathematicians who accept that one cannot avoid *deductive pluralism* (see Hosack [34]), as long as each axiomatic system proceeds consistently. Let's recall two famous examples:

Zermelo introduced the Axiom of Choice (AC) in 1904, though it was only in the late 1930s that Gödel was able to show it is consistent with the other axioms.[7] However the shock came in 1963 when Paul Cohen showed that it is fully independent from the other axioms since $\neg$AC is also consistent with the other axioms. Therefore do we work with the axiom of choice, or not?

Then there is Cantor's 1878 *Continuum Hypothesis*, the claim that there are no infinities that in size[8] lie strictly between the set of integers and the set of reals. In 1940 Gödel showed that one could not use the other axioms (including the Axiom of Choice) to show that there are such infinities, and in 1963 Cohen showed that the other axioms could not show that there are no such infinities. That is, the Continuum Hypothesis is independent of the standard axioms.[9]

Research mathematicians have little difficulty in accepting this plurality. Indeed some of the most interesting research programs in recent years in pure mathematics have needed to rethink some of the starting assumptions and definitions to make further progress on questions of interest (as we will discuss later). Nathanson [50] remarks that this plurality can lead to the misguided notion that

> mathematics is the logical game of deducing conclusions from interesting but arbitrarily chosen finite sets of axioms and rules of inference.

[7] For the experts, by the "other axioms", I mean Zermelo-Fraenkel set theory.

[8] Size here is defined in terms of 1-to-1 correspondences.

[9] Thus if ZFC denotes Zermelo–Fraenkel set theory together with the axiom of choice then *both* ZFC+CH and ZFC+$\neg$CH are consistent.

$M\times\Phi$ vol. I.2 © 2023

Indeed even Gödel [20] agreed that

> The truth of the axioms from which mathematics starts out cannot be justified, and therefore [one can argue] the drawing of consequences ... is a game with symbols according to certain rules, not insights.

Nonetheless there is broad agreement to begin with the nine axioms of ZF, perhaps adding the axiom of choice. To be widely accepted, any radically new starting points need to be selected for good reasons, really important reasons, or they will gain little traction. For example, there is much research today on which axiom or axioms need to be added to the standard ones to "decide" the truth of the Continuum Hypothesis.[(10)] Thus although Hilbert's dream has been widely accepted, Gödel [20] noted that we still make choices as argued by Dedekind and Cantor, just at a different stage and with a different level of confidence that what we do is coherent and consistent.

© 2023 $M\times\Phi$ vol. I.2

A key goal. Axiomatic systems are not only there to ensure correctness, but hopefully to provide a framework to prove all the theorems that are worth proving. After all if we can state a theorem, surely it should be provable in our "theory", in that the theory should not have any artificial limitations.

> Hilbert ... demand[ed] that an axiomatised theory be complete, meaning that the axiom system be able to derive all the important facts, or all facts of a certain sort.
>
> — Hallett [30]

As we wrote above, we wish to create a (finitely described) language/theory to quantify and work with all intuitive mathematics, perhaps enhanced by further axioms as we determine new issues that are independent of the axioms we are already using, so that any true theorem is accessible and provable. Hilbert's dream was that one could create axioms and a theory that would allow one to prove or disprove any given mathematical claim.

In 1931 Gödel showed that this dream is inescapably impossible.

[(10)] Woodin's (*)-axiom, or Martin's maximum for forcing, or variants of either, and whatever we choose needs to lead to a rich and consistent theory of infinities which can answer many of the important questions.

§ 2. — Living with, and ignoring, the Gödel crisis.

In 1931 Kurt Gödel shook mathematics (literally) to its foundations with his *incompleteness theorems*:

I) No consistent finite set of axioms and rules[(11)] can be used to prove all true theorems about the integers.

Even worse:

II) No consistent finite set of axioms and rules can prove itself to be consistent!

This absolutely contradicts Hilbert's dream that one could axiomatize so as to prove or disprove any given mathematical claim. The first is frustrating, while the second unveils inescapable limitations in the possible formulation of axiomatic theories and indeed mathematical foundations:

> Formal, axiomatic set theory ... cannot be a final foundational theory for mathematics, [yet] nothing else [can] be. — Hallett [30]

Or as von Neumann put it in 1930:

> There can be no rigorous justification for classical mathematics.

So how do mathematicians deal with this existential crisis in their subject? The only answer is that they learn to live with it.[(12)] Here's how:

How to (not) deal with incompleteness.

At the beginning there are axioms; consistent, independent and all powerful.

They are designed to be used for the rest of eternity to prove theorems, building up from these axioms a towering, self-justified, self-supporting edifice, a structure that allows one to objectively create a theory of everything.

[(11)] I am avoiding necessary subtleties. To be more precise, Gödel's theorem can be stated as: The set of provable statements, although not recursive, is recursively enumerable, whereas the set of true statements is not.

[(12)] From Janella Baxter on Twitter (@BaxterJanella, Oct 6, 2021):"When my husband was a math PhD [student], Gödel's theorem terrified him. He worried the dissertation topic he'd selected was impossible to solve. He made it through, but he's baffled why more mathematicians don't share this anxiety. I think there's a philosophy of math paper to be written..."

© 2023 $M\times\Phi$ vol. I.2

However humans have found that there are questions that those axioms could not answer, so had little option but to supplement the axioms with further axioms to answer those previously unanswerable questions. But, no matter how many axioms were added it seemed that not all questions could be answered, not all theorems could be proved, some theorem always seemed to be undecidable using just the given axioms.

Indeed it came to pass that Gödel established there would always be mysteries, no matter how many or what axioms or language one starts with, questions that cannot be answered, created out of the very axioms that one starts with, in the language one has chosen.

How should one handle this impasse? What are theorem tower builders to do?

When one occupies the penthouse suite of a very high and seemingly solidly built tower, and there is a small fire below but it is difficult to review the damage done, one has options:

• One can abandon the tower and try to rebuild it; if one can only work out how to do so given the new fire code. This is risky as it may mean never ascending as high again.[13]

• After the smoke and noise have mostly dissipated one can continue on as if nothing is amiss. Maybe the cinders will smoulder for a bit and then go out of their own accord? And hey, if the building collapses, which seems unlikely given how well built it mostly is, then we won't be around to learn more.

The latter option is the choice of most pure mathematicians to cope with the Gödel crisis. Investing in the building that is, not in the building that may be yet to come.

Inescapable logic. For many mathematicians, Gödel's objections are melodic whistling in the wind – perhaps practical questions in algebra and analysis would never face these sorts of epistemological problems? Surely we need a formal reasoning system that is reliable in all "reasonable circumstances" despite being incomplete (though seemingly incomplete only when one is truly looking for confounding problems). Mathematicians floated on that cloud until the devastating Ramsey theorem of Paris and Harrington [51]: Although only mildly more complicated than what people work with every day in Ramsey theory, it cannot be proved in Peano arithmetic since, roughly speaking, it implies that Peano arithmetic is

[13] And who knows? They might change the fire code halfway through reconstruction, adding delays and extra cost.

$M\times\Phi$ vol. I.2 © 2023

consistent (and Gödel's second incompleteness theorem says that you can't use a theorem in Peano arithmetic to prove its consistency). There is now a collection of beautiful and natural theorems with the same property.

The general mathematical culture is to not worry about these things too much. If one works on a question close to where some problem has been found that is undecidable within the "standard" axiomatic framework then one needs to act carefully,[14] but for the most part it seems like a distant problem, and not one that we meet in our day-to-day work.[15] As Peter Sarnak likes to say of mathematicians, "we are working people".[16] Indeed for the last century the basic axiomatic system (ZFC) has remained accepted as the essentially unchanged foundation for most of modern pure mathematics, surprising in the light of Gödel's results, and yet it works.

That there is no way to evade Gödel's incompleteness theorems is so counterintuitive to the prevailing culture that some top people (including two Fields medallists) have suggested that somehow one can evade these issues of what is provable by writing suitably

[14] By "undecidable" we mean that if it is true it is not provable using the axioms in the theory, and if it is false it is not refutable using those same axioms.

[15] One does hear people suggest that popular unsolved problems might be "undecidable" within our axiomatic framework, which I regard as hubris; just because we have not yet found a good understanding of something does not make it an eternal mystery. Knuth [37] even makes the tenuous argument "the Goldbach conjecture ... [is] a problem that's never going to be solved. I think it might not even have a proof. It might be one of the unprovable theorems that Gödel showed exist ... we now know that in some sense almost all correct statements about mathematics are unprovable," and goes on to claim that Goldbach must be "true because it can't be false" for which he then gives a standard heuristic. The only salvageable truth from this is that Goldbach might not be provable in Peano arithmetic since there might be a different model of integers that satisfy Peano arithmetic yet for which Goldbach fails; however if so then we tinker with our axioms and add one to ensure we remain in the usual integers and then Goldbach should be provable. People have made analogous fatuous claims about the Riemann Hypothesis, the twin prime conjecture, etc with no real substance to back their claims. There is a good discussion of all this at `https://mathoverflow.net/questions/27755/` .

[16] My own attitude to this is akin to my attitude to food sourcing: I would like to eat wholesome, healthy, unadulterated, non-exploitative food, and I make choices at the supermarket to do so by taking a superficial look at the labels. However I am quite aware that some of their choices, and their labelling, may be questionable. Indeed I do not really believe my local organic supermarket to be "trusted authorities", but I do not want to spend my time carefully following up each purchase, holding my purveyors accountable, unless something in particular makes me suspicious (or naseous).

© 2023 $M\times\Phi$ vol. I.2

capable computer programs. Voevodskǐ's evasion is that it would be enough to have a "reliable" system, but is vague on what he means [64].[17] Thurston [61] seems to suggest that it is difficult to hide errors and have a program run since one has to be syntactically correct, which seems to confuse two different issues: most mathematical papers that contain important mistakes are written in a coherent manner, obeying the syntax and grammar of our system of communication – a misinterpretation of one of these is rarely where the mistake lies (though see the Biss case discussed below). In 1950 Turing [62] had already remarked that one can adapt Gödel's proof to create questions that any given computer language cannot answer, in that the question is undecidable within the computer language's design.

One can interpret the *Church-Turing thesis* as the working hypothesis that all sensible computational systems are "equivalent" and "universal". That is, they can each calculate anything that is calculable, though time may be an issue. Thus there is theoretically little difference between the capability of different programming languages (although they might have been designed for different purposes[18]). Thus if a computational system is "universal" then one should be able to use it to emulate any other computational system, and that includes humans and their deductive systems (what we defined above is meant to cover human activity too). Indeed Turing [62] writes,

> The idea behind digital computers may be explained by saying that these machines are intended to carry out any operations which could be done by a human...[19]

$M\times\Phi$ **vol. I.2** © 2023

[17] Presumably the idea is to define a plausible concept of "formal reasoning system" that allows one to evade Gödel's result.

[18] Once, one of my summer undergraduate research students calculated twin primes, and statistics about twin primes, using Excel (much to my surprise), which makes this point well.

[19] He goes on to exhibit the sort of instructions you might give a computer via a "domestic analogy":

> "Mother wants Tommy to call at the cobbler's every morning on his way to school to see if her shoes are done, she can ask him afresh every morning. Alternatively she can stick up a notice once and for all in the hall which he will see when he leaves for school and which tells him to call for the shoes, and also to destroy the notice when he comes back if he has the shoes with him."

He then writes

> Digital computers ... have been constructed, according to the principles we have described, and ... they can ... mimic the actions of a human computer very closely.

Surely Thurston, Voevodskĭ and others were aware that this interpretation of the Church-Turing thesis means that there is no hope that a different calculation method will lead to a *theoretically* better way, and therefore all we have left is what we do *in practice*. It is important to note that Turing's remarks (discussed at the end of the last paragraph) do not contradict "universality" since computers can each answer everything that is answerable, but not questions that are not within its design parameters (i.e. its axiomatic framework).

A suitable timeframe. Can there be a mathematical statement that is provable within our standard axiomatic framework, for which every proof is too long for humans' timespan?[(20)] In practice, mathematical problems often come in families denoted $A_1, A_2, \dots$ (like A_n means "factor the integer n"). Such a family is in the complexity class P if there is a polynomial time algorithm[(21)] to resolve each A_n, and is in the complexity class NP if there is a polynomial length proof for each A_n. What is the difference? For NP we don't require an algorithm to find the proof, only that it exists and so P$\subset$NP.[(22)] For example, no one knows a fast algorithm to factor integers, but given the factors it is easy to multiply them together to verify a factorization, and so factoring is in NP.

Simple counting arguments show that the set of NP-families of provable correct statements is a rather tiny subset of the set of all families of provable correct statements. Thus in practice, no machine, no algorithm, no human or computer can hope to access

[(20)] In practice, computing resources are finite. If every atom of the universe were a computer, working at light speed, from now until the end of time, that would still be a finite number of steps, since the number of atoms, the speed of light, and the remaining time are all finite.

[(21)] That is, the algorithm takes no more than n^c steps, for some $c > 0$, for every $n \geqslant 2$.

[(22)] We did not require that we know how to find the shortest proof, just that there is one that is so short. Thus "NP" stands for **N**on-deterministic **P**olynomial time, where "*non-deterministic*" means that one can not necessarily find that proof in a short time. We believe that P$\neq$NP, the outstanding open question in theoretical computer science.

$M\times\Phi$ **vol. I.2** © 2023

all, or even a tiny fraction of, the correct and provable mathematical statements deducible from a plausible axiomatic system.

In [16], Wigderson states that "problems in NP are really all the problems we ... mathematicians, can ever hope to solve, because [we need to] know if we have solved them[(23)]... this is the very definition of NP: a problem is in NP exactly if you can check if the solution you got is correct."

"If P=NP [then] all these problems have an efficient algorithm, so they can be solved very quickly on a computer.... if P=NP then everything we are trying to do can be done. [which] is why P=NP ... would be so consequential. However, I think most people believe that P≠NP."

§ 3. — Formal proof vs culturally appropriate, intuitive explanation.

Venkatesh [63] states that "a proof is generally understood to mean an argument compelling consensus" indeed that

> A proof is defined by the fact that [it] should induce uniform agreement about [its] validity, without any need for replication.

De Toffoli's definition [15] is that "a mathematical proof is a correct deductive argument for a mathematical conclusion from accepted premises that is shareable and verifiable a priori", moreover that

> It is the public availability of arguments that allows other mathematicians to perform quality control. This is essential to filtering out non-proofs ... Shar[ing] mathematical arguments ... [is] a necessary condition of mathematical justification.

However if proofs are judged by the community, then can a proof be said to be objectively valid? Indeed De Toffoli [14] believes that "criteria of acceptability for rigorous proofs are not carved in stone ... but are indexed to a mathematical community in a particular time".

Even more, human proof verifiers are most satisfied when, after examining a proof, they can reproduce it, or something like it, in

[(23)] The same can be said, for example, for physicists positing the existence of fundamental particles, in that they need to propose an experiment that could falsify or verify their existence.

$M\times\Phi$ vol. I.2 © 2023

their own words. This way the new knowledge is not just part of the sum of all human knowledge but rather it is contextualized, and known of its own accord as part of much larger picture. In Plato's *Theaetetus* (148e-151d) Socrates makes the important point that a failure of rigour can lead to a

> miscarriage of ... thoughts [so] ... the lack of understanding becomes apparent to themselves and to everyone else.

I would claim this supports the positive consequences of deducing new ideas from old and contextualizing them, rather than considering the new ideas in abstract isolation.

But these considerations do not address "objective proof". Surely we want objective proof? Ideas that have been verified so that they work in any context. If so, how can we achieve that? By making the proof and its understanding part of a larger program of understanding, or by ensuring the proof follows from the axioms, what we might call a "formal proof". Nathanson [49] writes for many in claiming:

> How do we know that a proof is correct? By checking it, line by line.

Hales [28] explains what formal proofs are and why some people find them appealing:

> A formal proof is a proof in which every logical inference has been checked all the way back to the fundamental axioms of mathematics. All the intermediate logical steps are supplied, without exception. No appeal is made to intuition, even if the translation from intuition to logic is routine.

This goes a little beyond Hacking's *Leibnizian ideal* [26] of a proof:

> Every step is meticulously laid out, and can be checked, line by line, in a mechanical way,

in that a formal proof requires one (at least in principle) to create a logical path deducing the claim from the appropriate axioms.

Perhaps the most common fear is that in reading a formal proof, line by line, one can be convinced by each step yet not perceive the whole, not see global error.

$M\times\Phi$ vol. I.2 © 2023

One might also ask how such a formal (or Leibnizian$^+$) proof would be verified? Do we just believe it on the say-so of the formal prover (human or otherwise)? If not, who would be the "independent authority" that does the verifying? And what skills does that independent authority bring to the job of verification? Would they check it line-by-line, and perhaps miss the wood for the trees? Or would the authority have a strong intuitive notion of what is going on (and if so where do they get that intuition from)? Hales goes on to write

> Thus, a formal proof is less intuitive, and yet less susceptible to logical errors.

How does one interpret this claim? What errors are not "logical errors"? Why does the formal approach necessarily reduce the possibility of overlooking issues that arise, especially if the reader is not expected to see the big picture?

More logical fallacies about formal proof. Formal proof tries to avoid "intuition" as being imprecise. There is a belief that a presentation with only formal steps can help an independent authority more easily verify the proof. In 1895, Peano wrote

> Imprecise ideas cannot be represented by symbols,

which implies that ideas represented by symbols must be precise. Even if we agree with this extraordinary (and unproven) claim, it is easy to misinterpret it by believing that once you have translated your mathematical problem into symbols you are guaranteed to not be wrong (rather than not be imprecise); of course, one can be precise and wrong! Let's suppose that we are careful about meaning. Then who is the "independent authority" that does the verifying? And if that authority has limited understanding and little intuition, could a subtle error slip by?[(24)] So Peano suggested a formalization that necessitates identifying the fundamental mathematical ideas in an explanation and then finding a way to express these within a limited language ([56, section 2.3]). Moreover Peano wished to reduce proofs in his formal language (via various "identities") to as few symbols as possible to conclude that the mathematics has

[(24)] Many errors can occur, not only mathematical, but also, for example, in the interpretation of symbols and language. It is easy to invent an unambiguous protocol for any given identified issue, but can we know we have thought out all possible situations that need a protocol to avoid ambiguity or misunderstanding?

$M\times\Phi$ vol. I.2 © 2023

then been better understood. Even if we agree that the mathematics is better understood in this context, what happens when one translates Peano's shortest proofs into the standard lexicon so more humans can understand? Will this be the most desirable proof? Will it help us to see the next results?

Moreover, can a formal proof verifier see how to reproduce the proof on their own? Is it really verified, if it is not easily reproduced? It is then part of the union of human knowledge but is it really known of its own accord?

Formal proofs typically chase the details of a proof back to the axioms. It is like a child tirelessly asking "Why?" (until one gets back to immutable truths), but at the end of that process, does the child remember what they asked at the start and how they got to the end? A proof like this is little better than the answer "Because I said so", no matter who is the objective trusted authority. When we are functioning participants in a community we expect answers we can understand, interpret, appreciate, and even use if possible. We can also be excited to find an alternative or clearer proof, though that plays no role in a formal system.

What other kinds of proof? Eugenia Cheng [11] writes that philosophers[(25)] believe that

> Thanks to the notion of 'proof', we have an utterly rigorous way of knowing what is and isn't true in mathematics,

but that mathematicians perceive formal proof as an over-focus on precision. So if we are so skeptical about formal proof, what works instead? There is a wide gulf that separates traditional proof from formal proof. Hales [28] writes,

> Traditional mathematical proofs are written in a way to make them easily understood by mathematicians. Routine logical steps are omitted. An enormous amount of context is assumed on the part of the reader ... A trained mathematician [can translate] those intuitive arguments into a more rigorous argument.

[(25)] I think she means to write "some philosophers" and indeed some mathematicians.

$M\times\Phi$ vol. I.2 © 2023

That is, a substantial amount of *tacit knowledge* comes into understanding and contextualizing traditional proofs. This is reflected Hacking's *Cartesian ideal* [26] of a proof:

> After some reflection and study, one totally understands [the proof] and can get [it] in one's mind.[(26)]

But this does not quite explain what leads us to believe that a proof is correct. Cheng [11] remarks,

> Although proof is what supposedly establishes the undeniable truth of a piece of mathematics, proof doesn't actually convince mathematicians of that truth ... Something else does.[(27)]

Research mathematicians therefore write "proofs" that are a convincing argument which perhaps could be turned into a formal proof, written primarily to enhance knowledge and understanding, while maintaining some level of rigour that convinces the reader that more could be done in that direction. Mancosu [42] notes that this type of proof

> does not bear directly upon some of the traditional foundational concerns, such as certainty, which have dominated much of philosophy of mathematics.

Nonetheless, we believe that such a proof, especially if it is widely understood, is "robust" and so less prone to error. This has long been the "community standard":

> There is no ... mathematician so expert in his science as to place entire confidence in his proof immediately on his discovery of it...Every time he runs over his proofs, his confidence increases; but still more by the approbation of his friends; and is rais'd to its utmost perfection by the universal assent and applauses of the learned world.
> — David Hume (1739)

$M\times\Phi$ **vol. I.2** © 2023

[(26)] Hacking seems to suggest that to be satisfactory all details of a proof should be understandable at the same time. Although this is certainly desirable, I am not sure it is required for believing in a given proof, so long as one has at some time felt confident of each part of the proof, and of how the different parts join together.

[(27)] Indeed in a letter to Dedekind in June 1877, Cantor exhibits a bijective map from Euclidean 2-space to Euclidean 3-space, commenting "I see it, but I don't believe it".

Hume is claiming that confidence in a proof stems from its robust nature under enquiry from technically competent, interested peers, not from some abstract verification. This system is not perfect. Significant errors are sometimes found in published proofs that have been accepted to be true,[28] and yet we stick with our system; it must have its advantages.

The usefulness of a good proof. When reading a proof, a research active mathematician wishes to add to their own intuition and scope, not simply agree that the proof's argument is correct. The reader is not passive. She wants to understand, to synthesize and to use the ideas in her own research:

> Mathematics ...[does not] reward passive consumption. Understanding a mathematical paper is like visualizing a building based on the architect's drawings: the text and formulas are only a blueprint that the reader must use to reconstruct the author's imaginary world in her mind. If she does that, however, then the best mathematical theories have the same breathtaking quality as the image of Paris folded on itself.[29] The experience can be both exhilarating and addictive. — Izabella Laba [38]

Different people get different things out of a reading and therefore a new research article can inspire new ideas in hitherto unforeseen directions. Even the same person can, at different times, get different things from reading an article, our understandings do change over time, sometimes even how we approach the whole area.

Rather than chase proofs back to the axioms, most readers rely on the published literature (a *library* of reliable knowledge), on what is already known. This means the reader asks "Why?" a reasonable number of times to be satisfied of the correctness of a proof (unlike what might happen in a formal proof), at least if the reader has enough current knowledge, in that they have read and mostly understood all, or most of, the references on which the article is based.

To summarize, proofs are accepted by community standards. This means that they might be wrong since we don't expect the details to be incredibly carefully checked; so what is the purpose of proofs? Perhaps the plan of the proof is the primary thing,

[28] And we would guess that there are many more errors out there, yet to be found.

[29] A reference to the dream-architect scene in the movie *Inception*.

$M\times\Phi$ vol. I.2 © 2023

explaining what ideas are strung together to prove the proposed theorem, so the reader can learn, retain and re-use the ideas.[30] The details are usually of less interest, especially as an experienced reader can reconstruct them, though sometimes with an undesirable amount of effort.

Writing a proof. A proof is an explanation to a particular audience.

> Mathematical explanations are context-dependent since
> "in different contexts different features might be salient"
> — Mancosu [43] quoting Lange [40]

An explanation for an expert in the field (the first readers of a research paper) is different from that for a novice student in the area, which is again different from someone not in the area. Here we are talking about the level of detail required, and explanation of how arguments fit together. Today, the usual protocol is that papers are written so that the abstract and half the introduction is accessible to a broad audience, an overview in the introduction to specialists (including novices), and the details for experts.

Explanation can be given by examples and test cases, by analogy with already understood results, or by proving the result in some special cases that highlight the main ideas. A "good proof" to a mathematician is one that explains as well as proves, in fact the better the explanation the less danger there is in omitting cumbersome details. However there is a wider community that may wish to read and appreciate a proof of a result and then the details may become more useful.

Cheng [11] claims that the purpose of mathematical communication is to turn the author's beliefs, via a proof, into a *believed truth* of her reader. One cannot convincingly just state the belief; the colleague will only start to consider it to be true if there is plausible reasoning attached. But what is reasoning? Will our correspondent believe the reasons? Without the rigour and structure of proof, without fitting the reasons into an appropriate framework, the reader will probably remain skeptical, or at least is less likely to remain skeptical with a well reasoned argument in the form of a proof. This is because a carefully worded proof helps allay the fear of ambiguity or misdirection.

[30] Cheng [11] compares this to legality (the proof) and morality (understanding that the proof is correct in principle). It would be nice if proof and understanding were synonymous but they are not, and we have to appreciate and accept how they match and how they differ.

$M\times\Phi$ vol. I.2 © 2023

Mathematical truth is beauty, and beauty truth. What makes a piece of mathematics "feel right"? Cheng [11] claims it is about what "ought to be", and not to be confused with "useful, fun, intriguing, beautiful, proved in detail". This perspective helps understand what motivates the approach that many take to proving theorems.

The prolific and influential mathematician Paul Erdös claimed that an objective supreme being has a "Book" which contains the perfect proof for every true theorem, each of which is short and elegant. Short, so it is easy to verify, and elegant so one knows that the statement fits so well that it must be true. This is a wonderful conceit of professional pure mathematicians, when a concept is ripe to be understood then it should willingly yield its most succulent fruits.

The great Grothendieck did not believe in big steps or examples to elucidate progress: he believed that when a subject is ready the theory should be clear so there is no point in trying to push progress too fast.[(31)] Case-by-case analyses may be complicated and clumsy but can help the researcher understand the patterns that may lead to a proof that makes a case-by-case analysis unnecessary. Proofs develop over time, and different authors choose to publish at different stages in the potential process.[(32)] Some research mathematicians are loathe to publish anything but a "final proof" in which one can see why a concept is true in one fell swoop.[(33)] Others are less selective, publishing a less complete theory, but partial results can encourage others and enhance the sense of community. It takes all sorts.

The language of research articles. To a large extent mathematicians follow the library model of building upon past papers. A theorem is a key new result of the research paper. Since research papers build on the library of knowledge, they typically quote what is known, and perhaps make simple modifications of what is already known to fit their needs – these are the "lemmas". If such a simple deduction seems to be interesting in its own right, then it might be called a "proposition". There are no hard and fast rules for the use of these words but one will find few variants.

[(31)] To my taste, a sad reductionist theory of progress.

[(32)] This is one of many arguments against "citation indices". Half-baked good ideas will get improved by lots of people and so will be often quoted. A brilliantly thought out breakthrough, with a beautiful proof, may be difficult to improve, even marginally, and so get referred to far less often.

[(33)] Including Gauss whose "motto" was "*Pauca sed matura*" ("*Few but ripe*" in English).

$M\times\Phi$ **vol. I.2** © 2023

One can argue that no two people view language in the same way, no matter how well defined. Indeed we get used to words in common usage, like "epistemic" seemingly used for a rainbow of connected but subtly different meanings. But language is how we communicate proof, and it is rarely a perfect tool for that. Instead we hope that our description gives enough details that the reader is able to "figure out" or to re-construct what we meant from what we have written, perhaps in their own way which may be subtly different (and that perhaps leads to something new). This argues for formal proof and language but we have seen that that has its pitfalls.

§ 4. — What is an accepted proof in pure mathematics?

© 2023 $M\times\Phi$ vol. I.2

Hales [28] writes that "philosophers tell us that mathematics consists of analytic truths, free of all imperfection" and proceeds to "prove that $1 + 1 = 2$." He then writes

> If only all proofs were so simple. Mathematical error is as old as mathematics itself. Euclid's very first proposition asks, *on a given straight line to construct an equilateral triangle*. Euclid's construction makes the implicit assumption (which is not justified by the axioms) that two circles, each passing through the other's center, must intersect. We revere Euclid, not because he got everything right, but because he set us on the right path.

The central idea of accepted proof is simple enough: Starting from agreed upon axioms we construct a proof of given statements. Famously, Russell and Whitehead showed that if our axioms are consistent and logically independent then many seemingly simple statements take an inordinate amount of proving. So to advance far in mathematics we need to avoid going back to the axioms all the time. We need to build a library of statements that we know to be true and are unambiguously stated. Traditionally this library is stored in research articles, and synthesized in books.

Then a researcher can advance the sum of all human knowledge by making logical deductions from what they can quote from our library. By now there are millions of articles, and so there is a lot of scope for errors to creep in to the system. Articles might be flawed, or quoted statements might be misinterpreted (or even mistyped).

The refereeing process. To try to ensure that articles are correct, we don't take the researcher's word for it that they have made correct logical deductions. Good journals assure quality control by assigning one or several anonymous "expert reviewers" to a submitted article. These reviewers/referees usually begin by judging whether what is claimed is true, new and interesting enough to be worthy of publication in that journal. Famously, Littlewood would ask[(34)]

> Is it new? Is it correct? Is it surprising?[(35)]

Then the referee is supposed to pick the text apart looking for logical fallacies. It might be that what is claimed is true but there are mistakes, possibly rectifiable, or that explanations given are hard to interpret or somewhat ambiguous. The reviewer might not only identify errors and ask for further explanation, but might suggest significant changes (that they think will be improvements) and so get involved in helping the article develop to its final form. The idea is that, in this way, the "literature" is interesting and safe to quote and believe for future authors.

The dangers of expert reviewers. Referees are usually chosen because they are expert in the topic of the submitted article; often they have been referenced by the author. Some referees are defensive of their work and are very picky when reviewing work by others on their favourite topic. Other reviewers are happy to see other researchers participate to explore the same questions. The outlook of those reviewers obviously affects how they review and what they find acceptable. As an editor of several journals I look for reviewers who are keen, generous champions of their field but hold themselves to high standards, so when they review someone else's work they expect things to be done as well as possible but are willing to help some authors to perhaps more fully think through their arguments so that what is finally published does constitute progress in the field.

Any sentient reader will see the pitfalls in the process as I have just described it. Prejudices can enter the picture in any number of

[(34)] Ben Green, the managing editor at the *Mathematical Proceedings of the Cambridge Philosophical Society* also asks his editors to *persuade him* that an article that they recommend is indeed "interesting".

[(35)] Venkatesh [63] argues that "the value we assign to a work of mathematics is purely subjective, in the sense that it depends solely on the perception of that work, and not on any objective quality".

$M\times\Phi$ vol. I.2 © 2023

ways and editors must trust the reviewers to be fair, though occasionally something is obviously amiss in a review.[36] There are relatively few people in esoteric research fields,[37] and so all the experts probably know each other, or at least of each other, and regularly meet at conferences. They have many of the same research goals and so are often either collaborators or competitors (or both). There is little hope that they can be truly neutral when assessing each others' work.

So what are the alternatives? Perhaps non-experts can impartially judge submitted articles? After all, any submission should build logically on the literature, and if written clearly the reader should be able to simply follow the argument. This assumes a lot. For example how easy is it to read and appreciate what is there in the published literature? The latest breakthrough might quote several major advances from the last few years, so to follow the arguments of the new submission the referee must be able to understand and appreciate what was cutting edge not so long ago. In fact, to referee a top quality work, the referee will have to attain a fluency in the top works of the field in the recent past, and so become, in part, an expert herself.[38] For example, I was recently asked to referee a 104 page paper by a top journal, the authors building on their own and others' work, amounting to several hundred further pages all published in top journals in the last five years. Although I am an expert in the field it took me weeks to do an adequate job, even though I have enough confidence in the authors' technical skills that I do not feel that I need to check every detail. As an expert this is an appropriate use of my time (so as to keep up with the latest in my field of research), but one could hardly expect someone who has not immersed themselves in these questions and techniques to give up weeks of their time and perhaps find themselves incapable of doing an adequate job.

© 2023 $M\times\Phi$ *vol. I.2*

[36] Though rarely.

[37] Venkatesh [63] writes, "The size and complexity of modern mathematics means that most papers are almost incomprehensible to us; our opinion of them can then only repeat that of others. The only people who can be involved in the formation of opinion about a given paper or a given question are those who interact with it in some way. Now, the set of people who study the details of any argument themselves is very small; a much larger group acquire, instead, an awareness of its relationship to other existing work."

[38] And therefore refereeing can only be done properly by experts – this feels like a Gödel-inspired argument!

Checking details. The reader might flinch when reading that there are situations in which "I do not feel that I need to check every detail" when refereeing. From a pure perspective that is self-evidently problematic – how can one review an article for errors and feel it appropriate to not check details? But this is where expertise comes in. If one knows the area then there are often proofs that are more-or-less "standard", so the expert reviewer can see that it looks roughly correct, and perhaps can identify quickly what part of the proof might differ from previous similar proofs in the literature, and focus their efforts there. Everything is new to the non-expert reviewer so they must check every detail, as they have little idea what precisely to scrutinise for where potential errors might arise. Moreover, the expert reviewer is better equipped to help the author fix a faulty argument since they might have been stuck on the same issue in the past.

Mistakes commonly arise from mis-applying the literature, for example by quoting a result out of context: Most articles are written to be read from beginning to end, but in the pages before stating a result an author might have included an assumption to be used throughout the article.[39] Then the researcher quotes that result verbatim, neglecting the earlier hypothesis and so misapplies the result. The non-expert reviewer might verify the statement of the quoted result but is unlikely to read the whole referenced article, and thus is unlikely to be able to identify this issue, whereas an expert might have more feel for when a stated result is liable to be valid.

I hope we now agree that expertise has its advantages in refereeing, and is indeed unavoidable in practice, although refereeing remains onerous no matter what. I have largely been discussing "cutting edge work" above, the sorts of advances that truly make a difference. However the majority of submissions are not game-changers, but rather, small advances, and perhaps more easily reviewed. Nonetheless the same caveats hold, just to a lesser degree. But here the expert can truly make the difference in helping an author who has good ideas but perhaps has not yet developed the technical skills to take their ideas all the way. In my experience as an editor many referees are encouraging and helpful in these circumstances, particularly if the author is an "unknown", explaining

[39] For example, in my area, analytic number theory, one might write "Throughout we let $f(x)$ be a Schwarz-class function" rather than repeating that in the statement of every Lemma and Theorem.

$M\times\Phi$ vol. I.2 © 2023

how they might modify what they have done to make the argument correct, or the theorem stronger or more general.[40]

Finally if only experts are capable of reviewing big advances in cutting edge research what happens when a new submission contains a genuinely new idea? Something really different from what has gone before? If all the potential referees are reading from the same hymn book, aren't they all likely to be skeptical of the validity of the new tune? Particularly if it is not-so-well explained or contains non-lethal technical flaws? Certainly there may be initial (understandable) skepticism but the current system often works well. For example, when 57-year old unknown mathematician Yitang Zhang claimed in 2013 to have made an extraordinary breakthrough on gaps between primes using the deepest of ideas in new and surprising ways, the established community rapidly acknowledged that this was a research development of the highest calibre.[41] It also helps that different journals act independently: for example, the paper in which Mordell proved Poincaré's conjecture on the finite generation of rational points on elliptic curves, and posed his famous conjecture that would eventually be resolved by Faltings, was initially rejected by the London Mathematical Society [48], presumably because it was not then a fashionable subject, but subsequently accepted by the Cambridge Philosophical Society [47].

When building a theory that is not interesting. In [52] the authors write "To judge the originality of ... work on the basis of a conception of the 'existing body of knowledge' which comprises both secret and possible knowledge is intellectually callous". Certainly if one presents it like that.

Some authors come into conflict with the arbitration process by failing to take account of the culture of the community. For example, in combinatorics one learns proofs and then often adapts them to new circumstances (which may require substantial new inputs); few general theorems are of value since they will often cover only the known cases and uninteresting generalizations, and

[40] This is why I am against "double blind-reviewing" in which the authors' name is concealed from the referee, as it discourages the referee's generosity: It seems referees tend to assume an anonymous author "should know better" than to make that mistake, and so give a terse explanation about a mistake, rather than a helpful one.

[41] Initially the experts were hesitant about immersing themselves in such a difficult manuscript by an unknown, but once the experts started reading they quickly realized that here was something special.

$M\times\Phi$ vol. I.2 © 2023

not be usable for new and genuinely interesting cases.[42] In his last few years Serge Lang was upset that journal after journal independently rejected his general work of this nature in analytic number theory. His perspective was that if his precise result had not been explicitly stated in the literature then no referee had the right to say that it was "known" (an axiomatic viewpoint), whereas practitioners felt that the proposed generalizations added little of value (a community viewpoint). There is an old joke in refereeing:

> What is new is not interesting, and what is interesting is not new.

In my experience, if this applies to a submission then the author may have difficulty understanding the report, no matter how well justified.

Whose responsibility is the correctness of an argument in an accepted (for publication) paper? The author, the reviewer, the editor, the journal? It is generally agreed to always be the author's responsibility, no matter how much scrutiny the paper has gone through. It is not the referee's job to assure the paper is correct but rather that they are supposed to have made a good, serious effort to verify the details.

The robust nature of proof. We would like a system in which mathematical statements that are known as "proven", are reliably correct and always true, because a proof exhibits a theorem as the logical consequence of steps that are each verified. The proof should surely guarantee the theorem, so a competent mathematician does not need to verify each proof she uses herself (though that would be preferable). However this is not really feasible with the current system of proof verification of expert peer-reviewing.

We have discussed the advantages of expert peer-reviewing. It largely rests on the idea that familiarity simplifies scrutiny (at some risk of favouring conformity) so that papers are more quickly and accurately refereed. We mentioned the idea that a referee, having seen a certain type of argument before, does not feel it necessary to review every detail, perhaps only focusing on those details that

[42] Gowers [24] explains the value in such a perspective: "While the structure is less obvious than it is in many other subjects, it is there in the form of somewhat vague general statements that allow proofs to be condensed in the mind, and therefore more easily memorized and more easily transmitted to others."

$M\times\Phi$ vol. I.2 © 2023

might be most likely, in the experience of the referee, to cause concern. This supports a belief in the *robust* nature of proof, at least at a high level of research. We believe that not much can go wrong with well-used technical tools and so we are prepared to make assumptions about what needs verifying, usually true but perhaps not always. And then even if there is a mistake, experience shows that a simple modification should be enough to make the argument work. Any experienced researcher does this regularly in developing their own work, so when they encounter minor technical flaws in the work of others, they tend to believe that they are fixable. Indeed many referees write in their report, a sentence like: "I do not need to see a revision as I am sure that the author will be able to deal with these issues".

$M\times\Phi$ vol. I.2 © 2023

§ 5. — Mistakes.

Our expert peer-reviewing proof-verification system can and does go wrong. Big results are widely used and errors are sometimes found after publication. Usually not from someone checking the details even more carefully, but rather by another researcher applying the claimed result and deducing something that is wrong, or implausible or at least unexpected and therefore suspicious. Let's discuss some notable examples.

Goldston and Yildirim. In 2003 Dan Goldston and Cem Yildirim announced an extraordinary advance on what could then be proven about the shortest gaps between primes: We believe the *twin prime conjecture* that there are infinitely many pairs of primes that differ by 2, but at that time it was not even known that there were infinitely many gaps less than a tenth of the average gap, and Goldston and Yildirim claimed they could get that down to any positive multiple of the average, no matter how small. They distributed a preprint in which the novel part of the argument was written out in easily verifiable detail, though the proofs of some of the technical lemmas were sketched. Nonetheless those technical lemmas all had their roots in similar statements in the literature and so were believed; it was felt that Goldston and Yildirim would just fill in the details later, the proof seemed robust and their work was widely believed, and prominent mathematicians told the scientific press what a great breakthrough it was, etc.

Asking them for more details on two of the lemmas, they responded that one of the lemmas was proving stubborn but that they expected to sort it out soon and then would post a completely proved revision. That seemed plausible, so Soundararajan and I then proceeded to further develop the Goldston-Yildirim method and, to our surprise, managed to deduce that there are infinitely many gaps between primes of size $\leqslant 16$, almost the twin prime conjecture! However something did not feel right and when we carefully traced our steps we found that that stubborn lemma seemed to imply too much, which led us to a counterexample.[(43)] This story again highlights the issues in accepting proofs based on robustness but there is no obvious way to proceed otherwise.

This story though has a happy ending. Working with Janos Pintz, Goldston and Yildirim were able to find a correct version of what they had done, which eventually led to Yitang Zhang's breakthrough, and the Maynard-Tao modifications of the Goldston-Pintz-Yildirim method, and ultimately to the knowledge that there are infinitely many pairs of primes that differ by no more than 246, one of the great results of mathematical history.[(44)]

Wiles. In 1993 Andrew Wiles announced in a series of three lectures at a small conference at the Isaac Newton Institute in Cambridge, that he had proved Fermat's Last Theorem (FLT), using a very deep strategy suggested by work of Taniyama, Shimura, Weil, Frey, Ribet and Serre, to show that certain families of elliptic curves can be parametrized as had been conjectured (the *modularity conjecture*). Wiles had asked the conference organizer, John Coates, whether he could give three one hour lectures, refusing to tell Coates why but suggesting that it would be worthwhile. I was at the conference and rumours flew high. What was clear at the time was that Wiles presented a well-tuned sketch of his proof, even though having just three hours to explain the conclusions after seven years of thoughts was not enough. Wiles had the audacity to create a brilliant program of research to finish FLT from where Frey, Ribet and Serre had

(43) For the reader who knows complex analysis, the key issue revolved around how to move contours when integrating in very high dimension. There are aspects of high dimensional geometry that are different from the dimensions we are familiar with. The Goldston-Yildirim lemma worked in low dimension but went awry in high enough dimension.

(44) Moreover there is a second happy ending: Green and Tao tailored the original Goldston-Yildirim sieve weights to complete their proof that there are arbitrarily long arithmetic progressions of primes.

$M\times\Phi$ **vol. I.2** © 2023

left it. Every step was either plausible with the then current technology, or he had a plan to make it plausible, and at least sketched the details. We were all aware that filling in those steps might require some changes in the precise details of the plan, but he was confident that what he had was sufficiently robust that it could withstand a few missteps that might need correction. My impression was that he gave the talks a little before he was 100% ready, but this was a conference that would be attended by most of the experts in the world, in a brand new research institute at Wiles' Alma Mater, and he felt the participants would prefer to hear about it now, since they could confirm that the steps were all doable or at least plausible (and even perhaps make useful suggestions).[45]

As is well-known, when Wiles went on to fill in all those details, one part of the proof started to feel more shaky and stubborn no matter what his efforts, and he felt that he needed to retract his claim of FLT, at least temporarily. Why had the community been so prepared to accept Wiles' claims, before his work had been released and refereed? After all, anyone can claim such a result.[46] The answer is about community culture. Wiles had already several major results that had changed the way we thought about algebraic number theory; he was already a professor at Princeton, and his every new paper already produced excitement. Moreover he was not known to have made any serious mistakes in his earlier work.[47] His description of his attack on (part of) the modularity conjecture was feasible to the experts in the room, indeed they could not wait to get their hands on a manuscript.[48]

When Wiles was forced to retract his claim, he worked to find a way to "patch up" his argument. First with comparatively minor rethinking of that part of the argument, and subsequently with more

[45] In my memory one participant remarked that Wiles used a result that was known to be false but that something similar should be true and provable. (I tried to verify this with that person but he did not recall.)

[46] Indeed many have. Before Wiles' Theorem many of us used to be inundated with amateur claims to have proven FLT. Afterwards we were even more inundated, at least for a while, but felt less guilty about refusing to read them over.

[47] Actually I don't know if he had made any mistakes, but that is not relevant. I included the word "serious" to emphasize that no one worried too much about minor mistakes as long as the game plan of the proof was coherent and feasible, particularly for the experts.

[48] Ken Ribet, gave the lecture right after Wiles' third in which Wiles had announced FLT. Ribet's lecture began with an excited room of mathematicians not quite knowing what to do with themselves. Ribet started by laughing a little nervously and said, "After that, even I don't care what I am about to say".

$M\times\Phi$ vol. I.2 © 2023

dramatic changes, but he continued to work with the same overall plan. He knew his plan to be robust, indeed several parts of his proof plan have developed into interesting mini-subjects in their own right, that something should work to fill in the gaps, it was just a question of finding a workable path.[49] While Wiles worked on the "fix" (eventually collaborating with his former student, Richard Taylor) there were different reactions from leading researchers in the community. A few competitive souls told the press that what Wiles had was far from a proof, but the majority seemed to be willing Wiles on; praising him for what he had already accomplished, acknowledging the brilliance of the approach he had created and expressing hope publicly that he would complete his Herculean task.[50]

Biss. In the early 2000s, Daniel Biss made some important advances in homotopy theory, publishing papers in top journals like *Annals of Mathematics* (indeed his work developed ideas of leading mathematician and Annals editor, Robert MacPherson). Biss's proof underwent the tough scrutiny you would expect from a journal like Annals. And yet there were mistakes, in fact the key mistake was submerged in the paper, in Proposition 4.5. Biss's result had seemed "feasible" to the experts yet the mistake invalidated his main result. Similar mistakes appeared in other Biss papers in top journals like *Inventiones* and *Advances in Mathematics*, very likely refereed by other top experts. So what went wrong with the system here? Do all the experts think so similarly, and accept what seems plausible so readily that such a mistake slips by, even in the most reputable of journals? It gets worse though. Other Biss papers have mistakes; for example, in 2017 *Topology and its Applications* "retracted" Biss's 2002 paper as "the definitions in the paper are ambiguous and most results are false".

It is not wholly uncommon in some of the more abstract areas of pure mathematics that such issues arise with definitions or the simplest arguments (since they are often the least scrutinized). The flow of the overall argument is beautiful and persuasive so even if some of the details have not all been ironed out by the time the

[49] That path required further and deeper understanding than the original. In retrospect, he had made it more difficult for himself by announcing his proof publicly, as he then felt under pressure to get it done, whereas he had previously spent seven years working well away from the spotlight.

[50] There were also some who felt that Wiles should publish what he had and let others "have a go" at fixing the hole.

$M\times\Phi$ vol. I.2 © 2023

author has gone public, that is usually a technical issue that will be rapidly resolved. Biss's mistakes were identified by Nikolai Mnev in St Petersburg who quietly notified the author and other experts and expected Biss to rapidly resolve the issues or publish a retraction. Biss worked with colleagues who believed that these issues might be resolved and yet were frustrated. Time passed and Mnev, after four years, felt he should notify a broader community (by publishing his observations on the arxiv), as he was worried other people would develop Biss's work without realizing that it might be wrong. Four years can seem a long time to publish a correction or an erratum, but it might take a while in very deep and difficult research to identify the correct path forward, particularly under pressure.

In these last two examples the community held off for a long time passing judgement on a mistake until things are clarified. There is perhaps no correct answer as to when we should abandon a brilliant and seemingly robust proof plan.

© 2023 $M\times\Phi$ vol. I.2

§ 6. — Rethinking axioms and language.

Mochizuki's rethink. Shinichi Mochizuki is a highly respected researcher in arithmetic geometry who has made fundamental contributions to anabelian geometry. In 2012 he announced a proof of the abc-conjecture, perhaps the most fundamental open question in Diophantine arithmetic. The proof stems from the creation of "inter-universal Teichmüller theory" (IUTT), which he claims is a massive re-think of algebraic geometric aspects of Diophantine analysis. Mochizuki did something very different by discarding a lot of the development of arithmetic geometry and replacing it by his new theory that he believes is more "fit for purpose". Mochizuki did not compromise in his creation, the language is new, the concepts are, he claims, new and different, and he suggests that they are not translatable into the usual language and structures used in arithmetic geometry. Mochizuki has every right to create his own theory involving a new language and structures. As long as his proofs can be shown to follow from appropriate axioms in the usual manner via some path, then they are correct.[(51)]

[(51)] Indeed many top mathematicians (one associates Grothendieck in particular) have found that the correct language and definitions can change the complexion of

However, a practical problem arose in refereeing his works; Mochzuki had gone so far afield that there were no other experts to decide upon the correctness, and those who would usually seem best qualified to judge research in this general area were not prepared to do so without collegial help. Mochizuki uses a new language, and his four preprints giving his claimed proof of the abc-conjecture are about 500 pages long. He makes no attempt to compare his ideas with the standard lexicon, or to give motivating analogies with the usual arithmetic geometric objects. Moreover he turned down requests for more standard explanations from those experts stating that since everything is clearly defined in his work, those other experts have a responsibility to study Mochizuki's work from first principles. This would have meant putting their own research careers on hold for a significant time, though arguably they would have gained much in doing so if they eventually agreed there is high value in Mochizuki's proposed revolution.

Mochizuki's proof establishes a form of the abc-conjecture that is a bit different from the usual conjecture. Vesselin Dimitrov then deduced a more applicable version of Mochizuki's result, and found counterexamples, so that the result originally claimed by Mochizuki is provably incorrect. Mochizuki, claiming his proof was robust even if there were some minor glitches, then weakened one claim so that the eventual result could not be disproved by Dimitrov's example. Although it is not unusual that long and complicated proofs need some minor revisions, in this case the extremely minor change led people to wonder whether there might be counterexamples to the new version that one does not yet know about – what then? How many "tweaks" are acceptable?

A small coterie of established researchers studied Mochizuki's works and believed the ideas and logic to be plausible. In the process of these studies Mochizuki continued to revise what he had done to make that process easier. However that still left the majority of top arithmetic geometers unwilling to immerse themselves in IUTT, and so the community could neither satisfactorily accept nor satisfactorily discard Mochizuki's claims.[52]

deep problems from being far from what is doable to being something that follows inexorably from the theory. Indeed Scholze [32] recently wrote: "We perceive mathematical nature through the lenses given by definitions, and it is critical that the definitions put the essential points into focus."

[52] And this is not the place to comment on its correctness, since our focus is on how the community has functioned in this peculiar and complicated situation.

$M\times\Phi$ vol. I.2 © 2023

Peter Scholze, one of the top mathematicians in the world, took the responsibility of trying to translate enough of Mochizuki's manuscripts into more common mathematical language so that he could judge for himself. Believing they understood Mochizuki's strategy, Scholze and Jakob Stix identified a point (Corollary 3.12) at which they could not see how to proceed, and went to Kyoto in 2018 to discuss this with Mochizuki. Mochizuki met with them but claimed that their translations involved invalid oversimplifications of IUTT so that their objections were invalid. However he was unable to persuade them that they were missing the point. They published a report [59] in which Scholze and Stix claimed that "small modifications will not rescue the proof strategy".

This is a situation in which the usual system has broken down because the majority of mathematicians expect to be presented work that is presented as an addition to what came before and following certain traditions. Not only in the writing but also in the communication. Mathematicians expect to be able to ask each other questions about their latest work, either at conferences or by email, and expect clarifications in the most convenient language for all concerned.(53)

$M\times\Phi$ vol. I.2 © 2023

Voevodskĭ. Vladimir Voevodskĭ was a great intuitive mathematician. Trusting in his intuition he was shaken to discover some of his work with Kapranov was wrong. Somehow his intuition was insufficient and there were fatal mistakes in some of their arguments, and even counterexamples.(54) Voevodskĭ highlighted in [53] what can go wrong with our expert-peer review system: "A technical argument by a trusted author [like him], which is hard to check and looks similar to arguments known to be correct, is hardly ever checked in detail."

His mistake led him to propose a re-think of the foundations of his field: The *univalence foundation program* launched by Voevodskĭ attempts to reconcile Martin-Löf-style dependent type theory with

(53) For example, Nick Katz at Princeton writes with a great familiarity about certain profoundly deep algebraic-geometric objects in a way that I find difficult to appreciate. But in discussion he is prepared to discard all that and to give as down-to-earth an explanation as possible, recognizing my limitations. This is arguably the "community standard".

(54) It took him 15 years from being made aware of the mistake by Carlos Simpson (via Kapranov), to acknowledging it since he wanted to find out where his intuition had gone wrong, and not worry about a detail that might have been correctable.

the traditional mathematical treatment of proofs and categorical constructions. This involves a lot of rethinking of foundations with a new approach and new language. However in this case, one of the main architects of this program, Dan Grayson, has gone to great pains [25] to try to make the fundamental ideas accessible, and even translatable for other mathematicians, as one can surmise from the article's title, "*An introduction to univalent foundations for mathematicians*". We will discuss this new program of study again a little later.

Voevodskĭ also suggested a plan for the proof of the general Bloch–Kato conjecture in 1996. This provided a road map to the eventual proof in 2011. Along the way parts of Voevodskĭ's plan were proved and parts discarded and replaced (most notably by Chuck Weibel and Markus Rost). However the entire strategy and plan were robust enough to find a way around the obstacles (albeit involving the creativity of a number of remarkable mathematicians). This is a big advantage of a community approach, led by a plan.

§ 7. — Computers and proofs.

Frenkel is quoted in [53] as saying,

> As the towering mathematical enterprise reaches new heights, with more intricately and esoterically turreted peaks, each practitioner drilling down into minutiae that could never be understood by colleagues even in the same field, computers will be necessary for advancing the enterprise, and authenticating the integrity of the structure — a structure so complex in detail and so massive in scale that adjudication by a mere human referee will no longer be feasible.

There are three main uses of computers in proofs:

• They can be used for calculations in establishing a proof;

• They can be used to assist in verifying the logic of an author's arguments, perhaps interactively. These are called "computer-assisted proofs";

• They can be used to prove claimed theorems, so-called "computer-generated proofs".

We now briefly discuss these in reverse order.

© 2023 $M\times\Phi$ vol. I.2

Computer generated proofs. They are in their infancy. Common views include:

> I don't believe in a proof done by a computer ... I believe in a proof if I understand it. — Pierre Deligne

> I'm not interested in a proof by computer ... I prefer to think. — John H. Conway [53]

> Computer-generated proofs...teach us very little[(55)]. — Vladimir Voevodskĭ [53]

Since these top researchers do not want to believe a claim to be true just because the computer says so, we might counter these remarks by insisting that a computer-generated proof should be designed to be human-readable, and to attempt to *explain* the "why" as well as the "how". We will return to this viewpoint later in discussing recent developments.

Kahle [36] claims that once a computer finds and describes a proof it is relatively easy to verify it is correct, even if that is tedious (so a good job for a computer). However finding a proof is hard. The idea of "trying every possibility" suffers from well-known complexity problems but there are strategies to improve upon this, and we will discuss a different approach of Ganesalingam and Gowers.

Computer assisted proofs. Russell and Whitehead exhibited, in their arduous task, that for anything beyond a very trivial result, the number of logical inferences in a formal proof is too large to be adaptable, that the whole quest seems to only be of interest, in of itself. So it was not too sad when Gödel put paid to their original purpose. However formal proofs are making a comeback! No longer are they "roped-off museum pieces to be silently appreciated, but not handled directly" ([28]). Now we have computers that have the memory space to handle the length of more-or-less any proof, and (we hope) the logical resources to ensure that no steps are omitted (though this raises questions about trust in computer calculations, as discussed below).

In the future proof-verification might employ "computer-assisted proofs", since the author could interactively explain her

[(55)] And goes on to say "If we go toward computer-generated proofs then we lose all the good that there is in mathematics — mathematics as a spiritual discipline, mathematics as something which helps to form a pure mind."

© 2023 $M\times\Phi$ vol. I.2

proof to an appropriately designed proof-checker. Indeed in 2008 Harrison [33] at Intel, wrote that one of his goals for formalization is

> Supplementing, or even partly replacing, the process of peer review for mainstream mathematical papers with an objective and mechanizable criterion for the correctness of proofs.

Even now this would involve an inordinate drain on the author's time, but it might nonetheless be useful in situations where there are many new definitions that must be correct, for example in the works of Biss and Voevodskĭ that we discussed above.

We must surely be wary of believing in computer verified proofs for the old reason that we are translating mathematics into a specialized language.[56] To justify this one can resuscitate Peano's belief that in an appropriately designed language one can eliminate mistranslations and obtain "precision" so as to eliminate mistakes. However, even if this is possible and you have achieved this dream, how would you prove that you have succeeded?

In Voevodskĭ's work on his univalence foundation program he felt he no longer trusted himself on the details and designed a proof assistant. In [53] he claims that a proof-assistant can keep you honest: you lay down the plan, it builds the boring details. When it can't you have to refine your plan further; like working with a mythical pedantic enthusiastic colleague. Roberts [53] writes that Voevodskĭ was

> jousting with the computer. He instructs it to try this, try that

rather like playing a video game.

Calculations as part of proofs. When computers were first able to produce inordinate amounts of data and information that might be used in a proof, there was some struggle to decide what to require of an author for the reader to believe in the proof. The consensus now is that the author must produce a coherent explanation and

[56] In fact proof verification software is a spin-off of hardware verification software, and software now can formally verify that high level computer languages or microprocessors (or anything in-between) operate as claimed. For example Leroy [41] created a formally verified compiler for the C programming language.

$M\times\Phi$ vol. I.2 © 2023

justification of the calculations, and describe them so that the interested reader can easily repeat them independently.[57]

§ 8. — Uses of computers in major theorems.

Most proofs include calculations by the author. Indeed authors routinely claim properties of the objects they are working with, without feeling a need to justify them in detail, since anyone in the field will know or at least expect these things. These are not definitions but claims that the object worked with is an example of something familiar and therefore obeys the properties we expect of such objects. There is no precise rule about the amount of detail that an author needs to give: It depends on what she expects of her readers, and what she believes her readers expect of her! For example, one might assert that a given polynomial has no rational roots without further explanation, or one might state "by Gauss's rational root criterion", or one might work through the finite number of possibilities that Gauss's criterion yields. Sometimes the amount of calculation required to justify such a claim is substantial and so the author may feel compelled to explain how the calculations are done, so as to be helpful for a reader wishing to check the results, rather than give the data, which might not be enlightening at all.[58]

We begin by discussing examples where extraordinary calculations have been part of proofs, and then how proof assistants have subsequently helped.

Short gaps between primes. In Maynard's work on small gaps between primes [45] he needs to construct a sieve with certain properties which he showed follows from constructing a special

[57] As self-evident as this might seem now, one can find "philosophical discussions" of computer verification in the literature, both by research mathematicians and philosophers, which express discomfort in accepting such calculations, even with appropriate protocols. It is always initially difficult to accept technological change: For example, in 1994, a few years after the world wide web became ubiquitous, there was an MSRI workshop on "the Future of Mathematical Communication" to discuss its potential. Like now, the internet contained a lot of unedited nonsense and many of the more senior participants at the workshop seemed skeptical that one could create an online journal which maintained the standards of print journals. Even when the obvious solution was proposed (using referees), several remained unconvinced.

[58] Especially when there may be a large number of cases to verify for a simple property.

$M\times\Phi$ vol. I.2 © 2023

polynomial of degree $\leqslant 11$ in 105 variables, so that some function of this polynomial is > 4. To verify Maynard's proof one can follow his steps precisely, or one can veer off his path at some stage and see if one can get to the same conclusion via an alternative route.[59] By explaining things well, he thus gave the reader the opportunity to verify his game plan, and to infer that his proof is robust. The details of the calculation would have been less helpful than a description of the thinking behind each step.

God's number is 20. This refers to the number of twists needed to resolve Rubik's cube from an arbitrary starting position. That is, no matter how Rubik's cube is jumbled when you start you can get it back to the starting position in 20 moves, and there are positions that require 20 moves (as shown by Michael Reid in 1995). This claim was made by Morley Davidson et al in [12] in 2010. There are about 4.3×10^{19} possible positions of Rubik's cube. With a little group theory the collaborators reduced this to resolving about 56 million positions in 20 moves or less, and then used 35 CPU-years of idle computer time (donated by Google) to nail these down. Finding the 20 moves is a much bigger proposition than following the 20 moves already written down, so this is not so hard to check, and check independently, using their output file.[60]

The Classification of Finite Groups. This is another example of an overarching, robust plan, followed by an enormous number of parts to be filled in. In this case even the plan was and is of such extraordinary depth and complexity that only a few seemed to be confident of it (though this has improved over time). Filling in the steps was always going to be a massive job, with many participants and all sorts of room for error. However here many of the steps are interesting in their own right so that more people were willing to be part of this great project. Advances meant that new structures emerged and the plan got modified to explain more while becoming technically simpler. One senior participant, Michael Aschbacher [1, 2] honestly remarks that at times he has believed that the classification

[59] Having spent many years thinking about similar questions, I was skeptical when Maynard first told me he had got the constant > 4. I not only got the same answer proceeding as Maynard suggested, but also tried modifications since I worried that his persuasive writing had concealed a mistake.

[60] In practice a typical PC might be able to do 10^{15} basic computer steps (like adding two digits) in a day, which means that checking $< 10^8$ positions should be easy in a day, but more than 10^{14} impractical in a year.

$M\times\Phi$ vol. I.2 © 2023

has been complete, at other times not, and he is always certain that there are minor errors throughout. He believed in the robustness of the plan, that a full classification would be found along the lines claimed, albeit with some modifications, sometimes substantial. As yet the classification has not been computer verified.

The Four Colour Theorem (4CT). There is a lot of (philosophy) literature expressing epistemic dismay at the original proof of 4CT, some of which misunderstood what happens in the purported proof(s), so let me try to clarify. The key notion is, again, "robustness". In all proofs the idea is to show that if a counterexample exists to 4CT then one can "reduce" it to find a "minimal counterexample" belonging to some finite computable set, and then to prove that no such minimal counterexamples exist since they would have to have too special properties (or be further reducible).[(61)] This has, to date, been complicated.

In Appel and Haken's 1976 original proof of 4CT, their discussion of how to reduce is not entirely well organized and in parts is difficult to verify. They needed to show that 1478 different subgraphs can be reduced further; they used a computer but without convincing documentation. Moreover the computer algorithm, as described, is complicated (involving 487 steps) and difficult to verify as valid and non-self-contradictory. Nonetheless Appel and Haken did create a convincing plan for proving 4CT in which one could see the point of each major step and why it should work. Here the devil was in the details, and even if these were implemented correctly they were always going to be hard to verify, believe and build upon.[(62)] There is a robustness to the overall plan but the implementation was unconvincing.

In 1996, four leading graph theorists, Robertson, Sanders, Seymour and Thomas [55] (RSST) decide to rework the existing proof to make it more believable. They followed the same overarching plan as Appel and Haken but looked for simplifications in the implementation. They had to computer reduce just 633 subgraphs (still a large number, but the authors went to great pains to make their construction transparent). More important is that their

[(61)] This is not an uncommon strategy in graph theory.

[(62)] I have heard it said that quite a few mistakes were identified but each could be dealt with by relatively minor modifications of the details; however there were so many such fixes that few people in the area had faith in the details of the end-product.

$M\times\Phi$ vol. I.2 © 2023

computer algorithm only involves 32 steps, so that a strong mathematician in the area could spend a day or two and believe that the authors had successfully covered all the options. This new proof was consciously written to be easily verifiable (though still long and complicated) and no serious mathematician doubted that 4CT is proved, at least by community standards. That is, this RSST proof is sufficiently robust that one expects to be able to easily patch up any misunderstanding that some future researcher might unearth.

A few years later Gonthier [22] verified the RSST proof using the Coq v7.3.1 proof assistant, by developing a "formal proof" that covers both the 32 steps and the 633 subgraphs that needed to be reduced; the proof then "depends [only] on the correct operation of several computer hardware and software components" and is not specific to this proof, which feels more robust. Most importantly many of the programs used here had been used in other calculations, which makes one feel that no bug that is specific to what is done here could have crept in. Moreover Coq can produce a proof "witness" (albeit human-unreadable).[63]

3-dimensional sphere packing. Hales' 2005 proof [27] of the Kepler conjecture on packing unit balls in 3-dimensions[64] was assisted by large-scale calculations. Like in the proofs of 4CT, Hales' strategy was to reduce a minimal counterexample to Kepler's conjecture to some finite computable set of possible arrangements of spheres which would have to satisfy some extraordinary constraints. A large calculation could then enumerate these possibilities and rule them out. However there are so many cases and the refereeing was extremely onerous leading Lagarias [39] to write

> The nature of this proof ... makes it hard for humans to check every step reliably. ... [D]etailed checking of many specific assertions found them to be essentially correct in

[63] A *witness* is a relatively short verification that a problem has been correctly solved. For example, to prove to you that I have factored 147573952589676412927, I can simply produce the factors $193707721 \times 761838257287$; you do not need to repeat the steps that led me to these factors. To verify you can simply multiply the two factors together.

[64] In 1606 Sir Walter Raleigh asked a mathematical acquaintance to help determine how many cannonballs he could stack on board a ship; by 1611 Kepler conjectured, in his paper, "On the six-cornered snowflake", that the hexagonal pattern (and a closely related alternative) used by greengrocers to stack round fruit would be the best way to squeeze as many spheres as possible into a large space.

© 2023 $M\times\Phi$ vol. I.2

> every case ... [This] produced in these reviewers a strong degree of conviction of the essential correctness of this proof approach, and that the reduction method led to nonlinear programming problems of tractable size,

a conditional but appropriate endorsement. A second proof [29] in 2017 used the HOL Light and Isabelle proof assistants. The purpose of each step is fully described in [29] but the researcher certainly still needs to take a lot on trust.

§ 9. — Computer error.

One of my biggest concerns about computer proof systems is people's tendency to assume that once a program has been well-implemented it is reliable, and then to mistake "reliable" for "free from error".

We have all dealt with computer systems which supposedly "never make mistakes", yet they make them, whether it be the computer system for your credit card, phone company, bank, airline or your home university. The reason could be programming or implementation errors, a non-understanding of the possibility of your particular situation or even a hardware problem. Are these issues avoidable? Can we correct these programs and computers to be trouble free?

Dealing with all situations that can possibly occur. Turing [62] wrote:

> It is not possible to produce a set of rules ... to describe what [to] do in every conceivable set of circumstances. One might for instance have a rule that one is to stop if one sees a red traffic light, and to go if one sees a green one, but what if by some fault both appear together? One may perhaps decide that it is safest to stop. But some further difficulty may well arise from this decision later. To attempt to provide rules of conduct to cover every eventuality, even those arising from traffic lights, appears to be impossible.

$M\times\Phi$ vol. I.2 © 2023

And then, even if a program is impressively accurate and reliable, what about when upgrades appear? If we avoid upgrades,[65] can we eventually perfect a system? And even if a system is "perfect" in that it really does respond appropriately to all situations that can arise, how would we know? (*i.e.* How do we know we have listed all feasible situations? How do you prove such a thing?)

Computer hardware reliability. The gold standard for computer hardware and software in the non-academic world is that there are no longer any complaints,[66] but this does not translate to guaranteed proof verification! Former Intel President Andy Grove said

> No microprocessor is ever perfect; they just come closer to perfection

in discussing hardware reliability. It is feasible that a small mathematical error in a widely used computer chip could be exploited to defeat widely used cryptographic protocols which would put all e-security at risk. Indeed, in 1993 Pentium released a chip which they subsequently found had a hardware bug affecting its floating point processor. Rather than recall the flawed chips they decided to keep quiet and correct the problem in updates. In June 1994, Thomas Nicely discovered the error while computing a number theory constant as precisely as possible; and subsequently it was found that the error could be detected when dividing certain seven digit integers by each other in several different softwares that used floating point arithmetic. Even then Pentium resisted a recall until IBM refused to ship their product.[67] Shall we elaborate on how this story affects our discussion?

— Perhaps there are errors in chips today that are more obscure and so less likely to be detected. Moreover if we do not independently analyse the results of large scale calculations (as Nicely did by comparing the answers he got with those that were already known) then how will we spot such subtle errors?

[65] Which has the disadvantage that one cannot take advantage of new developments. In 1990 Donald Knuth made the decision to never again upgrade TeX, no new features, only bug fixes; most scientists prefer to use newer typesetting systems like LaTeX, based on Tex, which are regularly upgraded.

[66] In other words, outputs that are obviously inconsistent with other information.

[67] Pentium claimed the recall cost them $ 475 million. One can understand why they hesitated!

$M\times\Phi$ vol. I.2 © 2023

— Are seemingly small errors in a chip's output really worth the cost of fixing for their manufacturers? Moreover manufacturers will rarely reveal concerns about their products (so as not to put off potential purchasers), so even if they knew about a fault how likely is it that that information would be widely shared?

Computer software reliability. There are also many problems with computer programs. Hales [28] notes that commercial software contains about 1 bug per hundred lines of code, and perhaps 1 per 10,000 lines when programs are super-focused on being careful.[68] He writes that

> corporations ... keep critical bugs off the books to limit legal liability ... only those bugs should be corrected that affect profit.

Moreover correcting bugs can be problematic since that process often creates new bugs.[69] Harrison [33] remarks that there may be more mistakes in a physical engineering project but since the questions there are typically continuous in nature, a small error makes little difference, whereas computer engineering acts on problems that are discrete. Although this gives one the opportunity to find exact solutions, it also means that one minor error, one mistyped digit, is more likely to lead to a major fault.

Perhaps the only (sensible) answer proposed to these particular computer problems is to repeat the same calculations on different computers run by different chips, using different software. It would then be highly unlikely to get an error at the same point in the "proof" on several different systems. If results match from two such systems that are sufficiently independent then it seems extraordinarily unlikely that there is a problem. Therefore we feel that, in studying proof verification, we can safely ignore these computer implementation problems, though it is not clear what this means for guaranteeing (i.e. proving) that a proof is correct.

[68] Like in the space program.

[69] It is one thing to make a minor change in what you are working on when it is fresh in your mind and you can see the big picture as well as the small details, but it is difficult to regain that perspective when coming back to something complicated years later to make a hopefully minor fix that one does not wish to spend a lot of time on.

$M\times\Phi$ vol. I.2 © 2023

§ 10. — Protocols for automated theorem checkers/provers.

What features would be desirable in a proof checker? The proof verifications that we have discussed above (4CT and Kepler's conjecture) worked interactively with a person to construct a proof from the ground up. Those proofs are not human readable, but they can attempt to confirm further claims made by humans, for example minor variants to improve our trust in them. Indeed the more they show that they provide understanding, by helping us go further, the more faith we might have in a computer proof.

However, is a "proof" that cannot be understood in detail, really a proof? We have discussed how the purpose of a proof is not just to establish truth, but also to enhance understanding. If it cannot do that then what use is it to the community?

Why should we trust the output from a proof verifier or a prover if we can't read them? Can proof-assistants be self-correcting if they can only be checked by their own internal logic? Indeed, it may well be that they continue to propagate a subtle error.

Surely these programs need a community to verify their proofs? Perhaps their output may be independently verified by using different programs; in effect, we propose refereeing computer proof verifiers output within their own community! This assumes that they work in a common language which adds extra burden to the different designs. In this way humans might believe a computer-verified proof, via an independent computer verification, and so the proof verifier becomes a trusted, objective, expert authority (that is, a referee). The computer programs will use their community to obtain a worthwhile seal of approval. In this way we can design the future based on what already works.[(70)]

Proof presentation. To believe in a proof that goes back to axioms, we need to have a common language so that the proof can be independently verified. We have seen that it is not feasible for a human to do the verification, and that humans gain little from this process, so how can we make it more useful? It seems evident that a proof verifier could also output a human readable proof. It could learn the types of high-level arguments that humans understand and appreciate, and then present its proof not only for human verification but also to help enhance human understanding.

[(70)] Rather like how e-journals used the refereeing process to establish, in their new context, the integrity and standards of traditional print journals.

$M\times\Phi$ vol. I.2 © 2023

The automated prover might select results from the existing library of verified results to build a short, person-readable argument to deduce the latest advance and so fit into the well-established protocols of how the community agrees on the correctness of a proof. Anyone may use any result that has been previously established. Although each step in each proof is computer verified, back to axioms, one hopes that as more researchers contribute to the system, interaction will move towards something resembling the high-level practice of mathematicians. A system like this is user-friendly and should become an integral part of the mathematician's arsenal.

Typically theorem-provers are interactive, the user being able to give it hints. The user enters statements into the proof-verifier, based on simpler objects that the machine already knows about. The proof assistant will determine whether the statement is 'obviously' true or false based on its current knowledge. If not, the user enters more details. The proof assistant therefore forces the user to explain their arguments in a rigorous way, and to fill in simpler steps than human mathematicians might feel they need.[71] For example suppose a proof needs ten lemmas. Some the theorem-prover will see and resolve quickly. Others it might be stuck on and the user gives it more details until the computer can see its way to a proof. In so-doing the program learns more, and maintains a library and is perhaps more efficient when it next encounters similar issues.

For now proof assistants can't read a textbook, they need it all to be broken down for them by humans.[72] Proof assistants can't judge whether a mathematical statement is interesting or important, only whether it is consistent with what it has been shown. It should eventually require less help, perhaps much less help. We have no idea when (and whether) it will be able to generate its own proofs.

The uncertainty principle of objective proof verification. The history of mathematical practice suggests that

The less one questions a proof, the more susceptible it is to error.

[71] Users report that they often learn a lot during the input process!

[72] And this human/machine interface can lead to problems. Indeed some definitions input into Lean by different users have been inconsistent – those we know about have been corrected. But sometimes there are ambiguities in the literature. For example Kevin Buzzard pointed out to me that *topos* can mean different things to an algebraic geometer and a logician and sometimes the difference can be quite subtle. There is no clear way to deal with this dichotomy.

$M\times\Phi$ vol. I.2 © 2023

This important principle strongly suggests one must find a wide variety of ways to explain and to verify any given proof, even a computer proof, and to look at it from as many different perspectives as possible.

§ 11. — The future of proof.

Computer generated proofs give too many details (often by a factor of more than 100) to be human readable, whereas human generated proofs give too few details to be computer verified! Hence, we have to agree on what constitutes a proof, including how much detail is required. If we accept that the computer needs to work with human-style proofs[(73)] then it will need to be able to fill in details to justify the missing steps; the key difficulty is for the computer to independently determine what that requires. A famous example is the following proof of the irrationality of $\sqrt{2}$:

> If $\sqrt{2}$ is rational then we can write $\sqrt{2} = a/b$ where a and b are coprime integers, so that
>
> $$a^2 = 2b^2.$$
>
> Therefore a is even and we can write $a = 2A$ so that
>
> $$b^2 = 2A^2$$
>
> and b is even. But then 2 divides both a and b which are coprime, a contradiction.

In 2006, Wiedijk [65] noted that no computer-verifier could take this text as input and verify it as correct.

We find that we need to return to the question of what a proof is to help our proof checker. Avigad [4] explains that a proof is a communication which provides sufficient information to establish that the purported theorem is true. Beyond correctness, it can be evaluated with respect to background and interest. All of the major 2006 computer-verifiers could have constructed a proof of the irrationality of $\sqrt{2}$ based on the above argument, but these would have been of overwhelming length, and no one but themselves could check their own proofs. Moreover, since their languages are so different they could not even check each others' proofs. Thus, as Kahle [35] notes, the issue is not only to produce human readable output but to

(73) After all, isn't that our purpose in creating machines!?

$M\times\Phi$ vol. I.2 © 2023

work out how the proof-verifier can appreciate the structure of an abstract mathematical proof, and represent that for others to read and analyze. It also needs to *explain* the mathematics, not necessarily give the shortest proof, but to be understandable, convincing and potentially easily reproducible by the reader.

Robustness and fragility. We have discussed how traditional intuitive mathematical proofs are robust in that minor flaws can be fixed, and so we can have confidence in them, even without certainty. On the other hand, formal proofs are remarkably fragile in that if we find any errors, it puts much more into doubt – once a formal proof is in any way wrong it calls into question all sorts of aspects of its formulation, particularly with our inability to read the details and therefore thoroughly review its claims.

It is not uncommon to have several seemingly different proofs of the same theorem, and it is worth asking when two are really the "same proof in disguise". Is it when the key ideas are the same? Or should we expect more to be in common? It is unclear (see [14]). Moreover sometimes the disguise is quite convincing and one cannot easily recognize the common threads. This leads one to ask whether a formalization of a given intuitive proof is going to be the same proof? When an intuitive proof is dissected into what is required for, say, Lean to work with, it will look very different and rest on a rather different looking library of knowledge. And how different will the same proof look when modified for a different language?

Buzzard gives a couple of great examples (in private correspondence):

— Leaner A[(74)] might prove Pythagoras's theorem by assuming that the triangle lives in $\mathbb{R}^2$, changing coordinates to make the right angle at the origin and then proves that

$$\int_{(0,a)}^{(b,0)} 1ds = \sqrt{a^2+b^2}.$$

Leaner B, formalizing Euclid's actual proof, takes an abstract Euclidean plane, never defines area, but rather defines what it means for two shapes to have the same area. These are not just different proofs, these are different statements of Pythagoras's theorem, which emerge from using two different (but valid) models of the Euclidean plane.

[(74)] A *leaner* is someone who implements a proof in Lean.

$M\times\Phi$ **vol. I.2** © 2023

— In Lean real numbers are defined to be equivalence classes of Cauchy sequences; but another prover might define real numbers using Dedekind cuts. So a proof of, say, the mean value theorem will look very different in the two provers, even if they must be equivalent since they must both rely on the fact that a non-empty bounded set of reals has a least upper bound.

Formalizers hope that formalizations might help us see what is necessary to use in a proof (like which axioms). The idea is to create a "proof compacter" that somehow recognizes how to shorten proofs, compacting them into as small a space as possible. The human only participates once the proof compacter cannot do more. The hope is that this "shortest possible proof" will have its advantages though I do not think it is evident why the shortest proof should use a minimal set of axioms or necessarily be advantageous.

Can computers generate their own proofs? "Machine learning" typically develops its understanding in simple ways as a result of clever algorithms. Creating a large database and analyzing it with specially formulated tools can be startlingly effective (like Google Translate or ChatGPT) but this is not the same as developing intuition (or even simulating intuition effectively). There is a lot of money and a lot of publicity surrounding the subject of "machine learning" and some other forms of "artificial intelligence" but rest assured that many hyped advances are either exaggerated or easily explained in terms of well-designed algorithms and extraordinary computing power.[(75)] There are as yet no "thinking machines".

It is difficult to know how we can move forward in this direction, as relatively little is understood about creativity and intuition, and how we move from one understanding to a rather different one. To simulate this on a machine seems very far away.[(76)] Ada Lovelace (1815-1852), who posited the concept of (what we would now call) a computer program from Babbage's early calculation machine, and even supposed that a computer could be taught to compose music, wrote

> [Babbage's] Analytical Engine has no pretensions to originate anything. It can do whatever we know how to order it to perform.

[(75)] The impact of the Google search engine is more-or-less uncorrelated to how simple (though elegant) the mathematical ideas are behind it.

[(76)] Indeed the ideas behind the latest exciting developments in machine learning do not portend any real understanding; see Melanie Mitchell's wonderful book [46] for a forensic discussion of what underlies some of these developments.

$M\times\Phi$ vol. I.2 © 2023

Before computers, librarians were often credited with knowing a lot more than they really did (as the gatekeepers of so much knowledge). Computers are much bigger repositories for knowledge, more accessible and less proscribed by others, and can achieve some surprising feats; it is not surprising they get credited with powers they do not yet possess.

Can we design a computer verifier to learn and think like a human? Ganesalingam and Gowers [19] embarked on designing an automated theorem prover that proceeds rather differently than stripping everything back to the axioms. Rather, the idea is to proceed like a human, to model the way humans think, to produce proofs that read like a human proof. In their design they require input and output that is understandable and uses the standard lexicon (to be "user-friendly"),[(77)] informative solutions that are not just verifications, and that further capability can be easily added to the program to incorporate new concepts, problem-solving techniques etc. In short one should be able to interact with the program as one does with a (marvellously retentive) human. There have been previous human-oriented programs but the machine-oriented programs have seen much more progress over the last twenty years (see section 1.3 of [19] for some history).

Machine oriented search programs tend to try many options, which has the disadvantage of combinatorial explosion but with some *tactics* (like pruning search trees) this can sometimes be well-managed. Humans have to avoid severe combinatorial explosion so bring in their tactical awareness earlier in the process. For example if a theorem to be proved has several hypotheses and conclusions the theorem prover doesn't know which is most important (and so to be focused on) and basically tries the different statements in a random order to find connections; a human might quickly examine the important conclusion and start puzzling as to how one might get there from what is known. A computer might try very similar substitutions over and over in its search tree whereas a human might see from one example that a particular type of substitution (say where y is linear in x) cannot work in general. So there are strategies that humans have that are atypical of large search strategies. As yet we do not know how to list all these differences.

In [19] Ganesalingam and Gowers remark

[(77)] It is considerably harder for a machine to understand such input than to produce such output, a bit like my experience with speaking different languages.

$M\times\Phi$ vol. I.2 © 2023

> For the majority of proofs that mathematicians find, there is some kind of 'story' to tell of the ideas that give rise to the proof. Typically, such a story will be a high-level overview of the main difficulty and how it is overcome, where 'overcome' means that the problem is reduced to one or more problems where that difficulty no longer occurs. Often this reduction is achieved by means of a well-chosen intermediate statement that turns out to follow from the initial assumptions and imply the conclusion. The intermediate statement itself is typically found not by means of a brute-force search but by a process of approximation: one might make a guess, find that it is unhelpful, understand why it is unhelpful, and use that understanding to guide the search for a better intermediate statement. These characteristically human techniques enable mathematicians to penetrate deep into 'proof space', but the set of proofs that can be discovered in this way forms a tiny fraction of that space. It seems almost a truism that human methods will be useful for programs that want to find these special proofs that human mathematicians are so mysteriously good at finding.

Humans are pretty good at selecting which technique or techniques to try and a new one can improve their efficiency at proving things. On the other hand a new technique simply expands a computer's search space and might well decrease its efficiency.

In terms of describing human proof adequately we are still at the stage that we don't know how to describe what a good proof is, but we know it when we see it. The work of both humans and machines can be defined in terms of "tactics". For now we can best use interactive systems based on "tactics" that are designed to mimic human reasoning. However it seems to be difficult to get machines to recognize which tactics to use when, that is to provide an order in which to try different tactics and perhaps to adjust the future tactics or their order depending on intuition gained from a tactic that has just failed. In [19] they ask whether to let the computer learn from its past experience, to try to devise a theory that better mimics human choices, or work with a mix of the two. It is also hard to decide when humans use certain tactics, like proof by contradiction.

Eventually we will need to play Turing's "imitation game" [62] with machine created proofs; that is, their proofs should be indistinguishable from great human proofs. So we set the machine up

$M\times\Phi$ vol. I.2 © 2023

against a Fields' medalist, ask them both a tough but doable question, and see whether we can determine whose proof is whose.[78] Turing [62] notes that since computers are universal (via the Church-Turing thesis) they can perfectly imitate anything that can be computed including human interaction, so inevitably this will be doable (and indeed computers will eventually be able to perfectly imitate each other).[79]

In [19], Ganesalingam and Gowers selected problems to prove and got thousands of independent readers to try to distinguish which proofs were by their program and which by real people (see `https://gowers.wordpress.com/2013/04/14`). The results are encouraging, though of course this is not the Turing test since it is not an independent arbiter that selected the problems.

It is hard to predict the future here. Things are moving fast and brilliant people are getting involved. Kevin Buzzard (see the next section) takes the view that

> The more people are familiar with the software, the sooner interesting things will happen

which is a compelling perspective.

© 2023 vol. I.2 $M\times\Phi$

§ 12. — The Lean Theorem prover.

The Lean Theorem prover[80] has taken the research mathematical world "by storm" in the last year or two. Lean "is a functional programming language that makes it easy to write correct and maintainable code. *You can also use Lean as an interactive theorem prover.*" It has been made popular in pure mathematics by notable Imperial College arithmetic geometry professor Kevin Buzzard (who reports on his progress on the blog [9]). He writes

[78] In the imitation game [62] an interrogator tries to distinguish between a human who tries to prove she is a human and a computer who tries to fool the interrogator into believing it is a woman. (In the original, pre-Turing, game the part of the computer is played by a mischievous male.) This is now known as the "Turing test".

[79] He also made some predictions, for example that by 2000 the imitation game will have been mostly won by computers, and that by then "one will be able to speak of machines thinking without expecting to be contradicted". He was over-optimistic but, on the other hand, he did get a lot of things correct!

[80] `http://leanprover.github.io/`

> I believe that digitising mathematics is important, for the simple reason that digitising anything enables you to do new things with it.[81]

Buzzard's initial focus was to follow the axiomatizing dream by attempting to show that all the undergraduate syllabus in pure mathematics can be justified (partly interactively) in Lean (a bit like moving through levels of a computer game). Thus many of the technicalities could be verified by Lean, leaving the student to higher level thinking. Buzzard refrains from predicting too much of what is to come from this new capability.

Buzzard is well aware that this creates the possibility of an enormous shift in the proof culture of mathematics. He involved undergraduates in creating proofs in Lean for aspects of the undergraduate mathematical canon, but confessed (to me) that he was unsure how much of the proofs these undergraduates understood (though "perhaps they understood other things that are at least as valuable"). An independent project[82] is investigating how differently these undergraduates think about mathematical proof.

Research level mathematics. Buzzard writes

> In the near future I believe that maybe computers will be able to help humans like myself (an arithmetic geometer) to do mathematics research, by filling in proofs of lemmas, and offering powerful search tools for theorems ... but there is still a huge amount of work to do before this happens.

In early June 2021 this hope was realized when Lean verified the part of an argument that recent Fields' medallist, Peter Scholze, was unsure about in his notes on Analytic Geometry [58] with Clausen. In a guest blog post [57] Scholze writes

> I find it absolutely insane that interactive proof assistants are now at the level that within a very reasonable time span they can formally verify difficult original research.

[81] And goes on to write "Currently the computer proof systems we have are not good enough to tell mathematical researchers anything new about the p-adic Langlands program [on which Buzzard is an expert] or other trendy Fields Medally things, so the top mathematical researchers tend not to be interested in them. However [this does not mean] that they will never be useful to us, and the point of [this] project is to make it happen sooner."

[82] `https://www.researchgate.net/project/Learning-about-proof-with-Lean`

$M\times\Phi$ vol. I.2 © 2023

Developing Lean's formal version of the proof involved interactions between Lean and Leaners (the people formalizing the proof in Lean), and between the Leaners and Scholze. The Leaners input the Clausen-Scholze manuscript line-by-line into Lean, which created "a clear formulation of the current goal", and then the Leaners would refer back to the manuscript to figure out how to proceed with the next few steps. If necessary they would contact Scholze for clarifications. Scholze writes:

> Sometimes it was then realized that even on paper it does not seem clear how to proceed, and the issue was brought to [my] attention ... where it was usually quickly resolved.

Thus Scholze's intuition assisted Lean (and the Leaners), which never needed to look more than one or two steps ahead to follow and formalize the Clausen-Scholze proof.

Lean did pick up imprecisions, in particular that a certain infimum need not be a minimum (which had been assumed) and required some modification of the original proof. Scholze notes that

> This was precisely the kind of oversight I was worried about when I asked for the formal verification.

More interestingly Scholze felt that he learned "what actually made the proof work". His interaction with Lean via the Leaners allowed him to make various parts of the proof more explicit (and so elementary) and to evaluate certain key constants that he had wanted to better understand. Gonthier [23] found something similar back in 2005 while developing his formal proof (in Coq) of 4CT:

> Perhaps this is the most promising aspect of formal proof: it is not merely a method to make absolutely sure we have not made a mistake in a proof, but also a tool that shows us and compels us to understand why a proof works.

Indeed in the process of his interaction with the Leaners (particularly Commelin) Scholze realised that he could get away with a weaker theorem and therefore eliminate the stable homotopy theory used originally.

$M\times\Phi$ vol. I.2 © 2023

Scholze warns that he cannot read the formalized proof; allowing him to do so is an objective of the Lean developers.[83] There are obvious issues; for example it might be "obvious" for a human that $A = B$ from the definition of A and B, yet it takes quite a while for Lean to concur.[84] Nonetheless Scholze feels that this is a

> landmark achievement ... to take a research paper and ... explain lemma after lemma to a proof assistant, until [it has] formalized it all.

Are there epistemological advantages of Lean? Is there reason to claim that Lean makes less unjustified assumptions without realizing it (as long as they appear to be consistent) than humans? It seems that the process is similar to what happens when human mathematicians verify each other's work (asking pedantic questions, wanting to know more about the definitions and to see some details filled in, which can lead to revisions). Scholze explained to me[85] that

> It was exactly the interactions with the [Lean chat] that convinced me that a proper verification was going on. To me the [fact that it was a] computer didn't matter at all, it could also have been Ofer Gabber.[86] If I [had] simply got a blanket 'This is all correct' stamp, I would have been extremely suspicious, as I was completely sure that I made some slips.

Scholze had been nervous of a very complicated and technical Theorem, and had been aware of some nasty details that needed sorting out. He wrote

> The confidence that it is correct comes from ... seeing that during the process, [the Leaners] seemed to run into all the little nasty details that I expected (and sometimes [had] not [expected]).

[83] Lean's formalization of the Clausen-Scholze proof contains tens of thousands of lines of code, about 20 times longer than the original. It is not easily human readable.

[84] Akin to the old joke in which a professor is asked whether some step in a proof really is obvious. The professor goes silent, desperately thinking, until after half an hour he replies "Yes, it is obvious".

[85] In the comments on the blog.

[86] A mathematician who is known to insist on the right details.

© 2023 $M\times\Phi$ vol. I.2

He ended up by stressing that

> This whole experiment was a genuinely human experience, very similar to going through this with a very careful colleague.

Scholze's description of the proofs themselves being (currently) presented obscurely by Lean is frustrating (so it can feel like one's less communicative colleagues), but there is no reason for that aspect of Lean to not be improved. For now we have "no sense of the terrain", only where the peaks are and whether we have reached them, so what is it that can truly be said to have been learned? Perhaps now that Lean seems like a viable proof verification assistant, its design can find a more intuitive way to describe the proof it has constructed? Often after a first human proof scopes out the terrain, the next proof gives a more enriching description of the ideas involved, so why not Lean? This would be useful and move mathematics forward whereas,

> For now, I can't really see how [Lean] would help me in my creative work as a mathematician.
>
> — Peter Scholze in Nature

© 2023 $M\times\Phi$ vol. I.2

§ 13. — Myths of objectivity.

In confirming that a proof is correct we believe that we can recognize and establish an objective truth. But can we? It is easy to believe in one's own objectivity, or that of an "unbiased machine", but are such beliefs valid, or are they self-serving? To help highlight our potential fallacy, we now recall two famous examples of so-called objectivity, (arguably) objective in their own time and their original paradigm, but certainly not now.

These self-evident truths. For a long-time in our society biblical guidance was considered to be objective. Turing [62] remarks that in the time of Galileo, the quotations

> The sun stood still ... and delayed going down about a whole day — Joshua 10:13
>
> He laid the foundations of the earth, that it should not move at any time — Psalm 104:5

were considered by many to be an objective refutation of the Copernican theory.

Social commentators like Donna Haraway [31] explore the abuse of belief in "objectivity".[87] For her primary issue, sexism, it is not hard to make the case: for so long society has tolerated a belief that one's gender or race quantifiably affects one's ability, rather than societal issues, whether they be social class, upbringing, opportunities, etc. Few today would argue that those earlier beliefs in objectivity were anything other than self-serving and most people today believe that one should aim to replace those beliefs by a more objective understanding. However, Haraway remarks that *any* proposed objective standard is really a subjective "power move, not a move towards ... truth". Indeed to say something is "objectively proven" (that is, objective within some current paradigm) seems to quash any potential objections.

Poor scholarship continues to support this kind of prejudiced "objectivity". For example, in 2005 Larry Summers, while president of Harvard University, chose to try to explain the low numbers of women in STEM jobs: At that time some test scores of cognitive abilities for twelfth grade women showed less variation than for men, indicating fewer women at the top end (as well as at the bottom end). Identifying factors that might cause these different statistics, he focussed on "issues of intrinsic aptitude" (as well as "*lesser factors* involving socialization and continuing discrimination"). His excuse for selecting "intrinsic aptitude" to be the most compelling was that he "wanted to be provocative".[88] Today, and even then, one might view such provocation as "problematic".

To go beyond highlighting bias, and to not allow the issue to become bad people's "bias" versus good people's "objectivity", feminists [31] claim that bias is ingrained into any social construct, arguing that there is a "collective historical subjectivity".[89]

Haraway [31] highlights that "parables about objectivity" are told to mathematics students during their training, even though

[87] For recent in-depth discussions on the relationship between objectivity in mathematics and social ontology, see the theme issue [10].

[88] Sadly, representative groups censured him on the grounds of sexism, rather than inadequate scholarship. This led to their protests being ignored when President Obama selected Summers to be President of the National Economic Council, thus continuing to add an air of respectability to his poor scholarship.

[89] Haraway also notes the desire to find a legitimate feminist "objectivity", but this is evidently paradoxical, in that one is in danger of repeating the same errors that one has been at pain to identify in others.

$M\times\Phi$ vol. I.2 © 2023

this is not what mathematicians do in practice. She argues that there can be no truly "trusted authorities" but rather an earthwide network of connection, respectful of different perspectives. Indeed she claims that

> Science has been about a search for translation, convertibility, mobility of meanings, and universality,[(90)]

but one fails these criteria when work is only being compared to the hegemony. In the context of this article, Haraway's work suggests that the community standards that have served us well in the past should illuminate the path forward in the rapidly developing age of computer proofs. That these standards, verifying that the plan and immediate details of a proof appear to be correct, are the closest we can get to objectivity within any given paradigm.

$M\times\Phi$ *vol. I.2* © 2023

Objectivity and infallibility. Can there be infallible proofs? Proofs that are "unconditionally and eternally" correct? Leibniz and Newton's infinitesimals were mostly accepted in the 17th century, yet were replaced by the $\epsilon - \delta$ proofs as derived by Cauchy through to Weierstrass. Will our current proofs of the basics of calculus withstand the test of eternity? Are they infallible? Will they adapt to all the questions that will be asked of them? How can we know? We have a framework within which they seem to be incontrovertible, but will that framework seem appropriate in the light of future understandings?

> A triangle. This seems to be extremely simple, and you'd think we ... know all about it ... Even if we prove that it possesses all the attributes we can conceive of, some other mathematician, perhaps 1000 years into the future, may detect further properties in it; so we'll never know for sure that we have grasped everything that there is to grasp about the triangle. And this holds also for bodies, for their extension, for everything!
> — Descartes (16/04/1648) in conversation with Burman.

And what about objective truth within the current framework? From Gödel we know we can't come close to verifying all truths since the set of true statements is far larger than the set of provable

[(90)] All issues touched on in this article, and indeed part of what we have referred to as the "community standard".

statements so we need to restrict our attention to what is provable.[91] Even so how do we know whether a given proof is correct? Tarski [60] notes that "intuitive evidence is far from being infallible, has no objective character, and often leads to serious errors", and he views the "subsequent development of the axiomatic method ... as ... restrict[ing] the recourse to intuitive evidence". This argument suggests that we can only hope to use a formal proof to obtain sound verification of an objective truth, but can a proof verifier become infallible and so be the ultimate arbiter of what is correct? Without doubt a well written program may uncover new problems in established proofs, and create proofs that are more difficult to challenge, but that is not quite the same as infallibility.

The mathematical philosopher Avigad [5] claims that

> According to the standard view, a mathematical statement is a theorem if and only if there is a formal derivation of that statement, or, more precisely, a suitable formal rendering thereof.[92]

An attractive definition of a theorem, but resting on the assumption of the infallibility of a "formal derivation of that statement".[93] Obtaining an intuitive proof and a formal proof (based on the intuitive proof) are attractive goals, different ways of assuring that there are no easily recognizable errors, but it is a stretch to believe one can, in this way, assert objective truth. Avigad's claims rest on a view of objective truth that does not reflect any consensus (despite Avigad's confident phrasing), but rather serves to justify the focus of a certain sub-community. Indeed this is endorsed by the extreme claim of his colleague Azzouni [7],

> Formalized proofs have become the norms of mathematical practice.

[91] Even more we need to restrict our attention to the (much smaller) family of statements in the complexity class NP, which loosely means those for which there is a proof or solution that can be verified on a human time scale, a key issue in theoretical computer science. This was discussed with a few details at the end of section 2.

[92] So, it seems that the Clausen-Scholze result only became a theorem that Avigad would recognize, when Lean asserted the proof to be so.

[93] Avigad goes on to claim, "When a mathematical referee certifies a mathematical result, then, whether or not the referee recognizes it, the correctness of the judgement stands or falls with the existence of such a formal derivation," a claim to an extraordinary overview of mathematical process.

© 2023 $M\times\Phi$ vol. I.2

Not quite sure how this became Azzouni's "norm" as it is so far from true for the majority of the mathematical research community. The day may come when formalized proofs play something other than a peripheral role, but that day has not yet arrived.

Detlefsen [13] better understands mathematical practice:

> Mathematical proofs are not ... generally presented in a way that makes their formalizations either apparent or routine. This notwithstanding, they are commonly presented in a way that does make their rigor clear ... at least by the time they're widely circulated among peers ...

That is, the author of a community-accepted rigorous proof rarely concerns herself with formalization, although the formalizer must surely be concerned about rigour (but it is then odd of Azzouni to suggest that the formalization makes the proof significantly more trustworthy). Avigad [5] also argues for formal proofs since "providing less information only exacerbates the problem: if even a complete presentation of a formal derivation cannot be checked reliably, providing strictly less information can hardly provide more confidence." So how much explanation is enough? Personally I prefer a clear one page proof, than a lengthy turgid treatise that fails to appreciate what is important even if it dots a few i's,"filling a much-needed gap in the literature"!

Returning to the theme of the subsection "Robustness and fragility" in section 11, it is interesting that formalizers feel their proofs are more trustworthy when they are evidently so much more fragile. Moreover if, as Avigad seems to claim, a proof is not a proof until it can be formalised then one must ask which formalization is the correct one? Different Leaners might produce very different formalizations of the proof of Pythagoras's Theorem so which should be the accepted one? Do proofs come in "equivalence classes" (that is, proofs that are the same in disguise are in the same equivalence class)? If so, what are the criteria for deciding which proof belongs to which class, and if there is more than one equivalence class then what does that infer about the fundamental nature of proof? Should machines strive for a "Book" where only the best proofs are given?

Deus ex machina literally means "god from the machine". It describes the viewpoint of some in the formalized proof community. Believing in infallibility (of their own code) leads some

$M\times\Phi$ *vol. I.2* © 2023

programmers to not appreciate that their programming might occasionally be wrong! That only by communicating one's ideas will their ideas be accepted. Indeed in [8] Barendregt and Wiedijk confidently assert that a putative proof is verified if "the small number of logical rules are always observed", and so once they believe they have done that they create an unenticing description of the formalization (see figure 8 of [8]). They need to appreciate that a proof is a social compact, and so make the extra effort to inform interested readers.

The Clay Millennium prizes. In 2000 the Clay Foundation announced a prize of a million dollars for the resolution of any of seven famous mathematical problems. A solution can receive the prize only two years after it has been published in a refereed journal, and "has achieved general acceptance in the global mathematics community". These rules leave little doubt that the framers only have faith in community-based proof verification, and even then feel that it takes a while to be sure.

§ 14. — Will machines change accepted proof?

In this article I have asserted that proof verification does little to change the central tenets of proof as a social construction. Moreover that there is little added value in learning that a program claims a proof has been verified (without providing more helpful information to increase the reader's understanding). Nonetheless we can expect that efforts will be made to make those formal proofs more accessible and hopefully useful (for example, they might eventually remove extraneous ideas from intuitive proofs). Indeed Patrick Massot [44] recently announced that software tools are being developed to automatically convert formal proofs into human-readable interactive proofs, allowing a reader to dig progressively deeper until she reaches a claim that she believes with no further explanation, whether in a formal or intuitive proof.

Since the chess program Deep Blue defeated world chess champion Garry Kasparov in 1997, machine learning programs have become increasingly good at strategy in board games, by now easily beating the best in the world at both chess and Go. Indeed in late 2017, DeepMind's AlphaZero was switched on, played only games against itself for 24 hours, and achieved a grandmaster level

$M\times\Phi$ vol. I.2 © 2023

in both games (and in shogi), making it literally "superhuman". It has a relatively shallow search tree (80,000 positions per second in chess compared to 70 million for some other top software), but compensates for that with better tactics.[94] Can such "reinforcement learning" lead to a greater depth and variety of mathematical proofs? Might it create proofs that are more surprising at times than human ones, like its chess play?

In October 2022, DeepMind introduced[95] AlphaTensor, which was set the job of multiplying together two matrices of given dimensions as efficiently as possible; that is, with the least number of multiplications. AlphaTensor improved what had been known in about 20 cases [18] but never by more than 5%, and it is particularly striking that it does not appear to have deduced any new general theorems, nor indicated how new types of general theorems might be found.[96] This indicates that this new technology can surpass human observation that comes directly through calculation in areas where there are few strong theoretical ideas, but there is no indication yet that DeepMind's algorithms will lead to the creation of new, deep theorems. Perhaps one day it will be able to better identify the reasons that its tactics are so successful, which might help better understand what is going on.

No one has yet built a quantum computer that can calculate faster than a classical computer even in specially selected questions. Nonetheless the theory suggests that certain parallelizable problems may be *much faster* to resolve on a quantum computer, most famously Shor's quantum factorization algorithm, which has caused vast resources to be pumped into "post-quantum" cryptography.[97] For us the question is whether quantum computing could be adapted to the task of finding proofs (for example, if one uses a ridiculously large search tree). Carlos Simpson pointed out to me that we might well run into the problem of obtaining a proof

[94] Chess grandmasters found its play "alien" with "insane attacking chess", for example sacrificing a queen and bishop to exploit a positional advantage, something a human would be unlikely to do.

[95] `https://www.deepmind.com/blog/discovering-novel-algorithms-with-alphatensor`

[96] In pure mathematics, finding extensive examples is often a precursor to better understanding, and thus truly new Theorems.

[97] The security of several important cryptographic protocols are based on the difficulty of factoring so if factoring becomes easy then these protocols will no longer be secure. This worries powerful people so there is a lot of money going into research in this area.

$M\times\Phi$ vol. I.2 © 2023

but having no idea how it was found, and maybe so long as to be uncheckable on any classical computer.

Raw computing power extends our mathematical capacities in many significant ways and, paraphrasing Kevin Buzzard, the more that people invest in the possibilities, the sooner interesting things will happen. Venkatesh [63] suggests that mathematics might "be greatly altered; its central questions and values ... very different from those to which we are accustomed." This "will enhance our ability to do mathematics but also will alter our understanding of what mathematics is". Indeed it is only a matter of time before we learn how to uncover tremendous possibilities for mathematics and for proofs revealed by computing power, software, and brilliant programming ideas.

§ — Bibliography.

[1] M. Aschbacher, *The status of the classification of the finite simple groups*, Notices Amer Math Soc 51 (2004), 736–740

[2] M. Aschbacher, *Highly complex proofs and implications of such proofs*, Philos Trans R Soc A 363 (2004), 2401–2406

[3] J. Auslander, *On the roles of proof in mathematics*, in B. Gold & R. A. Simons (Eds.), *Proofs and other dilemmas: Mathematics and philosophy* (pp. 61-77). Washington, DC (2008) Mathematical Association of America.

[4] J. Avigad, *Mathematical method and proof*, Synthese 153 (2006), 105–149.

[5] J. Avigad, *Reliability of mathematical inference*, Synthese 198 (2020), 1–23.

[6] J. Avigad, *Varieties of mathematical understanding*, Bulletin AMS 59 (2022), 99–117.

[7] J. Azzouni, *Why do informal proofs conform to formal norms?*, Foundations of Science, 14 (2009), 9–26.

[8] Henk Barendregt and Freek Wiedijk, *The Challenge of Computer Mathematics*, Phil. Trans. R. Soc. A. 363 (2005), 2351–2375.

[9] Kevin Buzzard *The Xena Project*, a blog at `https://xenaproject.wordpress.com/`

[10] Paola Cantù and Italo Testa (eds.) *Mathematical Practice and Social Ontology*, Topoi **42** (2023), 187–344.

[11] Eugenia Cheng, *Mathematics, morally*, 2004 (preprint).

$M\times\Phi$ vol. I.2 © 2023

[12] M. Davidson, J, Dethridge, H. Kociemba and T. Rokicki, *God's Number is 20*, https://www.cube20.org/.

[13] M. Detlefsen, *Proof: Its nature and significance* in "Proof and other dilemmas: Mathematics and philosophy" (B. Gold & R. A. Simons, Eds.) Washington, D.C.: Mathematical Association of America (2008), 3–32.

[14] Silvia De Toffoli, *Reconciling Rigor and Intuition*, Erkenntnis (2020).

[15] Silvia De Toffoli, *Groundwork for a fallibilist account of mathematics*, The Philosophical Quarterly 71 (2021), 823–844.

[16] Bjorn Ian Dundas and Christian F. Skau, *Abel Interview 2021: László Lovász and Avi Wigderson*, Notices of the AMS **69** (2022), 828–843

[17] M. Eckert, *Water-art problems at Sanssouci – Euler's involvement in practical hydrodynamics on the eve of ideal flow theory*, Physica D: Nonlinear Phenomena 237 (2008), 14–17.

[18] Alhussein Fawzi, Matej Balog, Aja Huang, Thomas Hubert, Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, Francisco J. R. Ruiz, Julian Schrittwieser, Grzegorz Swirszcz, David Silver, Demis Hassabis and Pushmeet Kohli, *Discovering faster matrix multiplication algorithms with reinforcement learning*, Nature 610 (2022), 47-63.

[19] M. Ganesalingam and W.T. Gowers, *A fully automatic problem solver with human-style output* J. Automat. Reason. 58 (2017), 253–291.

[20] Kurt Gödel, *The modern development of the foundations of mathematics in the light of philosophy*, draft of a 1961 presentation to the American Philosophical Society, Collected Works, Vol III, Oxford (1995), 375–388.

[21] Dan Goldston and Cem Yildirim, *Small gaps between consecutive primes*, (retracted preprint)

[22] Georges Gonthier, *A computer-checked proof of the Four Colour Theorem* (2005), preprint.

[23] Georges Gonthier, *Formal proof – The four-color theorem*, Notices Amer Math Soc 55 (2008), 1382–1393.

[24] W.T. Gowers, *The Two Cultures of Mathematics*, in "Mathematics: Frontiers and Perspectives?, Amer Math Soc, 2000.

$M\times\Phi$ *vol. I.2* © 2023

[25] Daniel R. Grayson, *An introduction to univalent foundations for mathematicians*, Bull. Amer. Math. Soc. 55 (2018), 427–450.

[26] Ian Hacking, *Why Is There Philosophy of Mathematics at All?*, Cambridge University Press, 2005.

[27] T.C. Hales, *A proof of the Kepler conjecture*, Annals of Mathematics, 162 (2005), 1065–1185.

[28] T. Hales, *Formal proof*, Notices Amer Math Soc 55 (2008), 1370–1381.

[29] T.C. Hales and 21 others, *A formal proof of the Kepler conjecture*, Forum of Mathematics, Pi 5, e2 (2017), 29 pages.

[30] Michael Hallett, *Foundations of Mathematics*, The Cambridge History of Philosophy, 1870–1945, ed. Thomas Baldwin (Ch. 10 of Part I: 1870–1914), Cambridge University Press (2003), 128–156, 833–837.

[31] Donna Haraway, *Situated knowledges: The science question in feminism and the privilege of partial perspective*, Feminist studies, 14 (1988), 575–599.

[32] Michael Harris's blog: *Mathematics without apologies*, June 2nd, 2018, `https://mathematicswithoutapologies.wordpress.com`

[33] John Harrison, *Formal proof – theory and practice*, Notices Amer Math Soc 55 (2008), 1395–1406.

[34] John Hosack, *An inclusive philosophy of mathematics*, Notices Amer. Math. Soc. 66 (2019), 1433–1437.

[35] Reinhard Kahle *Towards the Structure of Mathematical Proof*, CEUR Workshop proceedings vol 1186 (2014), paper # 22.

[36] Reinhard Kahle *What is a proof?* Axiomathes 25 (2015), 79–91.

[37] Donald Knuth *All Questions Answered*, Notices of the AMS, 49 (2002), 318–324.

[38] Izabella Laba *Photography*, in "Art in the life of mathematicians" (ed. Anna Szemerédi), American Mathematical Society, 2015.

[39] J. Lagarias, *The Kepler Conjecture and its Proof*, (Springer, 2009), 3–26.

[40] M. Lange, *Because Without Cause: Non-causal Explanations in Science and Mathematics*, 2017, Oxford: Oxford University Press.

$M\times\Phi$ *vol. I.2* © 2023

[41] X. Leroy, *Formal certification of a compiler back-end, or: programming a compiler with a proof assistant*, 33rd ACM Symposium on *Principles of Programming Languages*, ACM Press (2006) 42–54.

[42] Paolo Mancosu, *Mathematical Explanation: Problems and Prospects*, Topoi 20 (2001), 97–117.

[43] Paolo Mancosu, *Explanation in Mathematics*, The Stanford Encyclopaedia of Philosophy (2018), Edward N. Zalta (ed.), `https://plato.stanford.edu/archives/sum2018/entries/mathematics-explanation/`

[44] Patrick Massot, *Formal mathematics for mathematicians and mathematics students* lecture on youtube.com (`https://www.youtube.com/watch?v=tp_h3vzkObo`).

[45] James Maynard, *Small gaps between primes*, Ann. of Math. 181 (2015), 383–413.

[46] Melanie Mitchell, *Artificial Intelligence: A Guide for Thinking Humans*, Farrar, Straus and Giroux, 2019.

[47] L. J. Mordell, *On the rational solutions of the indeterminate equations of the third and fourth degrees*, Proc. Cab. Phil. Soc 21 (1922), 179–192.

[48] L. J. Mordell, *Reminiscences of an Octogenarian Mathematician*, The American Mathematical Monthly 78 (1971), 952–961

[49] Melvyn B. Nathanson, *Desperately seeking mathematical truth*, Notices of the AMS (2008), 773.

[50] Melvyn B. Nathanson, *Desperately seeking mathematical proof*, Math. Intelligencer 31 (2009), 8–10.

[51] J. Paris, and L. Harrington, *A Mathematical Incompleteness in Peano Arithmetic* In Barwise, J. (ed.). Handbook of Mathematical Logic. Amsterdam, Netherlands: North-Holland, 1977.

[52] Colin J. Rittberg, Fenner S. Tanswell and Jean-Paul Van Bendegem, *Epistemic injustice in mathematics* Synthese 197 (2020), 3875–3904

[53] Siobhan Roberts, *In Mathematics, Mistakes Aren't What They Used To Be*, Nautilus online, 2015.

[54] Dan Rockmore, *Prove It!*, The New York Review of Books, Jan 13, 2022.

$M\times\Phi$ *vol. I.2* © 2023

[55] Neil Robertson, Daniel P. Sanders, Paul Seymour, and Robin Thomas, *A new proof of the four-colour theorem*, Electron. Res. Announc. Amer. Math. Soc. 2 (1996), 17–25.

[56] Dirk Schlimm, *Peano on Symbolization, Design Principles for Notations, and the Dot Notation*, Philosophia Scientae, 25 (2021), 95–126.

[57] Peter Scholze, *Half a year of the Liquid Tensor Experiment: Amazing developments*, `https://xenaproject.wordpress.com/2021/06/05`

[58] Peter Scholze (with Dustin Clausen), *Lectures on Analytic Geometry*, `https://www.math.uni-bonn.de/people/scholze/Analytic.pdf`

[59] Peter Scholze and Jakob Stix, *Why abc is still a conjecture* (preprint, 2018).

[60] Alfred Tarski, *Truth and Proof*, Scientific American 220 (1969), 63–77.

[61] William P. Thurston, *On proof and progress in mathematics*, Bull. Amer. Math. Soc. (N.S.) 30 (1994), no. 2, 161–177

[62] A. M. Turing, *Computing Machinery and Intelligence*, Mind, New Series, 59 (1950), 433–460.

[63] Akshay Venkatesh, *How we place value in mathematics*, preprint.

[64] Vladimir Voevodskĭ, *What If Current Foundations of Mathematics Are Inconsistent?* Lecture at `https://www.ias.edu/ideas/2012/voevodsky-foundations-of-mathematics`, 2010.

[65] Freek Wiedijk, *Formal proof – getting started*, Notices Amer Math Soc 55 (2008), 1408–1414.

$M\times\Phi$ vol. I.2 © 2023

Andrew Granville,
Départment de Mathématiques et Statistique,
Université de Montréal, CP 6128 succ Centre-Ville,
Montréal, QC H3C 3J7, Canada.

Annals of Mathematics and Philosophy
***vol.* I, *no.* 2**, 2023

$M \times \Phi$

Fregean abstraction in Zermelo-Fraenkel set theory: a deflationary account

Joel David Hamkins

Abstract. The standard treatment of sets and definable classes in first-order Zermelo-Fraenkel set theory accords in many respects with the Fregean foundational framework, such as the distinction between objects and concepts. Nevertheless, in set theory we may define an explicit association of definable classes with sets $F \mapsto \varepsilon F$ in such a way, I shall prove, to realize Frege's Basic Law V as a ZF theorem scheme, Russell notwithstanding. A similar analysis applies to the Cantor-Hume principle and to Fregean abstraction generally. Furthermore, because these extension and abstraction operators are definable, they provide a deflationary account of Fregean abstraction, one expressible in and reducible to set theory — every assertion in the language of set theory allowing the extension and abstraction operators εF, $\#G$, αH is equivalent to an assertion not using them. The analysis thus sidesteps Russell's argument, which is revealed not as a refutation of Basic Law V as such, but rather as a version of Tarski's theorem on the nondefinability of truth, showing that the proto-truth-predicate "x falls under the concept of which y is the extension" is not expressible.

I am grateful to Paddy Blanchette for insightful discussion and for pointing me to several relevant passages in Frege and grateful as well for insightful audience commentary at the talks I gave on this material in Oxford, Pavia, New York, and Paris. Commentary can be made about this article on my blog at http://jdh.hamkins.org/fregean-abstraction-deflationary-account.

© 2023

§ 1. — Introduction.

The treatment of sets and classes in Zermelo-Fraenkel set theory instantiates in many respects the principal features of the Fregean foundational framework, in which Frege distinguishes sharply between *objects* and *concepts*. The realm of objects, a universe of individuals, constitutes the domain of discourse, a domain over which the quantifiers may range, and this domain is considered under the guise of diverse Fregean concepts, each serving in effect as a predicate on that domain, picking out a collection of those objects, for a given object either falls under a concept or does not. Frege analyzes concepts through their course of values, viewing them in effect as functions mapping every object either to an object standing for truth or one for falsity.

The usual practice and development of sets and classes in Zermelo-Fraenkel set theory accords very well with this Fregean framework, for set theorists consider the cumulative set-theoretic universe V, the collection of all sets, as constituting a domain of individual objects, the sets, about which we make mathematical assertions, with our quantifiers $\forall x$ and $\exists x$ ranging over all the sets in this universe, and we consider these sets in the light of the various set-theoretic properties $\varphi(x)$ expressible in the language of set theory, forming the corresponding definable classes $\{\, x \mid \varphi(x) \,\}$, thereby undertaking predication by those properties. Thus, we have a universe of first-order objects, the sets, in the context of diverse Fregean concepts, predicates on that domain, the classes. We refer to or name such a concept or class $\{\, x \mid \varphi(x) \,\}$ by providing the definition φ along with any parameters, if any, used in that definition. Different intensional descriptions φ or ψ may end up describing the same class extensionally, if $\varphi(x) \leftrightarrow \psi(x)$ holds for every individual x. In this way, we may come to view the usual understanding of sets and classes in ZF set theory in a Fregean light.

Andreas Blass expresses this view, that proper classes are not part of the set-theoretic ontology, but rather are merely a convenience for predication.

> Proper classes are not objects. They do not exist. Talking about them is a convenient abbreviation for certain statements about sets. (For example, V=L abbreviates "all sets are constructible.") If proper classes were objects, they should be included among the sets, and the cumulative hierarchy should…continue much

$M\times\Phi$ vol. I.2 © 2023

> farther, but in fact, it already continues arbitrarily far. ...Considerations like these are what prompt me to view ZFC as a reasonable foundational system, in contrast to Morse-Kelley set theory. [Bla11]

Thus, Blass appears to work in precisely the set theory I have described, a set-theoretic realm having only sets as objects, and considering classes only as a shorthand manner of undertaking definable predication.

In his well-known treatise on large cardinals, Kanamori [Kan04] similarly works in this theory. Notably, he regards the famous Kunen inconsistency, asserting that there is no nontrivial elementary embedding $j : V \to V$, as "a schema of theorems, one for each j." That is, he considers the theorem as a scheme of statements making separate claims for each possible definition of such a class j.[(1)]

Thus, it appears that mainstream set theory is often undertaken in the Fregean style I have described, where the universe of sets constitutes a first-order realm of objects, and the various definable classes occupy the second-order realm of concepts. The classes "do not exist" in the sense that they are not themselves objects — they are not sets — but rather the classes serve as predications of the objects. Set theorists nevertheless use and refer to classes, of course, in that same Fregean manner, just as Frege uses and refers to concepts. In short, in set theory we have sets as objects and definable classes as concepts.

In this article, I aim to investigate set theory in this Fregean light, to see how well we can implement Frege's ideas in Zermelo-Fraenkel set theory. Let me emphasize, however, that my project here is not to elucidate Frege's exact system in set theory, but rather to be inspired by Fregean ideas as a way of understanding what is going on in set theory. I aim to investigate Fregean-inspired set theory, not to present a set-theoretical Frege.

$M \times \Phi$ *vol. I.2* © 2023

[(1)] I find it curious for Kanamori to emphasize this point with this particular theorem, however, in light of the fact that one may easily refute the existence of nontrivial *definable* elementary embeddings $j : V \to V$ with a soft minimality argument showing that no particular definition (allowing parameters) can admit a smallest critical point of such an embedding; see [Suz98; Suz99]. This soft argument furthermore makes no use of the axiom of choice, while for the class version this is a major open question leading to the so-called choiceless large cardinals.

A quick review of terminology. Before proceeding to the main argument, let me quickly review some of the terminology I shall be using, which is all standard. By ZFC I refer to the usual Zermelo-Fraenkel axioms of set theory, including the axiom of choice, expressed in the first-order language of set theory, whose only nonlogical symbol is the binary membership relation $x \in y$. This is a first-order theory, whose models have the form $\langle M, \in^M \rangle$, where M is a domain of individuals, the sets of the model, and $x \in^M y$ is the membership relation on those sets. The first-order quantifiers $\forall x$ range over the elements x in that domain M. The theory without the axiom of choice is denoted ZF.

Gödel-Bernays set theory GB (sometimes denoted NGB), in contrast, is expressed in the second-order language, allowing for both sets and classes, with models of the form $\langle M, \in^M, \mathscr{X} \rangle$, where $\langle M, \in^M \rangle$ is a model of Zermelo-Fraenkel set theory, and the second-order part $\mathscr{X}$ is a family of classes $X \subseteq M$. The first-order quantifiers $\forall x$ range over the elements of M, while the second-order quantifiers $\forall X$ range over the classes $X \in \mathscr{X}$. The theory GB posits the first-order class comprehension axiom, asserting that $X = \{ x \mid \varphi(x, z, Z) \}$ is a class for any set parameter z and class parameter Z, where φ is any assertion involving only first-order quantifiers. This is expressible by a single second-order axiom asserting that the classes defined by atomic assertions exist and the classes are closed under intersection, complements, and projections; it follows that GB is altogether finitely axiomatizable. The theory GB expresses the replacement axiom in the form: the image of a set under a class function is a set, and we use GBc to denote the theory with the axiom of choice for sets, while GBC is the theory with the (strictly stronger) global choice principle, asserting the existence of a global well order of the universe.

In the main theorem, I shall be using in effect the models of GBc that arise from a given model $\langle M, \in^M \rangle$ of ZFC, by equipping the model with all and only its first-order definable classes (allowing parameters). That is, we consider $\langle M, \in^M, \mathscr{X} \rangle$, where $\mathscr{X}$ consists of the classes $X \subseteq M$ that are definable $X = \{ x \in M \mid M \models \varphi(x, z) \}$ in $\langle M, \in^M \rangle$ by some first-order formula φ with parameter $z \in M$. This is easily seen to be a model of GBc, and furthermore since every model of ZFC can be expanded in this way to a model of GBc, it shows that GBc is conservative over ZFC for assertions about sets. A corresponding statement is also

$M\times\Phi$ vol. I.2 © 2023

true about GBC by adding also a generic class well-ordering of the universe, which can be done by forcing without adding any sets.

The theory GBC is altogether weaker than Kelley-Morse set theory KM, which has the second-order class comprehension axiom, asserting that $\{\, x \mid \varphi(x, z, Z) \,\}$ is a class, where z and Z are set/-class parameters, and where now φ is allowed to have second-order quantifiers $\forall X$, ranging over the classes $X \in \mathscr{X}$ of the model. This axiom is expressed as a scheme of statements in the second-order language.

§ 2. — Basic Law V.

One of Frege's central ideas was that we might associate every concept F with an object εF, the *extension* of F, an object that names or represents this concept. In this way, Frege is able to treat concepts in effect as objects by means of their extensions — he can quantify over them and refer to them using the objects that represent them.

The *extension* terminology carries today the connotation of a collection — the extension of a concept is nowadays commonly understood in set theory as the set or class of fulfilling instances of that concept. Nevertheless, this collection connotation was not necessarily part of Frege's conception. The extension object εF of a concept F was simply an object of some kind that served to name or represent the concept and needn't itself be understood as a set, class, or collection.

Importantly, Frege insisted that the assignment of extension objects to concepts should be a conceptual invariant — the extensions of equivalent concepts should be the same, and the extensions of inequivalent concepts should be different. In other words, εF should represent the concept F extensionally, rather than intensionally, and this is expressed by his Basic Law V.

Basic Law V. *The extension of a concept F is the same as the extension of concept G if and only if $Fx \leftrightarrow Gx$ for every x.*

$$\varepsilon F = \varepsilon G \quad \Longleftrightarrow \quad \forall x\,(Fx \leftrightarrow Gx).$$

Let me refer to the problem of finding a system of extension assignments $F \mapsto \varepsilon F$ fulfilling Basic Law V as the *extension-assignment problem*. Are we able in general to define extension objects for all our concepts in a way that fulfills this conceptual invariance requirement?

$M\times\Phi$ vol. I.2 © 2023

Basic Law V is an instance of the more general process of *Fregean abstraction*, by which we move from a notion of equivalence to the abstraction of that equivalence, a classification invariant for that equivalence. With Basic Law V, concepts F and G are logically equivalent in the sense that they have the same course of values $\forall x\ (Fx \leftrightarrow Gx)$ if and only if they have the same corresponding abstraction, that is, the same extensions of those concepts $\varepsilon F = \varepsilon G$. Similarly, according to the Cantor-Hume principle, considered in section 8, concepts F and G are *equinumerous* in that there is a one-to-one correspondence of the instances of F with instances of G if and only if they have the same corresponding abstractions, that is, the same number of elements $\#F = \#G$. And more generally, for any notion of equivalence $\sim$, the corresponding Fregean abstractions αF would be classification invariants in the sense that $F \sim G$ if and only if $\alpha F = \alpha G$.

In this article, I shall provide deflationary accounts for all of these various Fregean abstraction principles, providing ZF definitions of suitable extension objects εF, number objects $\#F$, and the various abstraction objects αF for any concept of equivalence, and prove as ZF theorem schemes that they fulfill the desired classification invariant requirements expressed by Basic Law V, the Cantor-Hume principle, and Fregean abstraction generally. I describe this account as deflationary because the abstractions I provide are expressible and defined in set theory, and as a consequence they are entirely eliminable — every assertion in the language with the abstraction operators, I shall prove, is equivalent to an assertion in the plain language of set theory without them.

§ 3. — Basic Law V holds in ZF set theory.

Let me prove the main result, namely, that Basic Law V holds in the set-theoretic Fregean framework I have described. I shall define an explicit association of definable classes F with sets εF representing them in such a way that the association $F \mapsto \varepsilon F$ is definable in set theory in various senses and such that Basic Law V is fulfilled. The theorem thus realizes Basic Law V as a provable theorem scheme in Zermelo-Fraenkel set theory.

$M\times\Phi$ vol. I.2 © 2023

Theorem 1. *In Zermelo-Fraenkel* ZF *set theory, there is an explicit association of definable classes F with sets εF, such that:*

1. *The association $F \mapsto \varepsilon F$ is definable in the following senses:*

(*a*) *For any explicitly given class $F = \{ x \mid \varphi(x, z) \}$, the extension object is first-order definable in set theory, using the same parameter z as provided in the description of F:*

$$\forall y \, \big(\, y = \varepsilon F \quad \textit{if and only if} \quad \theta(y, z) \, \big).$$

(*b*) *For any class I and class function $i \mapsto F_i$ for $i \in I$ and classes F_i, presented uniformly as a definable class $F = \{ (i, x) \mid i \in I, x \in F_i \} \subseteq I \times V$, the function $i \mapsto \varepsilon F_i$ is first-order definable in set theory, again with the same parameters used to define F.*

(*c*) *The full mapping $F \mapsto \varepsilon F$ is second-order definable in any model of* ZF *equipped with its definable classes:*

$$\forall F \forall y \, \big(\, y = \varepsilon F \quad \textit{if and only if} \quad \Theta(F, y) \, \big)$$

2. *The association $F \mapsto \varepsilon F$ fulfills Basic Law V:*

$$\varepsilon F = \varepsilon G \quad \textit{if and only if} \quad \forall x (Fx \leftrightarrow Gx).$$

Proof. To prove the theorem, I shall begin by explaining how to define the extension objects εF. Suppose that we are presented with a particular definable class $F = \{ x \mid \varphi(x, z) \}$, the class of all sets x fulfilling the first-order formula $\varphi(x, z)$, using a fixed set parameter z. To have provided a class concept means to have provided such a formula φ and parameter z explicitly. I shall associate this class F with a certain set εF in such a manner so as to fulfill all the claims of the theorem. Technically, I shall be providing for each formula $\varphi(x, z)$ a definable mapping $z \mapsto y$, where $y = \varepsilon F$ will be the extension of the class $F = \{ x \mid \varphi(x, z) \}$ defined by φ with that parameter z. We cannot in general take the class $\{ x \mid \varphi(x, z) \}$ itself as the extension, since this might be a proper class, not a set, but we need εF to be a set. Instead, what I propose is to use the data of the presentation itself to define εF.

Let us fix in the metatheory a particular enumeration $\psi_0, \psi_1, \psi_2, \ldots$ of the formulas of the first-order language of set theory. The formula φ appears somewhere in this enumeration. Let ψ_n be the earliest formula in that enumeration that is capable, with some choice of parameter, of defining the same class that φ

© 2023 vol. I.2 $M \times \Phi$

defines with parameter z. That is, ψ_n is the earliest formula on the list for which there is some set p for which $\forall x\big(\varphi(x,z) \leftrightarrow \psi_n(x,p)\big)$. We may find such a parameter p of minimal possible rank in the set-theoretic hierarchy. Indeed, let u be the set of all such parameters p of that minimal rank that define with that formula ψ_n the same class as φ does with parameter z. We now define εF be the pair $(\ulcorner\psi_n\urcorner, u)$, where $\ulcorner\psi_n\urcorner$ is the set coding the formula ψ_n. In short: the extension of a concept is the earliest formula on the list that is capable of defining it with some parameter, together with the set of minimal-rank parameters that do so.

Let us now prove that εF is first-order definable in the senses mentioned in the theorem. Suppose that we are presented with a class $F = \{\, x \mid \varphi(x,z) \,\}$, presented intensionally — that is, we have been provided explicitly with the defining formula φ and the parameter z. The formula φ itself appears as some ψ_n in the list, and there are only finitely many preceding formulas $\psi_0, \ldots, \psi_n = \varphi$ in the enumeration. To express the property $y = \varepsilon F$, therefore, we may take a disjunction over $i \leqslant n$ of the assertions that say,

> "y is a pair $(\ulcorner\psi_i\urcorner, u)$, whose first coordinate is the code of the formula ψ_i and whose second coordinate u is a nonempty set of parameters $p \in u$ for which
>
> $$\forall x\ \big(\varphi(x,z) \leftrightarrow \psi_i(x,p)\big), \text{ and}$$
>
> and furthermore u is the set of all such minimal-rank p for which this is true, and finally, none of the formulas ψ_j for $j < i$ admit such a parameter."

This property exactly describes what it means for y to the be the extension object εF as we defined it above, and since we have the formula φ as a metatheoretic formula in hand and there are only finitely many preceding formulas ψ_i, all of this is expressible by a formula θ in the first-order language of set theory

$$y = \varepsilon F \quad \iff \quad \theta(y,z).$$

Thus I have established claim (1a) as a ZF theorem scheme.

Next, suppose that we are provided with a definable class functional $i \mapsto F_i$, associating every object i in a class I with a class F_i. We assume that this class functional has been provided intensionally, by means of a uniform defining formula φ and parameter z, such that

$$x \in F_i \quad \iff \quad \varphi(i,x,z).$$

© 2023 vol. I.2 $M\times\Phi$

The uniformity of this definition means that it simply reduces to case (1a), for we can simply treat i as another parameter, thereby defining the extensions

$$y = \varepsilon F_i \quad \iff \quad \theta(y, i, z)$$

using the same analysis as above. This provides a uniform definition of the class function

$$i \mapsto \varepsilon F_i.$$

Thus, whenever we have a concept for a function from objects to concepts, we also have a concept for the extensions of those concepts, establishing (1b) as a ZF theorem scheme.

For statement (1c), let us assume we work in a model of ZF equipped for the Henkin semantics with all and only its definable classes, a model of Gödel-Bernays set theory GBc (not necessarily with the global choice principle). I claim that the map $F \mapsto \varepsilon F$ is second-order definable in this context. Here we assume that we are given the class F as a class only, without knowing the formula and parameter by which it was defined. Nevertheless, with a second-order assertion ranging over the definable classes, we shall be able to recover the defining formula and thereby recognize and provide the extension object εF. We may assume that the meta-theoretic enumeration of formulas $\psi_0, \psi_1, \psi_2, \ldots$ that we used in the definition is also available internally to the object theory, arising as the standard part of a definable enumeration. Note that for standard k, there is a definable Σ_k-truth predicate. That is, we have first-order definable classes providing these partial truth predicates. Using these classes, we can express the relation $y = \varepsilon F$ as follows:

> y is a pair $\left\langle \ulcorner \psi_n \urcorner, u \right\rangle$, where ψ_n is a Σ_k formula for some k, such that there is a class T that is a Σ_k truth predicate, u is the nonempty set of minimal-rank parameters p such that $\forall x \big(Fx \leftrightarrow T(\ulcorner \psi_n \urcorner, \langle x, p \rangle) \big)$, and no preceding formula ψ_i has this property.

The key point is that since F is definable, it is defined by some formula ψ_n having some complexity Σ_k, and because the least such k will be standard there will be a definable Σ_k truth predicate class T available, and with this class we can express truth for $\psi_n(x, p)$ and the fact that no earlier ψ_i succeeds in defining F. So we have proved (1c).

Finally, let me prove that this definition fulfills Basic law V. We had defined εF to be $\left\langle \ulcorner \psi \urcorner, u \right\rangle$ where ψ is earliest formula to define

$M \times \Phi$ vol. I.2 © 2023

F and u is set of minimal-rank parameters that do so. Thus, the extension object εF depends only on the class F considered as a predicate, and not on the original presentation of F as defined by a formula and parameter $\varphi(x,z)$. If we had used a different formula and parameter, but thereby defined the same class, then the extension object εF would still be picking out the least formula ψ_n in the metatheoretic enumeration that is able to define that class, and u would again be the set of minimal-rank parameters that work with this formula to define F. Therefore, different equivalent presentations of the definable class will get assigned the same object:

$$\varepsilon F = \varepsilon G \quad \text{if and only if} \quad \forall x(Fx \leftrightarrow Gx).$$

And this is precisely what it means for Basic Law V to be fulfilled. □

Notice how the method exhibits Frege's distinction between identity of concepts and extensional equivalence of concepts. Namely, we can often describe a class in two distinct manners $\{\, x \mid \varphi(x) \,\} = \{\, x \mid \psi(x) \,\}$, that is, using different formulas φ and ψ or perhaps using the same formula with different parameters. So we have given different names to this class, by picking it out by means of different descriptions, and in this sense we have different concepts, but those concepts have the same extension in the case that $\forall x\ \big(\varphi(x) \leftrightarrow \psi(x)\big)$, and in this case we shall have $\varepsilon\{\, x \mid \varphi(x) \,\} = \varepsilon\{\, x \mid \psi(x) \,\}$, or perhaps it is better to write it simply as $\varepsilon\varphi = \varepsilon\psi$.

I described in theorem 1 an interpretation of extension objects for unary classes $\{\, x \mid \varphi(x,z) \,\}$ defined by a formula $\varphi(x,z)$ with one free variable x ranging over the defined class and the other variable z held constant as a parameter. In the most general case, however, one would naturally want to consider higher-arity formulas $\varphi(x_1,\ldots,x_k,z_1,\ldots,z_r)$ defining a k-ary relation in the first k variables, using the other variables z_i as constants or parameters in the definition. For this one could introduce a formalism $\varepsilon_{x_1,\ldots,x_n}\varphi$ to indicate exactly which variables are taken as part of the defined relation, with the others held as parameters in the definition (and Frege has an analogous notation for exactly this kind of case). Without saying much more about the details of such a system, it is clear that the argument of theorem 1 works essentially just the same with this more general treatment.

What I claim about theorem 1, furthermore, is that it is deflationary. Every use of the extension operator ε in the semantics that

$M\times\Phi$ **vol. I.2** © 2023

are provided is entirely eliminable — every assertion in the language of set theory expanded to allow the extension operator $\varepsilon\varphi$ and more generally the higher-arity versions $\varepsilon_{x_1,\ldots,x_n}\varphi$ is equivalent to an assertion in the base language of set theory alone.

Corollary 2. *In the semantics for extensions of definable classes $\varepsilon\varphi$ provided by theorem 1, including the higher-arity versions $\varepsilon_{x_1,\ldots,x_n}\varphi$, every use of the extension operator ε is eliminable. That is, every assertion in the expanded language of set theory allowing the extension operator ε is equivalent to an assertion in the language of set theory. Furthermore, one can find the ε-free translation by a primitive recursive translation function.*

Proof. One can prove this by induction on formulas, using statements (1a) and (1b) of theorem 1. One can systematically eliminate $\varepsilon\varphi$ from formulas φ not involving ε. The point is statements (1a) and (1b) provide ε-free identity criterion for $y = \varepsilon\varphi$, and so any assertion made about this extension object can be made equivalently without need for any ε expression. The same idea works in the higher-arity case $\varepsilon_{x_1,\ldots,x_n}\varphi$. The translation process is clearly primitive recursive, being defined by a syntactic recursion on formulas. □

It follows immediately that Basic Law V, expressed in the language expanding the language of set theory with the extension operator $\varepsilon\varphi$, is a conservative extension of Zermelo-Fraenkel set theory in the base language of set theory. No new facts about sets will be proved by adopting the $\varepsilon\varphi$ formalism using the semantics I have defined.

This conservativity observation is essentially related to the conservativity of the Hilbert-Bernays epsilon operator (see [AZ20]), but there is an important difference: that operator is not definable or eliminable in the way I have just proved for $\varepsilon\varphi$ in corollary 2. My operator is eliminable, deflationary, but the Hilbert-Bernays epsilon operator is not.

One might want to consider the quality of the extension assignments in any proposed solution to the Basic Law V abstraction problem. The extension objects I provided in the proof above fulfill the biconditional requirement of Basic Law V, but there is no claim here of any canonicity. To be sure, perhaps one regards the abstraction requirement itself as expressing all that matters concerning a solution to the abstraction problem — every solution might be taken as equally good or valid as another.

$M\times\Phi$ vol. I.2 © 2023

Nevertheless, some may find a given solution dissatisfying or unnatural. My standard response when someone objects to a given solution as being unnatural is to inquire precisely which further features are desired? It is the duty of those who object on naturality grounds to articulate a precise account of what counts as natural. In the case of the extension-assignment problem, what further properties should we seek that are not expressed by Basic Law V?

Let me defend my extension assignment solution as natural in several further respects. First, my extension objects carry the natural information that one would expect when naming or presenting classes. Namely, whenever we present a particular class concept, we do so by presenting the class $\{ x \mid \varphi(x,z) \}$ with a particular defining formula φ and possibly a parameter z. And the extension object I define for this case similarly consists of a pair $\langle \psi, u \rangle$, where ψ is such formula (possibly simpler than φ) and u is a set of such parameters (of minimal rank), defining the very same class. Indeed, the formula ψ might be the very same formula φ used in the presentation, if that one had already been optimal. And conversely, every extension object $\langle \psi, u \rangle$ for a class naturally provides the means for us to present the class in that natural way, namely, presenting it as $\{ x \mid \psi(x,p) \}$ for any particular $p \in u$. These are senses in which my solution is natural.

But furthermore, more importantly, my solution is also natural in fulfilling the definability features achieved in theorem 1 statement (1). This is precisely the deflationary aspect of my solution. I shall argue in section 5 that we should naturally count such deflationary definability requirements as amongst the further desirable features we would want to see in any attractive solution of the extension-assignment problem — the extension assignments become almost useless or incoherent without this. This definability feature of my solution is missing in the prior art discussed in section 4.

In section 8 we shall see further instances where one seeks to place further naturality requirements on an abstraction problem. Namely, in connection with the Cantor-Hume abstraction principle for equinumerosity, we shall discuss that not every solution of the cardinal-assignment problem is also a solution of the cardinal-selection problem, whereas the latter solutions seem more desirable.

§ 4. — **Prior art.**

$M\times\Phi$ **vol. I.2** © 2023

Let me mention some relevant earlier work in the same direction as the project of this article. Terence Parsons [Par87] showed the consistency of the first-order fragment of Frege's system by providing a model in which Basic Law V is fulfilled for first-order definable classes. He begins with the observation that every countable model in a countable language admits, of course, only countably many definable classes. Given a countably infinite model, he divides the domain (externally to the model) into infinitely many infinite classes, reserving these objects to be assigned as extensions εF for the definable classes F in a series of stages stratified by the ε depth of the definitions. That is, at the first stage, using the first class of objects, he assigns extensions for the definable classes that are defined by formulas not involving extensions at all; at the next stage, using the next class of objects, he handles classes defined by formulas that make reference to the extensions, but only the extensions of extension-free formulas; and so on. The main point is that at each stage, all the relevant extension objects for a given formula to be successfully interpreted will have been already assigned at the earlier stages, and in this way he is able to assign all the extension objects εF in turn, realizing Basic Law V.

John L. Bell [Bel94] strengthens the Parsons result with a very clear argument, expanding the language Henkin-style with constants c_A for each formula A, in such a way that $c_A = c_B \leftrightarrow \forall x(Ax \leftrightarrow Bx)$, which is to say that the objects c_A serve as conceptually-invariant extensions for the classes A. See also his later work [Bel00], detailing a theory for the systematic labeling of classes.

John Burgess [Bur98] carries out the Parsons argument a little more explicitly, using compactness to achieve consistency in expanded language. Burgess [Bur05, p. 89] mounts a similar cardinality argument using Löwenheim-Skolem to reduce from full second-order to Henkin semantics with countably many classes.

I greatly admire this earlier work, but what I should also like to mention about it is that none of these prior arguments provide a deflationary solution of the extension-assignment problem in the sense I have claimed for my solution in theorem 1. In the Parsons construction, for example, the labeling of concepts with objects is performed from outside the model, as though by the hand of God (or at least the hand of Parsons), and similarly in effect for the Bell construction and the compactness arguments. These methods consequently do not provide a uniform definability of the identity

$M\times\Phi$ vol. I.2 © 2023

criterion, as in (1a), (1b), and (1c) of theorem 1. In the case of Bell's construction, of course one gets the expressibility of $y = \varepsilon A$ in the expanded language simply by using the formula $y = c_A$, but in general there is no guarantee that the individual named by c_A is definable in the original language (and indeed the original model may have no definable elements), and the method does not achieve a uniform identity criterion with respect to parameters. That is, if one defines classes A_i uniformly by a formula $A_i(x) \leftrightarrow \varphi(x, i)$, then there will generally be no uniform way to express $y = c_{A_i}$ in the manner of (1b) of theorem 1, since each parameter i leads to a syntactically different constant c_{A_i}, and these constants cannot all appear in one formula.

§ 5. — The identity criterion and definability.

I find it frankly strange to neglect the identity criteria and definability questions arising with the extension-assignment problem, for it seems to me that having an expressible identity criterion should be taken as a core part of what it means to solve the extension-assignment problem. Not only must there be an assignment of extension objects εF to concepts F, but one should also be able to recognize which objects are extensions, which objects go with which concepts, which concepts go with which objects and so on.

In order actually to use the extensions εF, after all, to form the extensions for our concepts at hand and to use them in our analysis, we would want to know not just that there are objects out there, somewhere, that have been assigned to concepts in a conceptually invariant manner, somehow, with nothing more to be said about which objects are used or how. Rather, we would want to know much more about it. Which objects get assigned as extensions for which concepts? How do we recognize for a given concept F which object is its extension?

$$y = \varepsilon F$$

How do we recognize for a given object y for which concept F it is the extension?

$$y = \varepsilon F$$

$M\times\Phi$ vol. I.2 © 2023

How do we recognize whether a given object y is the extension of any concept?

$$\exists F\ y = \varepsilon F$$

What use will an extension assignment $F \mapsto \varepsilon F$ be, after all, if performed externally to the domain of discourse, if it is invisible within it? If the assignments are made Parsons-style from outside the model, as it were, then we cannot seem to work with the resulting solution in the original object theory. But what good is that? The only way to find out which object is the extension εF of an expressible concept F, in those systems, is to already know what it is.

In my view, the identity criterion problem for a given solution to the extension-assignment problem is akin to the famous Julius Ceasar problem. Namely, when Frege despairs at the impossibility of determining whether Julius Ceasar is a number, at bottom he is pointing out his lack of an identity criterion, his lack of a decision procedure for determining whether a given individual is a number. Similarly, a solution of the extension-assignment problem without a corresponding identity criterion will lack a decision procedure for determining whether a given object or individual is the extension of a concept.

The definability of the extension assignments amounts to having what I have referred to as a deflationary account of extensions. When we have an extension-assignment $F \mapsto \varepsilon F$ that admits an expressible identity criterion, one for which we achieve the definability properties mentioned in theorem 1, then we are in effect able to refer to the extensions of concepts in the original language and theory. Precisely because the features of the extension assignment are expressible in the base language, we have no need to expand the language to include direct reference to extensions. In short, to have a solution of extension-assignment problem that admits of definable identity criterion is exactly to be able to reduce the treatment of extensions to the original language and theory, to deflate the extensions to the original theory, as described in corollary 2.

For this reason, I find the solution of the extension-assignments provided by my theorem 1 to be an advance on the earlier work mentioned in section 4, at least for the context of set theory as the base theory. Not only can we have a solution of the extension-assignment problem realizing Basic Law V in set theory, but we

$M\times\Phi$ *vol. I.2* © 2023

can do so in a deflationist manner, so that we can recognize in the object language which objects get assigned to which concepts.

Frege builds the expressivity of the extension assignment into his syntax for the expanded language with the extension operator ε, because he allows himself to form expressions such as $y = \varepsilon\varphi(x, p)$ for any formula φ, including any formula in the expanded language. This formal assertion expresses the relation of y being the extension of the concept defined by $\varphi(x, p)$. But ultimately one must recognize that Frege provided no semantics for these expressions and did not seem to make any claims about whether his use of ε should be eliminable.

Robert May had emphasized to me in discussion at the 2022 AILA conference in Pavia that Frege had sought a "logically perfect" language, one in which every concept has a name and every well-formed name denotes a concept (see [Fre03, §28]). To my way of thinking, however, merely having a syntactic expression $\varepsilon\varphi(x, p)$ serving as a name to refer to the extension object of a given concept is quite different from knowing which object it is or whether it is possible to have a semantics for these expressions having all the properties claimed of them. Ultimately, I shall argue, Frege's syntactic approach led him astray, because it was built into his language that he could express the concept "object x falls under the concept of which y is the extension," and it is this, rather than Basic Law V, that is the true downfall of his system, as I shall explain in theorem 3.

$M\times\Phi$ **vol. I.2** © 2023

§ 6. — Reconciling Russell.

Russell is commonly described as having refuted Frege's system and he is often described specifically as having shown Basic Law V to be inconsistent. How are we to reconcile this with my claim in theorem 1 that Basic Law V is a provable theorem scheme of Zermelo-Fraenkel set theory? Wouldn't this show ZF also to be inconsistent?

To be sure, Russell's argument is most naturally construed as a refutation of the general comprehension principle.

General Comprehension Principle. *For any property φ, one may form the set of all x with property $\varphi(x)$. That is,*

$$\{\, x \mid \varphi(x) \,\}$$

is a set.

Russell's argument refutes even very simple instances of this, in the case of the formula $x \notin x$. Namely, by the general comprehension principle, we may form the set R consisting of all sets x that are not elements of themselves.

$$R = \{\, x \mid x \notin x \,\}.$$

If this is indeed a set, then $R \in R$ if and only if $R \notin R$, which is a contradiction. Thus, the general comprehension principle has contradictory instances.

Although this is often how the Russell paradox is presented, the historical difficulty is that Frege does not state the general comprehension principle explicitly as part of his system. Rather, that principle is essentially hidden away into his manner of forming concepts and denoting them. In his letter replying to Russell, Frege blames Basic Law V, yet also hints at a systemic problem:

> It seems, then, that transforming the generalization of an equality into an equality of course-of-values (§9 of my *Grundgesetze*) is not always permitted, that my rule V (§20, p. 36) is false, and that my explanations in §31 are not sufficient to ensure that my combinations of signs have a meaning in all cases. I must reflect further on the matter. It is all the more serious since, with the loss of my rule V, not only the foundations of my arithmetic, but also the sole possible foundations of arithmetic, seem to vanish.
>
> –Frege, 22 June 1902 letter to Russell [Hei67, p.127]
>
> responding to Russell's fateful letter

Let me explain how Russell's argument can be used to refute Basic Law V in combination with a specific comprehension-like principle.

Theorem 3 (Russell)**.** *Basic Law V is inconsistent with the existence of the concept "x does not fall under any concept of which x is the extension."*

Proof. Assume Basic Law V and let R be the concept "x does not fall under any concept of which x is the extension." If this is indeed a concept, then let $r = \varepsilon R$ be the extension of it, then by Basic Law V we see that any concept of which r is the extension will agree extensionally with R, and so r falls under R if and only if it does not, which is a contradiction. □

$M \times \Phi$ vol. I.2 © 2023

The main point here is that Russell's argument is not at its core a refutation of Basic Law V as such, but rather a refutation of a certain concept-formation principle, which is an instance of class comprehension. It would suffice for Russell's argument if we had the *falling-under* concept: object x falls under the concepts of which y is the extension. In this case, we could negate and project onto the diagonal to form the forbidden concept mentioned in theorem 3.

In fact, the argument does not require the full biconditional of Basic Law V, but only the forward implication, namely, that if r is the extension of a concept, then that concept agrees with R. For example, it would suffice that every concept F had a distinct extension εF. This situation falls far short of Basic Law V, and it is similar to the situation of every formula φ having a distinct Gödel code $\ulcorner\varphi\urcorner$, but we don't generally insist that equivalent formulas get the same code. The point is that Russell's argument shows that even this weak form of the law is inconsistent with the existence of the concept "x falls under the concept of which y is the extension." In this sense, Russell's argument has little to do with Basic Law V as such.

I should like to emphasize that the falling-under concept amounts to a proto-truth predicate. That is, to say that object x falls under the concept of which y is the extension is exactly to say that that concept is true of x. In short, the extension-assignments εF are a kind of Gödel-coding for predicates F, and the falling-under concept is the corresponding truth predicate.

The true mistake of Frege's system, in my view, is not Basic Law V as such, but rather the fact that Frege has built this proto truth predicate into his syntax. Namely, one can easily express in Frege's system that x falls under the concept of which y is the extension by the assertion:

$$\exists F\,(y = \varepsilon F \text{ and } Fx).$$

Thus, Frege has made the falling-under concept available simply as a matter of notation. And he makes pervasive use of it throughout his work. He uses the falling-under concept, for example, when defining what it means to be a natural number: a natural number is any object that falls under every property that holds of 0 and is passed from every object to its successor.

Perhaps Frege looked upon the idea of "falling under" as innocent, it being merely a special case of functional application, which Frege took as fundamental. That is, if one understands concepts as functionals mapping objects to *true* or *false*, then for an object x to

$M\times\Phi$ vol. I.2 © 2023

fall under a concept F is simply to compute the value of $F(x)$. If this is *true*, then x falls under F, and if it is *false*, then it does not.

To be sure, the difficulty of the falling-under relation is not when it is used like this in a particular case with a particular concept F, for one can express that x falls under the concept F simply by Fx. Rather, the source of all the trouble is the falling-under relation as a relation of objects to objects, taking the latter objects to represent concepts, as in the relation "object x falls under the concept of which y is the extension." This uniform treatment of the falling-under relation enables Russell's refutation as described in theorem 3 and is what amounts to a truth predicate.

§ 7. — Proving Tarski's theorem directly from Russell.

$M\times\Phi$ vol. I.2 © 2023

Let me explain next how this kind of Russellian reasoning can be used to provide a direct proof of Tarski's theorem on the nondefinability of truth. Since Gödel's incompleteness theorem can be seen as a consequence of Tarski's theorem, this argument therefore also provides a proof of Gödel's theorem directly from Russell. In short, I would like to explain how truly close Russell was to proving Gödel's theorem. I find this remarkable, because contemporary accounts of Tarski's theorem usually invert this presentation, commonly proving Tarski's theorem only after Gödel's, and indeed often treating Tarski's theorem essentially as an afterthought to Gödel.

Tarski's theorem is commonly described in arithmetic or set theory, given a particular encoding of formulas φ with objects $\ulcorner\varphi\urcorner$, as the assertion that there can be no formula $T(y,x)$ in the same language expressing the truth of those formulas, that is, fulfilling every instance of the truth conditions:

$$\forall x\big[T(\ulcorner\varphi\urcorner,x)\leftrightarrow\varphi(x)\big].$$

A typical proof of Tarski's theorem proceeds by arguing that if there were such a truth predicate, then by the Gödel fixed-point lemma there would be a sentence asserting its own nontruth according to the predicate, and then by the Liar-paradox reasoning this sentence would be true if and only if the predicate declared it not to be true, contradicting the truth-predicate requirement.

That is fine, but allow me here to state a rarified version of Tarski's theorem and prove it directly by Russell's method, bypassing Gödel, with no explicit use of self-reference or the fixed-point

lemma. In particular, this formulation of the theorem avoids the need to consider any encoding of syntax whatsoever. We could state the theorem for arithmetic or for set theory, but in fact the argument applies to any first-order theory at all. It is a totally elementary argument.

Theorem 4 (Tarski)**.** *In any first-order theory, truth is not definable. That is, there is no formula $T(y,x)$ in the same language, such that for every formula $\varphi(x)$ it is consistent with the theory that there is some y for which*

$$\forall x\big[T(y,x) \leftrightarrow \varphi(x)\big].$$

Proof. Given any formula $T(y,x)$ with two free variables, let $R(x) = \neg T(x,x)$. This is a perfectly good formula with one free variable. There can be no r for which $\forall x\,\big[T(r,x) \leftrightarrow R(x)\big]$, however, since by its definition $R(r)$ asserts $\neg T(r,r)$, which contradicts $T(r,r)$. So there is no formula T such that every formula can arise consistently with the theory as a section of T. □

Let me elaborate in the case where we have fixed a representation of formulas $\ulcorner\varphi\urcorner$, which as I had mentioned is a common way of talking about Tarski's theorem. Given any proposed truth predicate T, let $R(x) = \neg T(x,x)$, which on the truth-predicate understanding of T could be seen as expressing, "it is not the case that x represents a formula that is true of x." Let $r = \ulcorner R\urcorner$ and form the sentence $R(r)$, which asserts $\neg T(r,r)$. But this contradicts $T(r,r)$, which is the same as $T(\ulcorner R\urcorner, r)$, and so $T(\ulcorner R\urcorner, r)$ is not equivalent to $R(r)$, which violates the desired equivalence. So there can be no such definable truth predicate T.

I find this argument to be a direct parallel of Russell's original argument against Frege. The proof here is essentially identical to Russell's proof of theorem 3, simply using $\ulcorner R\urcorner$ in place of εR. The Russell formula $R(x)$ asserting that x is not a formula that is true of itself is directly analogous to the Russellian concept of all concepts that do not apply to themselves, or the set of all sets x that are not members of themselves.

A very similar manner of presenting Russell's argument as a proof of Tarski's theorem is also provided in [Fit17], and accords with arguments of Smullyan, who had emphasized that much of the fascination surrounding Gödel's theorem may be more properly directed at Tarski's theorem. The inevitable road from Russell to Gödel is the related similar theme of Kripke's [Kri14].

© 2023 $M\times\Phi$ vol. I.2

Of course, one can find key elements of Gödel's argument here. The construction of the sentence $R(r)$, for example, where $R(x) = \neg T(x, x)$ and $r = \ulcorner R(x) \urcorner$, is exactly what ultimately appears in the usual proof of Gödel's fixed-point lemma, and so the resulting sentence in this Russellian argument is the same as what one finds in the Gödelian fixed-point proof of Tarski's theorem. In that sense, the two approaches to proving Tarski's theorem are not very different. Nevertheless, the straightforward Russellian treatment above is simpler than the Gödelian proof and entirely avoids confusing issues of self-reference. In any case, I find it far more natural to view Gödel's fixed-point lemma as generalizing Russell's method here, rather than viewing Russell's argument as an application of Gödel's lemma, which is how Tarski's theorem is often presented. Surely Gödel found his proof of the fixed-point lemma by contemplating and generalizing Russell's argument.

I have stated Tarski's theorem in a two-dimensional version, ruling out a definable satisfaction relation rather than sentential truth. But one also commonly finds a sentential version of the theorem, stating that there can be no arithmetically expressible predicate T such that

$$T(\ulcorner\sigma\urcorner) \leftrightarrow \sigma$$

for every sentence σ, using the standard Gödel coding of sentences $\ulcorner\sigma\urcorner$. Of course arithmetic satisfaction and sentential truth are closely related in arithmetic, since one expects that a formula φ is true at a number x if and only if the sentence $\varphi(1 + 1 + \cdots + 1)$ is true, using the term $1 + 1 + \cdots + 1$ with x summands. The point is that PA proves that every number x is the value of such a term, possibly nonstandard, in a nonstandard model.

Apart from that, however, the sentential formulation of Tarski's theorem can also be proved along Russellian lines, but the proof will require that we have an encoding of syntax in the object theory, rather than merely in the metatheory as above, and this admittedly makes the argument closer to Gödel. Namely, suppose that T is a sentential truth predicate. Let us form the Russell formula $R(x)$ asserting "x is the code of a formula φ that is false at the point x." This can be expressed as

$$R(x) \quad =_{\text{def}} \quad \neg T(\ulcorner \varphi(\overbrace{1 + 1 \cdots + 1}^{x}) \urcorner).$$

The term $1 + 1 + \cdots + 1$ with x summands is formed and then substituted into φ all in the object theory — this is where we use that

© 2023 vol. I.2 $M\times\Phi$

syntactic operations can be undertaken in the object theory, in order to know that R is expressible. Now let $r = \ulcorner R \urcorner$, and consider the sentence $R(r)$, meaning $R(1 + 1 + \cdots + 1)$ with r summands. By the definition of R the sentence $R(r)$ asserts $\neg T(\ulcorner R(r) \urcorner)$, but this contradicts the truth predicate property of T. So there can be no such sentential truth predicate.

Perhaps when thinking of this sentential version, some people argue that Tarski's theorem relies fundamentally on the use of Gödel codes $\ulcorner\varphi\urcorner$ and the arithmetization of syntax, whereas arithmetization is rightly considered a core contribution of Gödel. I am in complete agreement on the profound nature of Gödel's arithmetization ideas; see my remarks at [Ham21, p. 233]. A rebuttal to the specific point here, however, is that the core of Tarski's theorem simply requires far less coding than Gödel's. First of all, the set-theoretic version of Tarski's theorem has no need for coding specifically into arithmetic at all, since one can use set theory as a foundation of mathematics to represent essentially arbitrary mathematical constructions — including sequences and syntax and parse trees — a usage widely recognized by Hilbert and others before Gödel. And there is no need in Tarski's theorem to undertake the coding of any formal proof system. But furthermore, in the two-dimensional satisfaction form of Tarski's theorem, as I have stated in theorem 4, there is no explicit need for the coding of syntax at all. In particular, this is why the proof I gave above for theorem 4 is much simpler than the corresponding treatment in [Fit17], which aims instead at the sentential formulation. The simple fact is that Tarski's theorem in the two-dimensional satisfaction form applies in a sweeping manner to any possible representation of formulas in the object theory, even a very rudimentary representation, for which syntactic operations may not be expressible.

To my way of thinking, the idea of having an internal representation of formulas in the object theory can be traced to Frege himself. Namely, I see a very strong analogy between the extension $\varepsilon\varphi$ of a concept φ and the Gödel code $\ulcorner\varphi\urcorner$ of a formula, for in both cases we have an internal object-theory representation of a formula or concept from the metatheory. And as I have explained, Tarski's theorem already applies with such a representation, even without any additional involvement of Basic Law V. Namely, when proving Tarski's theorem we did not require that the representations respect logical equivalence, that equivalent formulas are represented by the same code.

$M\times\Phi$ vol. I.2 © 2023

Let me offer a slogan form of Tarski's theorem in a Fregean-style language:

"Truth is not a concept."

Just as Russell's argument showed in theorem 3 a sense in which the falling-under relation is not a concept, similarly the Russellian proof of theorem 4 shows that truth is not expressible. The core of the argument in both cases is about our lack of uniformity in expressing these notions. That is, in the context of a given instance, we can easily express that a given object x falls under a given concept F by asserting Fx. What we cannot do is express in a uniform manner that "object x falls under the concept of which y is the extension." Similarly, in the context of Tarski's theorem, there is no difficulty for us to express that a particular formula φ holds at a point x — we simply assert $\varphi(x)$ itself — what we cannot do is express in a uniform manner that "y represents a formula that is true at x."

Now that we have proved Tarski's theorem directly via Russell, let me deduce Gödel's incompleteness theorem as a consequence. Namely, since truth is not definable, but provability is definable (and for this one must implement the proof system in the object theory), they cannot be the same thing. QED.

In this way, one realizes the power of Tarski's theorem over Gödel's. Whereas Gödel is showing that truth is not the same as provability, a Σ_1 notion, Tarski shows the stronger result that truth is not Σ_n expressible for any n.

§ 8. — The Cantor-Hume principle.

Frege had used his extension objects εF and Basic Law V in order to prove other abstraction principles, including especially the Cantor-Hume principle, upon which he founded his arithmetic, which in many respects was the focal effort of his project. What I would like to do is explain how the ideas of theorem 1 can be used also to fulfill these other Fregean abstraction principles in Zermelo-Fraenkel set theory.

Let me begin with the central instance, the Cantor-Hume principle, also known simply as Hume's principle, which is the Fregean abstraction principle concerning numbers as a classification invariant of the equinumerosity relation. Specifically, one class X is

$M \times \Phi$ vol. I.2 © 2023

equinumerous with another Y, written $X \simeq Y$, if there is a one-to-one correspondence between them. The *cardinal-assignment problem*, which we might also call the *number-assignment problem*, is the problem of assigning a cardinal or number object $\#X$ to every class X, in such a way that it is an equinumerosity invariant, so that equinumerous classes get assigned the same number and nonequinumerous classes get assigned different numbers. The Cantor-Hume principle is the assertion that indeed there is a solution to this problem.

Cantor-Hume principle. *There is a number-assignment scheme, assigning to every class X a number object $\#X$, for which two classes get assigned the same number if and only if they are equinumerous:*

$$\#X = \#Y \qquad \text{if and only if} \qquad X \simeq Y.$$

Let me begin by mentioning several easy solutions to the cardinal-assignment problem that are available in Zermelo-Fraenkel set theory for special instances of the principle.

Consider, for example, the set-only version of the cardinal-assignment problem. This is completely solved in ZFC using the axiom of choice by the standard treatment of cardinalities in ZFC. Namely, in ZFC every set x can be well-ordered and is therefore equinumerous with some smallest ordinal, and so we may simply define the cardinality $|x|$ of the set x to be the smallest ordinal with which it is equinumerous. This not only solves the cardinal-assignment problem, but does so in a way that is simultaneously a solution of the *cardinal-selection problem*, the problem of choosing a unique representative from each equinumerosity class. That is, for any set x, the cardinality of x defined as the smallest ordinal equinumerous with x is not only an equinumerousity invariant, but it is one of the sets that is itself equinumerous with x — this ZFC definition provides a definable way to pick a canonical member in every equinumerosity class of sets.

This same solution works to an extent in ZF, without the axiom of choice, at least for the well orderable sets — for every such set x we can define the cardinality $|x|$ to be the smallest ordinal equinumerous with it.

But in fact, we do not need the axiom of choice at all to solve the cardinal-assignment problem for sets.

Theorem 5. *In Zermelo-Fraenkel set theory* ZF *without the axiom of choice, there is a definable solution of the cardinal-assignment problem for*

© 2023 $M\times\Phi$ vol. I.2

sets, thereby verifying the Cantor-Hume principle for sets. That is, there is a definable assignment

$$x \mapsto \#x$$

assigning to each set x a set $\#x$, in such a way that provides a classification invariant for equinumerosity:

$$\#x = \#y \quad \text{if and only if} \quad x \simeq y.$$

The theorem has become a folklore result in set theory, but it seems that Gödel had it by 1951, see [Göd48, footnote to Dfn 6.2, added 1951]. The main idea has become known as Scott's trick, in light of Dana Scott's discovery of it at around that same time, while an undergraduate student of Tarski's in Berkeley, but it would seem reasonable to me for it to be named the Gödel-Scott trick.[(2)] Scott's trick resolves an important issue in the definition of ultrapowers of the set-theoretic universe, and thus continually arises and rearises in large cardinal set theory, which is preoccupied with such ultrapowers as a central method, and it also arises similarly when defining the Boolean ultrapower of the universe by a forcing notion, thus becoming a key technique in the diverse mutual interpretability results provided by forcing.

The essence of Scott's trick is to handle a proper-class sized equivalence class for an equivalence relation defined on a proper class not by picking representatives of each class, which might not be possible without the global choice principle, but rather by considering the set of rank-minimal elements in each class, which will form a set and which will canonically represent that equivalence class.

Proof. The key idea is to drop the cardinal-selection aspect of the minimal-ordinal solution in ZFC, and strive instead merely for an equinumerosity invariant as required by the Cantor-Hume principle. For this, we can for any set x define $\#x$ to be the set of rank-minimal sets in the equinumerosity class of x. This is simply to use Scott's trick to select a subset of the equiumerosity class of the set. The set of rank-minimal members of the equinumerosity class determines the equinumerosity class and is invariant with respect to equinumerosity, and so it fulfills the Cantor-Hume principle for sets. □

[(2)] I corresponded with Dana Scott about his trick, and he explained that Tarski put to the class the problem of finding a set-sized invariant for a definable equivalence relation on the set-theoretic universe. Scott's minimal-rank idea solved it, and Tarski was taken with Scott's solution, dubbing it "Scott's trick," but Scott himself was sheepish about the honor, emphasizing to me his desire to be remembered for other mathematical achievements, several of which he itemized to me.

$M \times \Phi$ vol. I.2 © 2023

That seems to polish off the set version of the cardinal-assignment problem. But the Cantor-Hume principle, of course, is not just about assigning cardinals to sets, but requires us also to assign numbers to classes, and so we have not yet achieved a proof of the full Cantor-Hume principle in ZF.

To find a solution for the class case, let's observe first that if the global choice principle happens to hold, for example, if $V = \mathrm{HOD}$, then we can define a well-ordering of the universe in order type Ord, and from this it follows that all proper classes are equinumerous. In this event, therefore, there will be only one additional equinumerosity class to be handled, the proper class equinumerosity class, and so we can simply assign $\#X$ either to be the smallest ordinal equinumerous with X, if X is a set, or some default value ∞ that is not an ordinal, if X is a proper class. This will solve the cardinal-assignment problem under global choice, fulfilling the Cantor-Hume principle.

If the global choice principle fails, however, then not all proper classes are equinumerous, even in ZFC. For example, the whole universe V will not be equinumerous with the class of ordinals, since such a bijection would imply global choice.

So let me now explain a completely general solution to the number-assignment problem, for both sets and classes, one which fulfills the Cantor-Hume principle in ZF without using any choice principle at all. The main difficulty here is that we may have many different proper classes, not equinumerous, but we need to associate to each of them a cardinal number in such a way that fulfills the Cantor-Hume principle.

Theorem 6. *In Zermelo-Fraenkel set theory* ZF *with definable classes only, there is a number assignment scheme $F \mapsto \#F$, assigning to every definable class F a set $\#F$, called the number of the class F, with the following properties:*

1. *For any particular definable class $F = \{\, x \mid \varphi(x, z) \,\}$, the identity criterion for being the number $\#F$ is expressible by a formula Θ in the second-order language of set theory, using the same set parameter z, if any, appearing in the presentation of F:*

$$y = \#F \quad \text{if and only if} \quad \Theta(y, z).$$

2. *The identity criterion of whether a given object x is a number is expressible by a formula $\Lambda(x)$ in the second-order language of set theory.*

$M\times\Phi$ vol. I.2 © 2023

3. *The numbers #F are well-defined with respect to the equinumerosity equivalence of the classes, in the sense that*

$$\#F = \#G \quad \text{if and only if} \quad F \text{ is equinumerous with } G.$$

In other words, the Cantor-Hume principle holds for these number assignments.

Proof. The main difficulty, of course, is to describe how we are to assign numbers to each class. We shall provide a general procedure that works for any definable class F. Namely, suppose we are faced with a particular definable class $F = \{ x \mid \varphi(x,z) \}$, named intensionally by the formula φ with parameter z. Let us define $\#F$ to be the pair $\left\langle \ulcorner \psi \urcorner, u \right\rangle$, where ψ is the first formula on the enumeration of all formulas for which there is some z such that $\{ x \mid \psi(x,z) \}$ is equinumerous with F, and u is the set of all minimal-rank such sets z for which this is true.

Let's verify statement (1) of the theorem. Given any particular definable class $F = \{ x \mid \varphi(x,z) \}$, let $\Theta(y,z)$ assert that y is a pair $\left\langle \ulcorner \psi \urcorner, u \right\rangle$ whose first coordinate is the code of a formula ψ and whose second coordinate is the set of minimal-rank parameters p for which $\{ x \mid \psi(x,p) \}$ is a class that is equinumerous with F by a definable bijection, and such that no earlier formula ψ_i preceding ψ on the list can define with some parameter such a class equinumerous with F. This assertion is expressible in second-order set theory, in the context where the classes are all the first-order definable classes, because we can quantify over the partial truth predicates and thereby refer to the existence of a definable bijection between the classes.

For statement (2), the question of whether a given object x is the number $\#F$ of some definable class F is simply the question of whether x is a pair $\left\langle \ulcorner \psi \urcorner, u \right\rangle$ whose first coordinate is the code of a formula for which there is a truth predicate available and whose second coordinate is the set of minimal-rank parameters all defining via ψ classes that are definably equinumerous with one another, such that no earlier formula works in this way.

For statement (3), the number assignment $\#F$ clearly determines the equinumerosity class of F and depends only on that class and not on the particular presentation of F. So it will fulfill the full Cantor-Hume principle as desired. □

One can understand statement (2) as an answer of sorts to the Julius Caesar problem, for it provides a criterion deciding whether

$M\times\Phi$ **vol. I.2** © 2023

a given object x is a number on this account of number. Of course, Frege was concerned with the Julius Caesar problem in so far as that the Cantor-Hume principle, by itself, did not seem automatically to provide an identity criterion for what is a number. Here, in contrast, statement (2) is providing such an identity criterion for the specific number-assignment defined in the proof of this theorem; there is no claim, of course, that this is the only solution to the number-assignment problem and other assignments will naturally have a different identity criterion.

Perhaps one objects to this theorem in comparison with the realization of Basic Law V in theorem 1 because here we have only second-order definitions for the number identity criterion. But to my way of thinking, this objection must retreat in the face of the fact that the equinumerousity relation itself is a second-order relation. That is, the invariance property of Basic Law V is the first-order property of logical equivalence of two classes, $\forall x\,[Fx \leftrightarrow Gx]$, whereas with the Cantor-Hume principle the equivalence relation of equinumerosity $F \simeq G$ is itself asking for a definable bijection between F and G, a second-order quantification. So it is hardly surprising that the corresponding identity criterion is also second-order expressible.

$M\times\Phi$ vol. I.2 © 2023

§ 9. — Fregean abstraction in ZF.

Frege emphasized the process by which we abstract from a notion of equivalence to the abstraction of that equivalence. We have a notion of parallel lines, for example, which all have the same direction, but what is the *direction* of a line? According to Frege, it is an abstraction of parallelism. The direction of a line is an invariant of the parallelism relation — parallel lines have the same direction and nonparallel lines have different directions. Similarly, we have the notion of concepts being logically equivalent, and the *extension* of a concept is the corresponding abstraction — according to Basic Law V, logically equivalent concepts have the same extension and logically inequivalent concepts have different extensions. We have the notion of classes being equinumerous, and the *number* of a class is the abstraction — according to the Cantor-Hume principle, equinumerous classes have the same number of elements and nonequinumerous classes have a different number of elements.

Thus Fregean abstraction generalizes both Basic Law V and the Cantor-Hume principle. In the general case, we have an equivalence relation $x \sim y$, and we seek a system of abstractions $x \mapsto \alpha x$ with the property that equivalent instances get assigned the same abstraction and inequivalent instances get assigned different abstractions.

Fregean abstraction principle. *For any given equivalence relation $x \sim y$, the Fregean abstraction for this relation is a system of abstractions*

$$x \mapsto \alpha x$$

which respect the equivalence relation:

$$\alpha x = \alpha y \quad \text{if and only if} \quad x \sim y.$$

In short, Fregean abstraction for an equivalence relation $\sim$ provides a classification invariant for that relation, an assignment $x \mapsto \alpha x$ that is constant on each equivalence class and with different values on different equivalence classes.

In this section, I should like to investigate how well we might achieve the Fregean abstraction principle in Zermelo-Fraenkel set theory. Let me start as I had earlier with easier special instances of the principle. Let us begin with the easy case of a first-order definable equivalence relation $x \sim y$ on the sets.

Theorem 7. *In Zermelo-Fraenkel set theory* ZF*, suppose that $x \sim y$ is a first-order definable equivalence relation defined on a class of sets. Then there is definable system of abstractions, $x \mapsto \alpha x$, associating to each set x in the domain an object αx called the abstraction of x, which form a classification invariant for $\sim$ and thereby fulfill the Fregean abstraction principle:*

$$\alpha x = \alpha y \quad \text{if and only if} \quad x \sim y$$

Proof. In this easy case, the equivalence relation lives on sets only, and so we can simply employ Scott's trick. Namely, we define the abstraction αx of any object x in the relevant class to be the set of rank-minimal members of the $\sim$-equivalence class of x. This is an invariant for the equivalence relation, because if $x \sim y$, then they have the same equivalence class and consequently the same set of rank-minimal members of that class, and so $\alpha x = \alpha y$. And conversely, the set of rank-minimal members of an equivalence class determines that class, and so if $\alpha x = \alpha y$, then indeed $x \sim y$. So this is a full solution of Fregean abstraction in ZF for definable class equivalence relations on the sets. □

$M \times \Phi$ vol. I.2 © 2023

Model theorists will recognize this theorem as the elimination of imaginaries in ZF. This theorem covers the case of parallel lines, if one regards the lines as interpreted in set theory, and many further cases of that nature, including the case of equinumerosity for sets. In particular, this simple theorem covers the case of equinumerosity for finite classes, since these must be sets, and as a consequence this theorem provides a sufficient level of Fregean abstraction to undertake much of Frege's development of natural number arithmetic, which requires only sets rather than classes, since every finite class is a set.

But of course, several of Frege's other principal applications of Fregean abstraction arise not from equivalence relations on sets, but on classes. This is a more difficult realm, in which Scott's trick is not necessarily applicable. So let me make the step up to classes, treating first the case of a first-order expressible equivalence relation on definable classes:

$$F \approx G.$$

The instances F and G here will be definable classes in ZF set theory, and the notion of equivalence $F \approx G$ will be expressible by a formula $\varphi(F, G)$ involving only first-order quantification, in which those classes F and G appear as predicates. This is the kind of equivalence arising in Basic Law V, for example, because to express the course-of-values equivalence for concepts F and G, which is to say, that they are logically equivalent, is to say $\forall x\,(Fx \leftrightarrow Gx)$, which is therefore first-order expressible in terms of F and G.

Theorem 8. *In Zermelo-Fraenkel set theory, suppose that $\approx$ is a first-order definable equivalence relation on definable classes. Then there is a system of abstractions $F \mapsto \alpha F$, associating to each definable class F an abstraction object αF, which fulfill the Fregean abstraction principle:*

$$\alpha F = \alpha G \quad \text{if and only if} \quad F \approx G.$$

Furthermore, for any particular definable class $F = \{\, x \mid \varphi(x, z) \,\}$, the identity criterion $y = \alpha F$ is first-order expressible in set theory, using only the same parameter z, if any, appearing in the presentation of F:

$$\forall y\ \big(y = \alpha F \quad \leftrightarrow \quad \theta(y, z)\big).$$

If I is any class and $i \mapsto F_i$ is a definable sequence of classes, then the identity criterion $y = \alpha F_i$ is first-order expressible in set theory, using the same parameter used to define the sequence:

$$\forall y \forall i\ \big[y = \alpha F_i \quad \leftrightarrow \quad \theta(y, i, z)\big]$$

$M\times\Phi$ vol. I.2 © 2023

Proof. We use the same idea as in theorem 1. Namely, given any class $F = \{ x \mid \varphi(x, z) \}$, define the abstraction $\mathfrak{a}F$ to be the pair $\langle \ulcorner \psi_n \urcorner, u \rangle$, whose first coordinate is the code of a formula ψ_n that with some parameter p defines a class that is equivalent to F, that is, for which $\{ x \mid \psi_n(x, p) \} \approx \{ x \mid \varphi(x, z) \}$, and where ψ_n is the first formula to achieve this with some parameter and u is the set of rank-minimal parameters that do so with ψ_n. We don't insist that all parameters in u define the same class, but rather just that they define an $\approx$-equivalent class. This assignment will fulfill the Fregean abstraction principle, because $\mathfrak{a}F$ is determined by and determines the $\approx$-class of F, and so we will achieve $\mathfrak{a}F = \mathfrak{a}G \iff F \approx G$.

The identity criterion $y = \mathfrak{a}F$ will be first-order expressible, if we are provided with the explicit definition of $F = \{ x \mid \varphi(x, z) \}$, since there will be only finitely many earlier competitor formulas in the metatheoretic list of formulas, and to express that a given definable class is $\approx$-equivalent to F is first-order expressible.

That case immediately implies that we can also express the identity criterion $y = \mathfrak{a}F_i$ for a uniformly definable sequence of classes $i \mapsto F_i$, defined by a first-order formula $F_i(x) \leftrightarrow \varphi(x, i, z)$. The point is that this amounts simply to regarding i as another parameter. □

© 2023 vol. I.2 $M \times \Phi$

As I mentioned, this theorem has Basic Law V as a special case, since the equivalence relation of logical equivalence for definable classes is first-order expressible. In this sense, theorem 1 is a special case of theorem 8.

Let me finally treat the most difficult case, where we have an equivalence $F \equiv G$ on definable classes that is second-order expressible in Zermelo-Fraenkel set theory equipped with definable classes only.

Theorem 9. *In Zermelo-Fraenkel set theory, suppose that $\equiv$ is a second-order definable equivalence relation on definable classes, in the context where all classes are first-order definable. Then there is a system of abstractions $F \mapsto \mathfrak{a}F$, associating to each definable class F an abstraction object $\mathfrak{a}F$, which fulfill the Fregean abstraction principle:*

$$\mathfrak{a}F = \mathfrak{a}G \quad \textit{if and only if} \quad F \equiv G.$$

Furthermore, the identity criterion $y = \mathfrak{a}F$ is second-order expressible

$$y = \mathfrak{a}F \quad \textit{if and only if} \quad \Theta(y, F).$$

Proof. We may proceed just as in theorem 6. Namely, the abstraction object $\mathfrak{a}F$ is a pair $\langle \ulcorner \psi_n \urcorner, u \rangle$ where ψ_n is the first formula on

the metatheoretic list that is able with some parameter p to define a class $\{ x \mid \psi_n(x, p) \}$ that is $\equiv$-equivalent to F. This is easily seen to fulfill the Fregean abstraction principle requirement, since αF determines and is determined by the $\equiv$-equivalence class of F. The identity criterion $y = \alpha F$ will be expressible in second-order set theory in our context, because first of all, we have available the partial truth predicates sufficient to determine the truth of any particular formula $\psi_n(x, p)$, and the $\equiv$-equivalence relation was assumed to be second-order expressible. □

One must disparage the fact that the identity criterion $y = \alpha F$ is merely second-order expressible here, which is surely unsatisfying, but it should not be surprising at all in light of the fact that this is the case where the equivalence relation $\equiv$ itself is merely second-order definable. Theorem 9 includes the case of equinumerosity of classes, and thus has theorem 6 as a special case.

§ 10. — Gödel-Bernays and Kelley-Morse set theories.

Let us now investigate the extent to which we may achieve the Fregean abstraction principles in the various standard second-order set theories, such as Gödel-Bernays set theory and Kelley-Morse set theory. In the earlier main theorems, we had considered Zermelo-Fraenkel set theory in the context of all its definable classes, which is a model of Gödel-Bernays GB set theory (without the axiom of choice or global choice). All those arguments generalize, however, to the context of a *principal* model of Gödel-Bernays set theory, which is a model in which there is a class Z from which all other classes are first-order definable. An important example of these kinds of models arising when performing class forcing over a model of ZFC, for in such a case one gets a principal model of Gödel-Bernays set theory when taking as classes everything that is definable from the generic filter class G. Indeed, the principal models are preserved generally by tame class forcing, since one can amalgamate the new generic filter with the old class parameter.

Observation 10. *All the results of theorems 1, 6, and 9 hold not just in all models of Zermelo-Fraenkel with definable classes, but also in all principal models of Gödel-Bernays set theory, provided one allows a fixed class parameter in the definitions.*

© 2023 $M\times\Phi$ vol. I.2

The earlier proofs simply relativize, accommodating the fixed class parameter Z into the definitions of the abstraction objects. The extension εF of a class $F = \{ x \mid \varphi(x, z, Z) \}$, for example, will be the pair $\left\langle \ulcorner \psi_n \urcorner, u \right\rangle$ where ψ_n is the first formula on the list able to define F with some parameter p and class Z, and u is the set of rank-minimal such parameters p that do so. And similarly with the other Fregean abstraction invariants $\#F$ and εF for the other instances of Fregean abstraction.

The main result of [HLR13] shows that every countable model of Gödel-Bernays GBC set theory (with countably many sets and classes) has an extension to a pointwise definable model, a model of GBC in which every set and class is first-order definable without parameters. Such a model falls under the scope of the original theorem, since the first-order ZFC content of the model suffices to recover definably all the original classes. In this sense, any countable set-theoretic universe can be extended to one where there are definable solutions for Fregean abstraction.

What I should like to observe next, however, is that one cannot achieve this in a model of Kelley-Morse set theory, for Russell's argument can be implemented in KM with force.

Theorem 11. *Kelley-Morse set theory is inconsistent with a definable solution of the extension-assignment problem (even second-order definable), which is to say,* KM *is inconsistent with a definable solution to Basic Law V.*

Proof. Assume in Kelley-Morse set theory that we have a definable assignment $F \mapsto \varepsilon F$, in the sense that

$$\forall x \forall F \left[x = \varepsilon F \quad \leftrightarrow \quad \theta(x, F) \right]$$

for some formula $\theta(x, F)$ in the second-order language of set theory. It would be fine for the argument if this definition involved second-order quantifiers. It follows that the Julius-Caesar-like property of x being an extension $\exists F\, x = \varepsilon F$, is second-order expressible in Kelley-Morse set theory.

Let us also assume that this system of extension-assignment fulfills Basic Law V:

$$\forall F, G \left[\varepsilon F = \varepsilon G \quad \leftrightarrow \quad \forall x \, (Fx \leftrightarrow Gx) \right].$$

In the Russellian style, let R be the class of all sets x that do not fall under any concept of which it is the extension. That is,

$$R(x) \quad \Longleftrightarrow \quad \forall F \left[x = \varepsilon F \rightarrow \neg Fx \right].$$

$M \times \Phi$ vol. I.2 © 2023

If $r = \varepsilon R$, then $R(r)$ asserts that $\neg F(r)$ for any F for which $r = \varepsilon F$, and by the Basic Law V requirement all such F agree on whether $F(r)$ or not. But R is such an F, and so $R(r)$ if and only if $\neg R(r)$, which is a contradiction. □

Theorem 1 shows that the corresponding theorem is not provable in mere Gödel-Bernays set theory, because that theorem shows that every model of Zermelo-Fraenkel ZF set theory, when equipped with its definable classes, does admit a definable system of extension assignments $F \mapsto \varepsilon F$.

§ 11. — Fregean ontology without infinity.

Although I have provided the analysis of this paper for Zermelo-Fraenkel ZF set theory, which includes the axiom of infinity, nevertheless the entire analysis can be undertaken in finite set theory $\mathrm{ZF}^{\neg\infty}$, which replaces the infinity axiom with its negation and which includes the foundation axiom in the form of the $\in$-induction schema. I find finite set theory to be quite a nice realm for Fregean thinking.

Observation 12. *All the results of theorems 1, 6, and 9 hold in models of finite set theory* $\mathrm{ZF}^{\neg\infty}$*, equipped with their definable classes.*

The proofs go through essentially unchanged. One thing that is a bit nicer in finite set theory is that there is a definable global well-ordering — this theory proves that the whole universe is in bijective correspondence with the finite von Neumann ordinals (just as V_ω can define a bijection of itself with ω), and so we are in the easy case for definably picking representatives in every definable class.

$M\times\Phi$ vol. I.2 © 2023

§ — References.

[AZ20] Jeremy Avigad and Richard Zach. "The Epsilon Calculus". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2020. Metaphysics Research Lab, Stanford University, 2020. https://plato.stanford.edu/archives/fall2020/entries/epsilon-calculus/.

[Bel00] John L. Bell. "Sets and Classes as Many". *Journal of Philosophical Logic* 29.6 (2000), pp. 585–601. DOI: 10.1023/a:1026564222011.

[Bel94] John L. Bell. "Fregean extensions of first-order theories". *Math. Logic Quart.* 40.1 (1994), pp. 27–30. ISSN: 0942-5616. DOI: 10.1002/malq.19940400104.

[Bla11] Andreas Blass. *Are proper classes objects?* MathOverflow answer. 2011. https://mathoverflow.net/q/71773 (version 01 August 2011).

[Bur05] John P. Burgess. *Fixing Frege*. Princeton Monographs in Philosophy. Princeton University Press, Princeton, NJ, 2005, pp. x+257. ISBN: 978-0-691-12231-1; 0-691-12231-8.

[Bur98] John P. Burgess. "On a consistent subsystem of Frege's *Grundgesetze*". *Notre Dame J. Formal Logic* 39.2 (1998), pp. 274–278. ISSN: 0029-4527. DOI: 10.1305/ndjfl/1039293068.

[Fit17] Melvin Fitting. "Russell's paradox, Gödel's theorem". In: *Raymond Smullyan on self reference*. Vol. 14. Outst. Contrib. Log. Springer, Cham, 2017, pp. 47–66.

[Fre03] Gottlob Frege. *Grundgesetze der Arithmetic, Band I/II*. Verlag Herman Pohle, 1893/1903.

[Göd48] K. Gödel. "The consistency of the axiom of choice and of the generalized continuum hypothesis with the axioms of set theory". *Uspehi Matem. Nauk* (*N.S.*) 3.1(23) (1948), pp. 96–149. ISSN: 0042-1316.

[Ham21] Joel David Hamkins. *Lectures on the Philosophy of Mathematics*. MIT Press, 2021. ISBN: 9780262542234. https://mitpress.mit.edu/books/lectures-philosophy-mathematics.

[Hei67] Jean van Heijenoort. *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*. Harvard University Press, 1967, xi+660 pp. (1 plate).

$M\times\Phi$ *vol. I.2* © 2023

[HLR13] Joel David Hamkins, David Linetsky, and Jonas Reitz. "Pointwise definable models of set theory". *Journal of Symbolic Logic* 78.1 (2013), pp. 139–156. ISSN: 0022-4812. DOI: 10.2178/jsl.7801090. arXiv:1105.4597[math.LO]. http://jdh.hamkins.org/pointwisedefinablemodelsofsettheory/.

[Kan04] Akihiro Kanamori. *The Higher Infinite*. Corrected 2nd edition. Springer-Verlag, 2004.

[Kri14] Saul Kripke. "The Road to Gödel". In: *Naming, Necessity and More: Explorations in the Philosophical Work of Saul Kripke*. Ed. by Jonathan Berg. Palgrave-Macmillan, 2014, pp. 223–241.

[Par87] Terence Parsons. "On the consistency of the first-order portion of Frege's logical system". *Notre Dame J. Formal Logic* 28.1 (1987), pp. 161–168. ISSN: 0029-4527. DOI: 10.1305/ndjfl/1093636853.

[Suz98] Akira Suzuki. "Non-existence of generic elementary embeddings into the ground model". *Tsukuba J. Math.* 22.2 (1998), pp. 343–347. ISSN: 0387-4982.

[Suz99] Akira Suzuki. "No elementary embedding from V into V is definable from parameters". *J. Symbolic Logic* 64.4 (1999), pp. 1591–1594. ISSN: 0022-4812. DOI: 10.2307/2586799.

$M\times\Phi$ vol. I.2 © 2023

Joel David Hamkins,
O'Hara Professor of Logic, University of Notre Dame, 100 Malloy Hall, Notre Dame, IN 46556 USA
& Associate Faculty Member, Professor of Logic, Faculty of Philosophy, University of Oxford, UK

⋆
⋆ ⋆

Annals of Mathematics and Philosophy
***vol.* I, *no.* 2**, 2023

$M \times \Phi$

Virtues of Priority

Michael Harris

In memory of Serge Lang

§ — Introduction.

In the *Republic* Plato has Socrates insist to Glaucon that geometry, and by extension mathematics as a whole, "is the knowledge of what always is, and not of what at some particular time comes into being and passes away"[(1)]; Glaucon readily agrees. Even the most dogmatic of contemporary platonists must nevertheless acknowledge that the "knowledge" to which Socrates refers first does come into being at some particular time, and to some particular person or group of persons. Socrates and Glaucon may have been surprised had they known how aggressively many of today's mathematicians would defend their claims to *priority* in the development of new mathematical knowledge. We, on the other hand, are not surprised; we know the importance today's Academy attaches to priority in assigning the marks of prestige that constitute a mathematician's standing in the community.

The arXiv includes a timestamp with every submission, going so far as to indicate time of day as well as date of submission. Since what is posted on arXiv only counts as "knowledge" when it has been accepted by a journal for publication, journals provide additional timestamps for the timeless truths their pages contain. Publications in the prestigious *Annals of Mathematics*, for example, list four "Milestones": date of submission, date of revision, date of acceptance, and date of publication.

[(1)] Translation in G. E. R. Lloyd, "Mathematics and Narrative: An Aristotelian Perspective", in A. Doxiadis and B. Mazur, eds., *Circles Disturbed*, (Princeton University Press, 2012), p. 389, adapted from *Republic* 527ab, ed. Paul Shorey, Loeb Classical Library 276 (Cambridge, MA: Harvard University Press, 1935).

© 2023

These dates provide objective markers for adjudicating potential priority disputes, when it comes to mathematical *texts*. Priority in the creation of mathematical *ideas* is much less clear-cut. In this article I will consider a priority dispute concerning a famous *conjecture*. A conjecture is something more than an *idea*: it states a proposition that aspires to the status of a *theorem*. But it can be, and in the present instance is, considerably less than a *text*; if and when it first appears as a formal conjecture with a precise date, it may well already have been circulating informally for some time as something different from but recognizably on the path to its ultimate shape.

It is perhaps because of this liminal character of the *Modularity Conjecture* that the dispute over the priority for its creation has never been resolved. I will not attempt to resolve the dispute here. Instead, I will examine the arguments advanced in favor of each of the three attributions of priority — to Yutaka Taniyama, Goro Shimura, and André Weil. On my reading, each of the attributions is justified by the invocation of a different mathematical *virtue*. Each of these virtues, in turn, can be understood as loosely reflecting the priorities of a different school of philosophy: cartesianism in the case of Taniyama, realism or phenomenology for Shimura (and his champion Serge Lang), and Popperian falsificationism for Weil (and his champion Jean-Pierre Serre). My conclusion will be that the priority dispute cannot be settled independently, but only on the basis of the relative importance one is willing to assign to each of these virtues.

$M\times\Phi$ vol. I.2 © 2023

§ — Originality and other virtues.

If hiring committees are arbiters of mathematical virtue, then letters of recommendation should give a good sense of the virtues most appreciated by mathematicians. You will not see "proves true theorems" among them. That's merely part of the job description, and drawing attention to it would be analogous to saying an electrician won't burn your house down, or a banker won't steal from your account. I don't know how electricians or bankers recommend themselves to one another, but I have read a lot of letters for jobs and prizes in mathematics, and their language is revealing in its repetitiveness. Words like "innovative" or "original" are good, "influential" or "transformative" are better, and "breakthrough" or "decisive" carry more weight than "one of the best". Best, of course,

is "the best", but it is only convincing when accompanied by some evidence of innovation or influence or decisiveness.

When we try to answer the questions: what is being innovated or decided? who is being influenced? — we conclude that the virtues highlighted in reference letters point to mathematics as an undertaking relative to and within a community. This is hardly surprising, because those who write and read these letters do so in their capacity as representative members of this very community. Or I should say: members of overlapping communities, because the virtues of a branch of mathematics whose aims are defined by precisely formulated conjectures (like much of my own field of algebraic number theory) are very different from the virtues of an area that grows largely by exploring new phenomena in the hope of discovering simple underlying principles (like much of the field of dynamical systems). The interpretation of letters is left to the experts who apply not only their knowledge of their field's problems but also their familiarity with its value system. The same vocabulary (original, influential, breakthrough) can thus apply to developments whose character differs radically from one field to another. Some of the five mathematicians who in 2014 were the first laureats of Silicon Valley's *Breakthrough Prizes* were honored for solving longstanding unsolved problems, while others defined new classes of objects or raised original questions.

"Originality" is a virtue recognized across disciplinary boundaries. The ability to understand and explain the details of another mathematician's difficult proof is undoubtedly admired as a virtue, but it's almost never enough to get you a job. For that you have to have brought something *new* to your field. This sounds so banal to be unworthy of attention, but the philosophical problem is glaringly obvious: how do we determine what is and is not *new*? Here the word *new* is used as shorthand for the more complex notion of "bringing something new to your field". Bare newness is not enough. You can take 99 theorems in your field and string them together into a single sentence connected by "and"s but that's not *new* even if no one has ever thought to do it before. As I already hinted, most true theorems are not *new* in the virtuous sense I share with my fellow members of the hiring or prize committee.

Surely an element of taste is involved. It would be easiest to leave the clarification of the virtue of newness to specialists in mathematical aesthetics, who are in as difficult a position as any philosophers of mathematics. Specifically, they face the difficulty common to all

$M\times\Phi$ vol. I.2 © 2023

philosophers who deal with collective characteristics of communities: how to distinguish the abstract (and presumably permanent) principle they wish to emphasize from the (presumably transitory) idiosyncracies of individual members of the community? Or more crudely, is a mathematical contribution *innovative* because it is claimed to be so by influential or powerful members of the community — the kind who are asked to write letters of recommendation? Are reference letters just politics by other means?

These considerations can be left in the background when it has somehow come to be accepted that answering a certain question — solving a famous conjecture, for example — meets the community's standards of importance. No member in good standing of the community of geometers doubted that Grigori Perelman's claim to have solved the Poincaré conjecture was *new* in the relevant sense, because (a) the problem had long been considered central to the field; (b) no one had solved it before Perelman; and (c) several teams of experts worked through the proof and vouched for its correctness.

It is already more difficult to characterize the newness of a proof of a theorem that was not long-awaited. Judgments in such cases are obtained as the result of a process of implicit negotiation among those recognized as leaders within the community. On the basis of the standards invoked in such negotiations it should be possible to form a reasonably coherent (or perhaps chaotic) picture of the community's virtues.

Still more problematic are the mathematical ideas that express themselves in the form of *conjectures*. It is at least conventionally believed that the *goal* of mathematics is to produce proofs of theorems.[2] So faced with such a proof, one at least knows where the negotiation starts. The place of conjectures is much more ambiguous: they can guide the thoughts of mathematicians for generations, or even for centuries, but in principle a conjecture that turns out to be mistaken can also vanish without a trace.[3]

[2] However, William Thurston famously wrote that "The product of mathematics is clarity and understanding. Not theorems, by themselves". `https://mathoverflow.net/posts/44213/revisions`.

[3] The canonical reference for the role of conjecture within mathematical practice is still Barry Mazur's article on the subject: Mazur, B., Conjecture. *Synthèse*, vol. 111, no. 2, 1997, pp. 197–210. JSTOR, www.jstor.org/stable/20117628.

$M\times\Phi$ vol. I.2 © 2023

§ — The Modularity Conjecture: Weil's 1967 paper.

In 1967 André Weil published an 8-page article entitled "*Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen*" — "On the determination of Dirichlet series by means of functional equations". The title was an intentional allusion to a 1936 paper of Erich Hecke — "*Über die Bestimmung Dirichletscher Reihen durch ihre Funktionalgleichung*": only one functional equation is used in Hecke's title and in his paper for determination, whereas the result of Weil's paper requires a collection of functional equations. The main theorem of the paper, now known as *Weil's converse theorem*, provides conditions that guarantee that a Dirichlet series — an infinite series of the form $\sum_n a_n n^{-s}$, where s is a complex variable — arises by a well-known procedure from a *modular form* of a specified type. I will return to this last point later.

Weil must have considered his result respectable, because it was his contribution to a special issue of the journal *Mathematische Annalen* in honor of the 70th birthday of Carl Ludwig Siegel, a mathematician whom Weil greatly admired. The last page or so of his brief text was devoted to "reflections on the zeta functions of elliptic curves that perhaps also merit some attention" (*Überlegungen über Zetafunktionen elliptischer Kurven anknüpfen, die vielleicht auch einige Aufmerksamkeit verdienen*). These reflections came very close to, but stopped short of proposing the conjecture, now solved, whose name has since been a matter of controversy:

Conjecture I. The zeta function of any elliptic curve with rational coefficients coincides with the zeta function of a modular form of the type specified in the main theorem of Weil's paper.

I choose to formulate this sentence as a conjecture for the sake of clarity, and to set the conjecture in boldface type, but Weil presented his "reflections" in normal type. Moreover, instead of coming out and stating a conjecture, he explained that this is sometimes the case — here he cited a result of Goro Shimura — but added that "whether things always behave that way" appears for the moment to be problematic (*scheint im Moment noch problematisch zu sein*) and may be recommended to the interested reader as an exercise (*mag dem interessierten Leser als Übungsaufgabe empfohlen werden*).

These last sentences came at the end of a paragraph that began with Weil's very strong hints in favor of believing Conjecture I: "On

$M\times\Phi$ vol. I.2 © 2023

certain theoretical grounds one may with reasonable certainty suspect"[4] that something is true that entails the claim, in view of the main result of the paper. This "suspicion" — in a rather stronger form, mentioned in that same paragraph of Weil's paper — soon was circulating under the name of the *Weil Conjecture*. I will have to provide some background in order to explain the nature of these hints, but for the moment let me focus on the superficial aspect of the situation. Weil has drawn his readers' attention to a series of steps, some of them proved in his paper, some "reasonably certain", leading to the suspicion and, with another step that is "of course obvious to expect" (*natürlich naheliegend zu erwarten*), to the stronger form that will be explained in a moment.

§ — Lightning introduction to elliptic curves, modular forms, and their Dirichlet series.

But first I had better review the mathematical notions invoked in the "suspicion". An *elliptic curve with rational coefficients* is a geometric object E with number theoretic meaning: it is the set of solutions to a polynomial equation in two variables of the form

$$y^2 = x^3 + ax + b \qquad (\star)$$

where a and b are rational numbers (with the property that the polynomial on the right-hand side of the equation has no multiple roots). A modular form is a different kind of geometric object with a very different number-theoretic meaning. To explain this, we consider the set of complex numbers $\mathfrak{H} = \{z = x + iy, y > 0\}$– the upper half-plane in the complex plane. The 2×2 matrices

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with determinant 1 form a group $SL(2,\mathbb{Z})$ that acts on $\mathfrak{H}$ by the formula

$$z \mapsto \frac{az+b}{cz+d}$$

[4] In the original German: "*Aus gewissen theoretischen Gründen darf man mit ziemlicher Sicherheit vermuten*". The German admits more than one translation. The online dictionary `https://www.linguee.com/german-english/translation/` provides sentences in which *mit ziemlicher Sicherheit* means "almost certainly". Moreover, *Vermutung* is also the first and best-known German equivalent of the English word *conjecture,* which leads one to wonder whether Weil was being entirely sincere when in 1979 he argued against conjectures in general.

$M\times\Phi$ **vol. I.2** © 2023

Whenever Γ is a subgroup of finite index in $SL(2,\mathbb{Z})$, the quotient $\Gamma\backslash\mathfrak{H}$ can be identified with a compact Riemann surface with some points missing; adding the points yields the *modular curve* $X(\Gamma)$. A modular form of weight 2 and level Γ is a complex analytic function $f(z)$ on $\mathfrak{H}$ that is not quite invariant under Γ; rather the differential form $f(z)dz$ is invariant under Γ.

There is a procedure A for associating a Dirichlet series to such an f, and there is also a procedure B for associating a Dirichlet series — Weil denotes it Λ — to an elliptic curve E. Weil proves that any Dirichlet series that satisfies a property (A3) — the "functional equations" (plural) of Weil's title — arises from procedure A — with specified *level subgroup* Γ — and alludes to "certain theoretical grounds" for believing that the Dirichlet series arising from procedure B also satisfy property (A3). This motivates the original suspicion, which I now present in full:

Aus gewissen theoretischen Gründen darf man mit ziemlicher Sicherheit vermuten, dass Λ eine Funktionalgleichung besitzt... Weiter lässt sich vermuten [that the functional equations of condition (A3) also hold upon twist by characters].

On the other hand, E can also be viewed as a Riemann surface, and there is a stronger form of the conjecture:

Conjecture II. Any elliptic curve E with rational coefficients admits a modular parametrization: a non-trivial map of Riemann surfaces from $X(\Gamma)$ to E, again with specified Γ.

For nearly 20 years, such a map was called a *Weil uniformization*, in conferences and seminars (including many that I attended) as well as in a number of influential papers, reflecting the consensus that Weil had formulated an idea whose implications included the existence of such a map. But then things got more complicated.

§ — One conjecture, many names.

Actually, things were already more complicated within a few years of the publication of Weil's papers, because it turned out that Yutaka Taniyama had listed something very like Conjectures I and II as items 12 and 13 of a collection of 36 problems. This list was published in Japanese in his collected works, but it was handed

$M\times\Phi$ vol. I.2 © 2023

out in English at the 1955 Tokyo-Nikko conference on number theory, attended by Weil, Shimura, and Jean-Pierre Serre,[5] who also has a part to play in this story. Great things were expected of Taniyama, but he committed suicide a few years after 1955 and is now mainly remembered for the theory of complex multiplication that he developed with Shimura, and for his part in formulating the boldface statement Conjecture I, whose official name had become the *Taniyama-Weil Conjecture*, no later than 1977. By then it was generally understood that the statement would follow from the *Langlands program*, a vast and ambitious network of conjectures and theorems that linked *automorphic forms* — generalizations of modular forms — to generalizations of the number-theoretic structures derived from elliptic curves. And it was generally believed that the Langlands program was a distant fantasy, in large part because no one had any idea how to begin to prove the (then-so-called) Taniyama-Weil Conjecture — which I will henceforth call the *Modularity Conjecture*.

Positions hardened about halfway between the announcement by Frey and Ribet that Fermat's Last Theorem would be a consequence of the Modularity Conjecture, and the announcement by Wiles of his proof (later completed in a paper of Taylor-Wiles) of enough cases of the Modularity Conjecture to derive Fermat's Last Theorem. In February 1988, before Wiles announced his result, Joseph Oesterlé had published a Bourbaki seminar talk whose first theorem, attributed to Ken Ribet, asserts that "*La conjecture de Taniyama-Weil implique le théorème de Fermat*". No paper by Shimura was cited in Oesterlé's bibliography. Ribet himself, writing in the United States two years later, published a paper in France with the title "From the Taniyama-Shimura conjecture to Fermat's last theorem", in which his reference to Weil's 1967 paper draws attention to what he considered to be Weil's contribution to the story.

Finally, A. Weil proved in [35] *that an elliptic curve E over $\mathbb{Q}$ is modular provided that its L-function $L(E,s)$ has the analytic properties one expects of it.*

Mathematicians' choice of terminology for the conjecture largely broke down along national lines, and remains contested more than 20 years after a complete proof of the Modularity Conjecture was published by Christophe Breuil, Brian Conrad, Fred Diamond, and

[5] I am following Serge Lang's account in "Some history of the Shimura-Taniyama Conjecture", *Notices of the AMS*, **42**, November 1995, 1301–1307.

$M\times\Phi$ vol. I.2 © 2023

Richard Taylor.[6] Public disputes over priority were acrimonious on occasion, private disputes much more often. My own discomfort with the situation, and my dissatisfaction with the quality of the arguments put forward in support of one position or the other, led me to seek clarity and consolation in philosophy. Thus, by a circuitous route, I find myself writing an article for a philosophical journal on the topic that revived my interest in philosophy in the first place.

§ — One conjecture, many virtues: comparing the cases for Taniyama, Shimura, and Weil.

In fact, although the difference of opinion had all the characteristics of a vulgar personality conflict — it was common knowledge that relations between Serre and Shimura were difficult, and the pages Shimura devotes to Serre in his memoirs are so gratuitously nasty that one wonders why Springer-Verlag agreed to publish them — arguments were advanced in support of each of the attributions. It is these arguments that deserve to be evaluated for their implicit or explicit assumptions about the virtues attributed to priority.

1. The case for Taniyama. By all accounts, Taniyama was the first to have expressed an expectation along the lines of the conjecture, in items 12 and 13 of his list of problems at the Tokyo-Nikko conference. It is also agreed that his version was not literally correct as stated.[7] I copy the English translation of Taniyama's questions from Lang's article in the *AMS Notices*:

$M\times\Phi$ vol. I.2 © 2023

[6] Wikipedia is an international battleground. The English, German, French, Russian, and Arabic pages are entitled "The Modularity Theorem" but disagree on its former terminology as conjecture: in English and German it was the "Taniyama-Shimura Conjecture", in Russian the "Taniyama-Shimura-Weil", and in French either "Taniyama-Weil", "Shimura-Taniyama-Weil", or "Shimura-Taniyama". The Arabic page mentions Taniyama-Shimura-Weil "and other names". The Spanish, Italian, Swedish, Hebrew, Catalan — and, as far as I can tell, the Chinese — pages are entitled "The Taniyama-Shimura Theorem"; the Portuguese page calls it the "Shimura-Taniyama-Weil Theorem"; and the Dutch page is "The Shimura-Taniyama Theorem". Wikipedia pages on mathematical topics are generally maintained by mathematicians, who, as far as I know, do not report to their national authorities. The proof by Breuil et al. is in the article "On the modularity of elliptic curves over Q: wild 3-adic exercises", *Journal of the American Mathematical Society*, 14 (4): 843–939 (2001).

[7] This opinion is shared by Lang, Serre, and Shimura. In a letter to Lang, Shimura wrote that Taniyama "was not completely careful, and if someone had

12. Let C be an elliptic curve defined over an algebraic number field k, and $L_C(s)$ denote the L-function of Cover k. Namely,

$$\zeta_C(s) = \frac{\zeta_k(s)\zeta_k(s-1)}{L_C(s)}$$

is the zeta function of C over k. If a conjecture of Hasse is true for $\zeta_C(s)$, then the Fourier series obtained from $L_C(s)$ by the inverse Mellin transformation must be an automorphic form of dimension -2, of some special type (cf. Hecke). If so, it is very plausible that this form is an elliptic differential of the field of that automorphic functions. The problem is to ask if it is possible to prove Hasse's conjecture for C, by going back this considerations, and by finding a suitable automorphic form from which $L_C(s)$ may be obtained.

13. Concerning the above problem, our new problem is to characterize the field of elliptic modular functions of "Stufe" N, and especially, to decompose the Jacobian variety J of this function field into simple factors, in the sense of isogeneity.

It is well known, that, in case $N = q$ is a prime number, satisfying $q = 3 \pmod 4$, J contains elliptic curves with complex multiplication. Is this true for general N?

2. The case for Shimura. (a) The case for Shimura's priority was made by Lang in his 1995 article: Lang established that Shimura had actually stated a version of the conjecture in the presence of Serre and Weil, at a specific place (the Institute for Advanced Study in Princeton) and during a specific time interval ("most likely in 1964", according to a private message[8] Shimura sent me in 1998). Weil mentions this conversation in his 1979 comments on his 1967 article, included in his collected works.[9] The chronology is important for Lang, since the conversation definitely took place before Weil published his result.

When I first read Weil's answer about "one and the other set being denumerable", I characterized it as "stupid". I have since also called it inane.

pointed out this, he would have agreed that the problem would have to be revised accordingly".

[8] Shimura, private emails to the author, May 6-12, 1998.

[9] Weil, *Oeuvres scientifiques*, Vol III, p. 450.

$M\times\Phi$ vol. I.2 © 2023

But actually, Weil's answer gives further evidence that he did not think of the conjecture himself. Indeed, as a result of his conversations with Serre and Weil, **Shimura was directly responsible for changing the prevailing psychology** *about elliptic curves over* $\mathbb{Q}$ [my emphasis; we will be returning to psychology].[(10)]

(b) Shimura added two arguments that need to be considered separately. In the first place, there is the important theorem that he first published in 1971 in his textbook *Introduction to the Arithmetic Theory of Automorphic Functions*: namely, that

Theorem (Shimura)**. To a modular form of weight 2 for** $\Gamma_0(N)$ **with rational Fourier coefficients one can associate an elliptic curve with rational coefficients.**

This is the evidence Weil cited in 1967 (as a *Mitteilung* from Shimura) and about which he asks "whether things always behave that way".

Twenty years ago, Shimura was attaching a great deal of importance to being credited with the proof of this theorem. He was unhappy with Wiles's discussion of this and a related point in one of his articles on Fermat's Last Theorem:

"Strangely Wiles says 'Any such elliptic curve has the property... has an analytic continuation ...' without saying that the result is due to me. This is another defect. Does he think that it is trivial?"[(11)]

(c) Shimura's third argument is of a different nature again. In his remarks on receiving the American Mathematical Society's Steele Prize for Lifetime Achievement, Shimura explained his motivation:

...at first I thought that ... curves obtained from a division quaternion algebra B over $\mathbb{Q}$ *might not be modular... but I realized that no nonmodular* $\mathbb{Q}$*-rational elliptic curves could be obtained for the following reason: Eichler had shown* [*that*] *the Euler products on B are already included in*

[(10)] Lang, p. 1304.

[(11)] Shimura, op. cit., May 11, 1998. In his letter to Serge Lang, February 14, 1999, appended to this article, he refers to "one more annoying item" in Wiles's article: "He attributes that statement to nobody. I wonder if he considers it trivial. You may call it a non-reference by Wiles".

The letter to Lang mainly complains that Wiles failed to appreciate the importance of the boldfaced Theorem, and only mentions the result about analytic continuation in the final paragraph. This observation could well serve as the starting point for an analysis of the virtues inherent in a new theorem; but I don't have enough information to reconstruct Shimura's thoughts on the matter.

$M\times\Phi$ vol. I.2 © 2023

those obtained from elliptic modular forms [2]. *The Tate conjecture on this was explicitly stated much later, but the idea was known to many people, and so* **it was natural for me to think** *that two elliptic curves with the same zeta function are isogenous. This fact concerning B, in addition to the results I had about the zeta functions of modular curves, may have been the strongest reason for my stating the conjecture that every* $\mathbb{Q}$*-rational elliptic curve is modular.*[(12)] [my emphasis, to be explained later]

3. The case for Weil. Of course Weil gets credit for writing and publishing the paper that includes his converse theorem, an insight at the basis of an entire research program. A vast generalization of the converse theorem by James Cogdell and Ilya Piatetski-Shapiro has been indispensable in 20 years of progress on the Langlands program.

Weil may have remembered the conjecture on elliptic curves from that 1964 conversation with Shimura. This doesn't strengthen the case for attaching his name to the conjecture. But Weil's paper included a major advance, already mentioned above, on the previous versions of the conjecture: he suggested that an elliptic curve E is a quotient of the Jacobian of the modular curve with level subgroup $\Gamma = \Gamma_0(N)$, where N is an explicit integer, the *conductor* of E, defined in terms of the arithmetic of E. This makes the conjecture considerably more precise: there is an equivalence between two sets of objects that matches a specific numerical property; moreover for a given N, the set of such objects is finite. Serre wrote that this was *a beautiful new idea ; it was not in Taniyama, nor in Shimura* (*as Shimura himself wrote to me after Weil's paper had appeared*)". But he then added an explicit value judgment:

Its importance comes from the fact that it made the conjecture **checkable** *numerically* (*while Taniyama's statement was not*). *I remember vividly when Weil explained it to me, in the summer of 1966, in some Quartier Latin coffee house. Now things really began to make sense. Why no elliptic curve with conductor* 1 (*i.e. good reduction everywhere*)*? Because the modular curve* $X_0(1)$ *of level* 1 *has genus* 0*, that's why! I went home and checked a few examples of curves with low conductor : I did not know any with conductor* < 11 *nor with conductor 16? No surprise, since* $X_0(N)$ *has genus* 0 *for such values of N, etc. Within a few hours, I was convinced that the conjecture was true.*[(13)]

[(12)] Shimura, *Notices of the AMS*, November 1996, p. 1345.

[(13)] Jean-Pierre Serre, "L'histoire de la 'modularity conjecture'", *Gazette des mathématiciens*, **91**, Janvier 2002, 55–57.

$M\times\Phi$ **vol. I.2** © 2023

The virtue Serre identified in Weil's formulation — that it lends itself to numerical verification — has been taken up by defenders of the claim for Weil's priority; most of the defenders I know happen to be French.

§ — "Conceptual evidence".

We should also hear what Andrew Wiles himself has to say about the origin of the modularity conjecture. Wiles may have been thinking of the virtue mentioned by Serre in the first page of his paper on Fermat's Last Theorem:

A well-known conjecture which grew out of the work of Shimura and Taniyama in the 1950's and 1960's asserts that every elliptic curve over $\mathbb{Q}$ *is modular. However, it only became widely known through its publication in a paper of Weil in 1967..., in which, moreover, Weil gave conceptual evidence for the conjecture.*

At the time, though, I understood that the "conceptual evidence" was the derivation of modularity from the characterization of modular forms given by Weil's converse theorem, together with the Hasse conjecture, in Weil's version (A3), and the Tate conjecture, mentioned below.

It's important to emphasize that all three protagonists — Taniyama, Shimura, and Weil — justified their speculation in roughly the same terms. Taniyama's 12th problem includes the following sentence:

If a conjecture of Hasse is true ... then the Fourier series obtained from $L_C(s)$ *by the inverse Mellin transformation must be an automorphic form of dimension* -2*, of some special type* (*cf. Hecke*).

Shimura never published any version of his conjecture, but Lang reported in 1995 that

The rationale for Shimura's conjecture was precisely the conjectured functional equation (*Hasse*)*, along the lines indicated in Taniyama's problem 12, suitably corrected. Shimura's bolder insight was that the ordinary modular functions ... suffice to uniformize elliptic curves defined over the rationals.*

Shimura confirmed this a few years later in a private communication.[14] And Hasse's conjecture constituted the "certain theoretical

[14] Shimura, unlike Wiles, did not see Weil's paper as a conceptual breakthrough. In his e-mail to me dated May 11, 1998, he wrote "'Conceptual evidence' is rather

$M\times\Phi$ vol. I.2 © 2023

grounds" [*gewisse theoretische Gründe*] to which Weil alluded (elliptically) in his paper. However, Hasse's conjecture alone would have indicated that the function $L_C(s)$ satisfied some functional equation, but not necessarily the condition (A2) that had been shown by Hecke to characterize modular forms. So Weil actually took his speculations to the next step: "*Weiter lässt sich vermuten*",[(15)] as indicated above, that $L_C(s)$ satisfies Weil's condition (A3), which Serre thought added the necessary precision to make the speculation of Taniyama and Shimura into a checkable and therefore full-fledged conjecture.

§ — Discussion: Naming Virtues.

1. Taniyama's clear and distinct idea. Taniyama's reasoning follows a pattern that is familiar to working mathematicians but that to my knowledge has not been taken up by philosophers. "If a conjecture of Hasse is true", he says, then something else must follow, because of Hecke's well-known characterization of modular forms. Taniyama's 12th problem can be taken to be a sketch of a proof of a theorem:

Taniyama's "theorem". *If Hasse's conjecture is true, then the L-function of any elliptic curve over a number field k is an automorphic form of weight* 2 *of some special type.*

In more detail: (i) Hasse's conjecture implies that $L_C(s)$ satisfies a functional equation, and (ii) Hecke's theorem then implies that $L_C(s)$ is the L-function of an automorphic form. Step (ii) is false for at least two reasons: (a) Hecke's theorem only applies when k is the field of rational numbers, and (b) it only applies when the conductor N is 1, which is never the case (a point Serre mentioned in the text cited above). So the virtue of Taniyama's contribution seems to

misleading. After all Taniyama and I were familiar with the results of Hecke on the characterization of an automorphic form by the functional equation of its Mellin transform, and our ideas were partly based on that correspondence. I always viewed Weil's result as its nice refinement. But "conceptual"? Really? I don't get it".

The word "conceptual" clearly continued to bother him, because in the letter to Serge Lang cited in footnote 10, he complained vigorously about Wiles's failure to understand his (Shimura's) contribution. Shimura sent me a copy of this letter and I have included a reproduction of the letter at the end of this article. The letter to Lang makes it clear, however, that Shimura was especially annoyed that Wiles and others failed to give him credit for the boldface Theorem cited above, under point (b) of the case for Shimura, and which he calls a "geometric construction".

[(15)] Hard to translate. On www.linguee.de one finds "it looks as if", "it can be assumed that", and "suggests that" among the translations, depending on the context.

$M\times\Phi$ vol. I.2 © 2023

involve the sketch of an incorrect proof of an imprecisely stated and undoubtedly false theorem (see (a)). The mistakes do not enhance the contribution, of course; the imprecise and partially incorrect reasoning served as the basis of Taniyama's imaginative leap, making him the first to intuit an unexpected connection between elliptic curves and modular forms.

The imaginative leap was not totally unprecedented. Dirichlet series with Euler products and functional equations arise in number theory in two ways. One has to do with Galois theory, the other involves constructions in complex analysis that can be interpreted as some sort of harmonic analysis on topological groups.[16] Class field theory, the major accomplishment of algebraic number theory in the first half of the 20th century, proves that both ways give rise to the same family of Dirichlet series when the structures involved (the Galois groups and the topological groups) are both abelian. Taniyama can be credited with the insight that something analogous is going on when the groups are not abelian. To the best of my knowledge, this sort of insight is considered contingent and thus not suitable as a topic for philosophical analysis. It is undoubtedly the most prized of all the mathematical virtues, however, and for that reason alone deserves the attention of anyone who is interested in mathematical practice. For example, the mathematical community would see no virtue in the publication of a new speculative intuition every week on the hope that one of them will turn out to be right. A new conjecture not accompanied by a compelling motivation — what we might call a *clear and distinct idea* — will win no credit for its author. In this respect the reasoning that accompanied Taniyama's imaginative leap qualifies as *cartesian* in Ian Hacking's sense: any reader familiar with both Hasse's conjecture and Hecke's theorem will immediately see the pertinence of Taniyama's 12th problem.

The apparent importance of the cartesian character of Taniyama's insight has an apparently paradoxical implication that deserves to be stressed: there is more virtue in Taniyama's incorrect formulation of his 12th problem, accompanied as it was by steps (i) and (ii) above, than there would have on the counterfactual

[16] This perspective was already visible in Erich Hecke's theory of L-functions of Grössencharaktere, and was made systematic in John Tate's influential thesis of 1950. Kenkichi Iwasawa had independently discovered an essentially identical method slightly earlier. Although Iwasawa never published his results and never wrote them up in detail in English, there has never been a priority dispute.

$M\times\Phi$ vol. I.2 © 2023

assumption that he had presented with no justification whatsoever the fully-formed and correct modularity conjecture, as later verified by Breuil-Conrad-Diamond-Taylor. Something about this paradox feels wrong, but I am not sure what it is.

Of course, mathematicians generally show more appreciation for a conjecture when it is accompanied by a proof, however uncartesian it might be, than when it is merely accompanied by a clear and distinct idea. Nevertheless, although Hacking thought cartesian proofs are rare, I would argue that cartesian insights are fundamental to mathematical communication and progress. Everyone remembers that Caesar crossed the Rubicon; only eyewitnesses and specialized historians remember the details of what happened when he got to the other side.

Before turning to the other two claimants to priority, I remind the reader that the 13th problem on Taniyama's list was a version of Conjecture II. I have alluded indirectly to the possession of the "clear and distinct idea" underlying this intuition on the part of each of the three protagonists. In Taniyama's case, it is implicit in the words "Concerning the above problem" that form the transition between his problems 12 and 13, together with his use of the word "isogeneity". This is a basic concept in the theory of *abelian varieties*, a class of algebraic varieties to which elliptic curves belong. An abelian variety is a projective algebraic variety which is also an algebraic group — its group structure is given by algebraic maps. To an algebraic curve is canonically assigned an abelian variety called its *Jacobian*; the Jacobian of the modular curve $X(\Gamma)$ is denoted $J(\Gamma)$. The theory of abelian varieties is one of the best understood parts of algebraic geometry. In particular, every abelian variety is a product of "simple factors" but only "up to isogeny". An isogeny between abelian varieties is an isomorphism up to finite error: it is a surjective homomorphism of algebraic groups with finite kernel. Taniyama's problem 13 says the Jacobian $J(\Gamma)$ for Γ of "stufe" N "contains" a certain elliptic curve; by "contains" he means that it is one of its simple factors up to isogeny.

We have seen that an L-function can be attached to any elliptic curve, but in fact Hasse's conjecture (or the Hasse-Weil conjecture) concerns the L-function attached to any algebraic variety. In particular, one can attach an L-function to an abelian variety. It has been known for a long time that two isogenous abelian varieties have the same L-function. Recall what Shimura wrote in 1996: *it was natural for me to think that two elliptic curves with the same zeta function*

$M\times\Phi$ vol. I.2 © 2023

are isogenous. For want of a better account of the *nature* to which Shimura is referring here, let's assume that this "idea", of which he claimed that it was "known to many people", was "clear and distinct" in their minds. It was this idea to which Weil alluded when we wrote that it was "of course obvious — *natürlich naheliegend zu erwarten*". Here is the full citation:

"It is of course obvious to expect that under these conditions C' [the elliptic curve factor of the Jacobian attached to the modular form by Shimura] is isogenous to C [the original elliptic curve that on "certain theoretical considerations" should be modular]; this is in fact confirmed in some cases". (*Es ist natürlich naheliegend zu erwarten, da unter diesen Umständen C' mit C isogen ist; das bestätigt sich tatsächlich in einigen Fällen.*)

With this in mind, problem 13 can be read in two ways. First, is it true that every $J(\Gamma)$ of "stufe" N has elliptic curves among their simple factors? This is the most obvious reading of the text, but it is a strange question, because there are many modular curves whose Jacobians do not involve any elliptic curves. The question was given a precise answer by Shimura's analysis of the Jacobians of modular curves in his 1971 textbook.

The alternative reading is that every elliptic curve with rational coefficients is isogenous to a factor of the Jacobian of some modular curve. This is now known to be true, as a consequence of two very substantial theorems: the theorem of Breuil-Conrad-Diamond-Taylor on the modularity of elliptic curves, and the 1983 theorem of Gerd Faltings that includes the *Tate conjecture* that Shimura had mentioned: that two abelian varieties over a number field whose L-functions coincide are necessarily isogenous.[(17)] This is the expectation that Shimura found "natural" and Weil

[(17)] At some point in the 1980s I had the disconcerting experience of being asked by Shimura whether or not it was true that Faltings had proved this theorem. Unfortunately, I don't remember whether it was before or after Faltings had joined Shimura in the Princeton mathematics department, nor whether it was before or after the work of Frey and Ribet, around 1985-86, that showed how Fermat's Last Theorem would follow from the Modularity Conjecture, in view of Faltings's theorem. Either way, Shimura was surrounded by people who were far better qualified than I to answer his question.

Tate's conjectures on algebraic cycles, including the one about isogenies of abelian varieties that Faltings proved, were made public in the 1960s. They are fundamental to our thinking about the Galois representations attached to the l-adic cohomology of algebraic varieties, but I am not familiar with their prehistory and cannot comment on why Taniyama, Shimura, and Weil found it "natural to think" that it was valid for elliptic curves.

$M\times\Phi$ **vol. I.2** © 2023

found "obvious", once Conjecture I has been established. While one needs a good deal of expertise and sophistication in order to perceive Conjecture II as an obvious or natural consequence of Conjecture I and Tate's Conjecture, the choice of words suggests that neither Weil nor Shimura is claiming priority for this particular aspect of the conjecture. So I will henceforth restrict my attention to Conjecture I.

2. Lang's realism, Shimura's phenomenology. Lang's case for Shimura's priority is primarily forensic in nature: his detective work places Weil and Serre in conversations with Shimura on this topic during a precise period in the early 1960s. Weil and Serre were skeptical when Shimura asserted that he "believed [that an elliptic] curve [over $\mathbb{Q}$] should always be a quotient of the Jacobian of a modular curve" (both Conjectures I and II, in other words). Since Weil was not only a skeptic at the time but even responded in a way Lang characterized as "stupid" and "inane", how could anyone assign him any credit for the conjecture?

If Lang's approach can be assigned to any philosophy, I would tend to call it *empiricist* and *realist*: to the extent that there is a fact of the matter (and Lang clearly believes there is), it can be established by gathering evidence. But Lang made a comment that calls this assignment into question: as we have seen, he claimed that "Shimura was directly responsible for changing the prevailing psychology about elliptic curves over $\mathbb{Q}$". We can agree that something changed about the way number theorists were thinking about elliptic curves over $\mathbb{Q}$ without commiting ourselves to Lang's use of the word "psychology", nor do we have to accept his judgment that Shimura had direct responsibility for the change by speaking with Weil and Serre. Let's say that we could ascribe to the community of number theorists the *disposition* to think about elliptic curves over $\mathbb{Q}$ in certain ways in 1960, and a *markedly different disposition* in 1970. Then we can say that Lang counts among the prized virtues of mathematical practice *the ability to alter dispositions* with regard to important topics, and that this is one of the main virtues he identifies in Shimura's role in proposing the modularity conjecture. This doesn't suffice as a characterization of Shimura's contribution, however. Other distinguished number theorists, including Barry Mazur, Yuri Manin, and especially John Tate, as well as Serre himself, were busily altering dispositions with regard to elliptic curves over $\mathbb{Q}$ in the decade following the publication of Weil's article, by

$M\times\Phi$ vol. I.2 © 2023

giving lectures and writing survey articles[18] promoting a very precise and detailed international research program that continues to frame thinking about elliptic curves even now. One presumes that they did so with dispositions altered as a direct result of Shimura's intervention with Serre and Weil a few years earlier — and, I should add, with Tate ("I told it to Tate around the same time, and I remember that he was impressed".[19]).

Lang is no longer among us, and we don't know what he had in mind when he refers to Shimura's "changing the prevailing psychology". My guess is that he would like his readers to believe that Weil's conversation with Shimura inspired him to think about the questions that led to his 1967 paper, inviting us to interpret his failure to mention either the contents of the conversation or Taniyama's 1955 questions as a deliberate attempt to claim the virtue of the original idea of Taniyama and Shimura for himself. That interpersonal conflict does not lend itself to philosophical analysis, whereas "dispositions to think about" various mathematical objects can be legitimately understood as targets of virtuous intervention within the community of mathematical practitioners. The dispositions, and the interventions to change them, can thus be incorporated in a philosophy of mathematical practice, whether or not Shimura's private conversations were as effective as Lang claimed. At the very least, this way of looking at things raises interesting questions about the relative virtue of private and public interventions within the community — and we should not underestimate the importance of the letters of recommendation with which this essay began in creating dispositions in the first place.

Shimura's own arguments (b) and (c) make no reference, explicit or otherwise, to the community of mathematical practitioners. To put them in perspective, it's helpful to rewind to the early 1950s, when Hasse's conjecture was known in very few cases but was the basis of a very active research program. Weil, Taniyama, and Shimura were reporting on their contributions to this program as early as the 1955 Tokyo-Nikko Conference, but their results (like those of Max Deuring, also present at the conference) were based on what was already known about abelian class field theory. The first non-abelian example had been treated in a

[18] Especially J. Tate, The arithmetic of elliptic curves, *Invent. Math.*, **23** (1974), 179–206, which is where I and everyone else I knew developed our dispositions at the time.

[19] Shimura, letter to the author, May 11, 1998.

© 2023 $M\times\Phi$ vol. I.2

1954 article by Martin Eichler. Shimura spent much of the next 20 years generalizing Eichler's observation and creating what is now called (following Pierre Deligne) the theory of *Shimura varieties*, of which the simplest (but by no means simple) case is the one treated in his 1971 textbook. In that book, and in a series of papers dating back to 1958, Shimura developed a systematic method for verifying new non-abelian cases of Hasse's conjecture for elliptic curves, in the process providing evidence for both Conjectures I and II. Lang puts it this way:

Shimura himself in the late fifties and sixties extended Eichler's results and proved that elliptic curves which are modular have zeta functions which have an analytic continuation. (*Cf. the three papers* [*Sh 58*], [*Sh 61*], *and* [*Sh 67*].)[(20)]

Thirty years earlier, this had already been mentioned in a survey article of J.W.S. Cassels that was largely responsible for creating the "prevailing psychology" about elliptic curves in the 1960s:

Hasse's conjecture has also been verified for a few C without complex multiplication which arise in the theory of modular functions [*Shimura* (*1958a, 1961a, 1964a*)][(21)]

What should we call the virtue, exemplified by this evidence, on which Shimura bases his claim (b) to priority? Wikipedia suggests one possible name:

Shimura's approach, later presented in his monograph, was largely **phenomenological**, *pursuing the widest generalizations of the reciprocity law formulation of complex multiplication theory.*[my emphasis][(22)]

Wikipedia's history page informs me that on March 31, 2007, Charles Matthews inserted the word that I have set in boldface, together with its philosophical overtones. The word is not used in Husserl's sense but rather as it is understood in philosophy of science. Shimura developed the theory in order to establish the facts of the matter.[(23)] In the course of this development, he made

[(20)] Lang, p. 1303.

[(21)] Cassels, J. W. S., Diophantine equations with special reference to elliptic curves. *J. London Math. Soc.*, **41** (1966) p. 280.

[(22)] Wikipedia, "Shimura variety", `https://en.wikipedia.org/wiki/Shimura_variety`.

[(23)] A Wikipedia editor named Arcfrk emphasized (in 2010) the contrast with Deligne's definition, which served "to isolate the abstract features that played a role in Shimura's theory".

$M\times\Phi$ vol. I.2 © 2023

the observation reported above in argument (c): that the phenomena (elliptic curves) observed in his study of the Shimura varieties attached to quaternion division algebras provided additional evidence for Conjectures I and II.

Thus Shimura claimed the virtue that Matthews calls *phenomenological*. As he himself acknowledged, it built upon Taniyama's original *cartesian* insight, but from the standpoint of philosophical analysis it is conveniently different. Nevertheless, I don't want to leave Shimura without pointing out that mathematical practice sees an unbridgeable gulf between *providing* (*phenomenological*) *evidence* for a conjecture and *proving* the conjecture. This is obvious, of course, and since no one is claiming that any of the triad (Taniyama, Shimura, Weil) came close to proving the conjecture, its relevance to identifying the virtues of a conjecture may not be clear. This may best be understood *a posteriori*, in the light of the Langlands program for automorphic forms, specifically the program Langlands outlined in 1979 for the study of Shimura varieties. During the 1960s, Alexander Grothendieck and his school had developed the theory of étale cohomology, which made it possible to attach Dirichlet series to automorphic forms in both of the ways mentioned previously. Étale cohomology provided the connection with Galois theory, while the relation to *representation theory* — what we previously called harmonic analysis on topological groups — gave a variety of constructions of Dirichlet series. The generalization of Conjecture I (no uniform generalization is possible for Conjecture II) asserts roughly that

A. *Galois-theoretic Dirichlet series are representation-theoretic Dirichlet series*

Specialists call this the "Galois-to-automorphic" direction of Langlands's conjectures; current work on this direction largely derives from the methods Wiles developed in his proof of Fermat's Last Theorem. The "automorphic-to-Galois" direction, naturally, asserts

B. *Representation-theoretic Dirichlet series are Galois-theoretic Dirichlet series*

Shimura's phenomenological efforts established **B** for the class of modular forms that is relevant to elliptic curves with rational coefficients, and provided the starting point for all subsequent work on **B**, which by now is established in considerable (though by no means complete) generality. But it's not immediately obvious why even a complete proof of **B** should help to establish a disposition to believe **A**, as it apparently did for Shimura.

$M\times\Phi$ vol. I.2 © 2023

Shimura actually did prove **A** — Conjecture I for the class of *elliptic curves with complex multiplication* — in a paper published in 1973.[24] Even though this class of elliptic curves has especially favorable properties (the problem in this case is very close to being abelian), phenomenological results of this kind typically do count as significant evidence for **A** within the framework of the Langlands program, since it is generally believed that we will have to wait several generations, if not centuries, for a complete proof of **A**.

3. Weil's falsificationism, on Serre's reading. Unlike some of his Bourbaki colleagues, Serre has never displayed an interest in philosophy for its own sake, but the virtue he discerns in Weil's contribution to the conjecture is the easiest to translate into familiar philosophical terms. If we replace the word "checkable" with its synonym "verifiable", then Weil's virtue, according to Serre, is that he has endowed the Modularity Conjecture with a *meaning*, following the *logical positivist* dictum that (in the words of Schlick) "The meaning of a proposition is the method of its verification".[25] In practice, though, *falsification* of the Modularity Conjecture by finding a counter-example would have carried more weight than its verification in special cases, if for no other reason than that a proof by (philosophical rather than mathematical) induction would require infinitely many verifications, which is impossible, whereas a single counter-example suffices for disproof. Thus the virtue Serre highlights may just as well be *popperian*. Either way, by insisting on the virtue of making a conjecture "checkable", Serre seems to be in agreement with Mach, when he asserted that "where neither confirmation nor refutation is possible, science is not concerned".[26]

The clarity of Serre's argument simplifies the philosopher's task; he has identified a virtue in conjecture-making that is both unambiguous and easy to map onto familiar concerns in the philosophy of science. I would go further and say that Serre has rendered a service to philosophy: while both the verificationism of the logical positivists and Popper's falsificationism are known to be vulnerable

[24] Goro Shimura, On the factors of the jacobian variety of a modular function field, *J. Math. Soc. Japan*, **25** (1973), 523–544. This theorem can be proved using Weil's converse theorem, but this was not the method Shimura chose, although his paper on modular forms of half-integral weight, published the same year, was based on a highly original application of Weil's 1967 paper.

[25] Moritz Schlick, "Meaning and verification", *The Philosophical Review*, *45*, No. 4 (Jul., 1936), 339-369.

[26] Mach, Ernst, 1883, *Die Mechanik in ihrer Entwicklung*, Leipzig: Brockhaus, 9th ed. transl. by T.J. McCormick, *The Science of Mechanics*, Chicago: Open Court, 1960, p. 587.

$M\times\Phi$ vol. I.2 © 2023

to critique, on the grounds that there is more to the world than its logical *Aufbau*, Serre's "checkability" may well be a viable criterion for the meaningfulness of a conjecture. At any rate, it has not been subject to philosophical analysis.

Most importantly, perhaps, Serre has shown some consistency in following his own dictum: Serre formulated his famous Conjecture on modular representations of Galois groups in 1973 but only published it in 1987, after, as he put it in his thanks,

Jean-François Mestre [...] *a réussi à programmer et vérifier un nombre d'exemples suffisant pour me convaincre que la conjecture méritait d'être prise au sérieux.*[27]

The examples occupy 11 of the 52 pages of his manuscript.

§ — Conclusion: the "prevailing psychology".

Lang indirectly hinted at a model of the *mathematical subject* by assigning to Shimura direct responsibility "for changing the prevailing psychology" by privately conveying his belief in the truth of what we are calling the Modularity Conjecture to two or three individuals. This subject is structured hierarchically, so that impressions acquired by those (like Weil, Serre, or Tate) at the pinnacle of the hierarchy have the power of detectably altering dispositions throughout the subject. Lang doesn't explain the means of this alteration, so we don't know the relative weight in the process of private exchanges, seminar presentations, or publications — to mention a few of the mechanisms with which most members of the community would be familiar.

But even granting that the dispositions of a collective subject can be altered as a result of a few brief conversations involving a group of exceptionally influential individuals, the question remains: why should there have been a "psychology about elliptic curves over $\mathbb{Q}$" in the first place? Whether or not there is a virtue ethics for mathematics, in which every conceivable topic merits its own psychology, in practice the attention of the mathematical subject is rather narrowly focused. Why were elliptic curves so much on the minds of number theorists in the 1960s?

[27] Jean-Pierre Serre, Sur les représentations modulaires de degré 2 de $Gal(\overline{\mathbb{Q}}/\mathbb{Q})$, *Duke Mathematical Journal*, **54** (1): (1987) p. 180. Its proof some 20 years later, by Chandrashekhar Khare and Jean-Pierre Wintenberger, made extensive use of the methods introduced by Wiles, Taylor, and others in proving the Modularity Conjecture, to which Serre's conjecture is very closely related.

$M\times\Phi$ vol. I.2 © 2023

It's easier to understand this attention when we remember that number theory at its core is concerned with understanding *Diophantine equations* — polynomial equations in several variables with integer coefficients– and more specifically with characterizing their set of solutions in integers, or in rational numbers. Equation $(\star)$ that we encountered at the beginning of the article, the one whose solution set is an elliptic curve, is a Diophantine equation. It is among the simplest possible classes of Diophantine equations, whose theory has been studied intensively since the 19th century (if not longer), and yet the nature of its solution set remains an outstanding open question. Much was known by the 1960s, however. It was proved by L. E. J. Mordell nearly 100 years ago that its set of rational solutions forms a finitely generated abelian group, a result generalized in André Weil's thesis and now known as the *Mordell-Weil Theorem*. This landmark in number theory, arguably the deepest result known at the time about Diophantine equations, posed a challenge to number theorists: how many generators does the group have? And how can we tell whether the group is infinite or finite? The challenge was more acute in view of the special status of elliptic curves among all algebraic curves[(28)] with rational coefficients. Every such curve has a genus, a non-negative integer; elliptic curves are the curves of genus 1. The rational solutions of curves of genus 0 — like the circle with equation $x^2 + y^2 = 1$ — are well understood (rational points on the circle are essentially the same as right triangles with integer sides). The *Mordell Conjecture* — proved by Faltings in 1983 — is the statement that curves of genus 2 have only finitely many solutions. For elliptic curves, simple as they are, there is no elementary way to determine whether the (rational) solution set is finite or infinite.

But in the early 1960s, when Shimura was privately conversing in Princeton with Weil, Serre, and Tate, over in England Bryan Birch and H. P. F. Swinnerton-Dyer were busy formulating and testing a *conjectural answer* to the question. The *Birch-Swinnerton-Dyer* (*BSD*) *Conjecture* — the appropriateness of its name has never been called into question! — asserts that the number of generators of the group of rational points of the elliptic curve C can be determined by examining its associated Dirichlet series — Taniyama's $L_C(s)$, Weil's Λ — but only if the Dirichlet series had (at least some of) the properties conjectured by Hasse.

(28) For our present purposes, an algebraic curve is defined by an equation in two variables, like $(\star)$.

$M\times\Phi$ **vol. I.2** © 2023

The (eminently checkable) BSD Conjecture, which was widely known[29] by the time Weil's article was published, gave such an appealing answer to the questions raised by the Mordell-Weil Theorem that one can easily understand why number theorists wanted the Conjecture to be meaningful — and thus for Hasse's conjectures to be true for elliptic curves, and thus for Taniyama's "theorem" to be true. With the help of the BSD Conjecture, the Modularity Conjecture thus becomes an object of number theorists' collective desire — a symptom of our mass psychology, as Lang intuited.[30]

Along with the Modularity Conjecture's implications for Fermat's Last Theorem, spectacularly realized by Wiles, this helps to explain why the controversy over the Conjecture's origin stirred up so much interest. On its own, however, it cannot settle the controversy. The virtues examined in this article — Taniyama's cartesian insight, Lang's realism, Shimura's phenomenology, and Weil's falsificationism — are by no means mutually exclusive. But nor does any of these virtues subsume any of the others. I would be delighted if my consideration of the virtues of the respective attributions helps to resolve the priority dispute in the minds of some of my readers. It has not settled the matter in my mind in one direction or another. For that reason I'm not sorry to see my colleagues have grown more relaxed about these attributions in recent years. I hope

© 2023 $M\times\Phi$ **vol. I.2**

[29] Thanks to the 1966 article by Cassels cited in note 20, and of course to lectures by Birch and Swinnerton-Dyer themselves, especially the latter's chapter in the book *Algebraic Number Theory* edited by Cassels and A. Fröhlich (London: Academic Press (1967)) which for 50 years has been required reading for every student of number theory.

[30] The importance of the BSD Conjecture in promoting the Modularity Conjecture is beyond the scope of this article, but two landmark articles written in the decades following the publication of Weil's paper give a sense of the excitement it generated. The introduction to "Arithmetic of Weil Curves", by Mazur and Swinnerton-Dyer, published in 1974 (but available several years earlier) explicitly identifies the "aim of Part I of this paper": to "examine... structures" that are "associated to the Weil parametrization and the bearing they have (via the conjectures of Birch and Swinnerton-Dyer) on the arithmetic" of elliptic curves over $\mathbb{Q}$. Part II as well as Part I became instantly applicable to all elliptic curves over $\mathbb{Q}$ with the proof of the Modularity Conjecture. In 1983 Benedict Gross and Don Zagier announced their proof of their famous formula the first that explicitly constructs infinitely many rational points on an elliptic curve using information about its Dirichlet series. Since the Modularity Conjecture was not yet established, their article (published in 1986, before the controversy over priority broke out) explains that the results apply to "strong Weil curves" and refers to the Mazur-Swinnerton-Dyer article for the terminology. B. Mazur and H. P. F. Swinnerton-Dyer, Arithmetic of Weil Curves, *Inventiones Math.*, **25** (1974) 1–61; B. Gross and D. Zagier, Heegner points and derivatives of L-series, *Inventiones Math.* **84** (1986) 225–320.

I have demonstrated that there are legitimate reasons to attach the names of one, or two, or three authors to the conjecture. As for myself, I prefer the version with no authors.

Princeton University **Department of Mathematics**
Fine Hall—Washington Road
Princeton, New Jersey 08544-1000

February 14, 1999

Professor Serge Lang
Department of Mathematics
Yale University
New Haven, Connecticut 06520-8283

Dear Serge:

This may not be your concern, but in my case the problem is the attitude of those people who make inaccurate or insufficient statements about the conjecture. What Weil did poses no longer any problem. Let me give examples.

In his paper *Modular elliptic curves and Fermat's Last Theorem* Andrew Wiles wrote "Weil gave conceptual evidence for the conjecture." I have been puzzled by this statement, and so recently I asked him about its meaning. His explanation was as follows: Given a **Q**-rational elliptic curve C, assume that the twists of the zeta function of C satisfy the functional equations of suitable types; then Weil's result gives a modular form f. Therefore C is modular, that is, my conjecture follows from the assumption concerning the functional equations combined with Weil's result.

I pointed out to him that the argument would produce a modular form, but it wouldn't say that C is a modular elliptic curve. He conceded that it didn't give any geometric construction of a modular elliptic curve, but added "Doesn't one call C modular if such an f exists?" He contradicts himself, because, the first line of his paper says "An elliptic curve over **Q** is said to be modular if it has a finite covering by a modular curve of the form $X_0(N)$." Therefore, if he really wishes to speak of *conceptual evidence* in the above sense, then he has to say something more. I would advise him to read Weil's 1967-paper again. Weil was more accurate than Wiles about this, since he at least mentioned my work which associated a modular elliptic curve E to such an f. Besides, even after finding E, we have to invoke the result of Faltings, which was of course established many years later. I was careful about this point in my 1971-Nagoya paper concerning elliptic curves with complex multiplication.

There is one more annoying item in the paper of Wiles: After the definition of modularity, he says "Any such elliptic curve has the property that its Hasse-Weil zeta function has an analytic continuation and satisfies a functional equation of the standard type." He attributes that statement to nobody. I wonder if he considers it trivial. You may call it a non-reference by Wiles.

These are examples of what I mean by inaccuracies and insufficiencies. Worse, there are also some people who accept or even copy that statement about *conceptual evidence*.

Sincerely yours,
Goro Shimura
Goro Shimura

1

$M\times\Phi$ vol. I.2 © 2023

Acknowledgement. — This article was originally written in response to an invitation by a group of philosophers as part of "a proposal for a special issue of the philosophy journal *Synthèse* on virtues and mathematics". The author thanks the editors of *Annals of Mathematics and Philosophy* for allowing me to rework the text for publication in the new journal, and especially to the referees for their suggestions for improving the exposition.

This work was partially supported by NSF Grant DMS-1701651. This work was also supported by the National Science Foundation under Grant No. DMS-1440140 while the author was in residence at the Mathematical Sciences Research Institute in Berkeley, California, during the Spring 2019 semester.

$M \times \Phi$ vol. I.2 © 2023

Michael Harris,
Department of Mathematics, Columbia University,
New York, NY 10027, USA



Annals of Mathematics and Philosophy
***vol.* I, *no.* 2**, 2023

$M \times \Phi$

Rota's variation on the eidetic identity

Carlos Lobo[1]

Gian-Carlo Rota's conception of identity is based on his own understanding of Husserl's eidetic variation. This is evidenced by numerous publications.[2] Most of them refer explicitly to phenomenology, but others, more discreet, belong to another side of his activities which is not philosophical, but mathematical. Some of his colleagues, for reasons we will discuss later, hardly recognize such philosophical implications *in* a mathematical practice, or prefer to consider them as a mere ornament, to be left aside. Yet, the philosophical problem of identity is at the heart of any scientific identity and more precisely of mathematical identity.

To understand how Rota works and searches for mathematical identities, and what he means by "identity as eidos", I will begin by referring to when Rota became himself, that is, a *self-reflective* mathematician. I do not mean to say that working mathematicians are not reflective,[3] especially in modern times, but Rota is certainly reflective in a special way: phenomenologically reflective. Nor do I mean to say that he is the first: Husserl, as a mathematician, was certainly of this kind and others after him, like Hermann Weyl or Brouwer.

© 2023

[1] This paper integrates part of a presentation given at the *Rota Memorial Conference*, 7/8 May 2019, organized by Albino Lanciani, Fabrizio Palombi, Marco Rigoli, held at the Università degli studi di Milano Dipartamento de Matematica Federigo Enriques, and completes analysis given in previous contributions (Lobo, 2017, and 2018).

[2] In his articles on Husserl in *Discrete Thoughts*, and in *Indiscrete Thoughts*, more particularly a chapter from an article written with Sokolowski and Sharp, entitled *Syntax, Semantics and the Problem of the Identity of Mathematical Items* (*Philosophy of Science*, 55, 1988, pp 376-386). A part is published in *Indiscrete Thoughts*, under a slightly different title *Syntax, Semantics and the Problem of the Identity of Mathematical Items* (Rota, 1997b, pp. 151-158). For this reason, we shall follow the first version (*i.e.* Rota, 1988).

[3] About reflectivity and the critical turn in mathematics, see H.-B. Sinaceur, Scientific philosophy and Philosophical Science, in *The Philosophers and Mathematics*, *Festschrift for Roshdi Rashed*, Ed. Hassan Tahiri, Springer, 2018, p. 56-63. And in the same volume, J.-J. Szczeciniarz, For a Continued Revival of Philosphy of Mathematics, especially p. 273-294.

But Rota is among the few who could so easily and explicitly articulate philosophically the series of motivated acts and decisions that constitute the ordinary work of mathematical thought, as well as the resources of mathematical virtuosity. Rota is undoubtedly such a virtuoso as well, but what sets him apart from many of his colleagues is that his scientific performances, at least after 1964, are accompanied by continuous and conscious critical reflection. This reflective turn in mathematics has been underlined, but this reflection is declared and explicitly phenomenological in Rota's case, that is referred to Husserl's transcendental and eidetic phenomenology, as it is evidenced by his philosophical lectures at the M. I. T.

I will begin with a brief presentation of his conversion to phenomenology. I will then present his philosophical account of identity as *eidos*, and of the phenomenological performance of eidetic variation, as well as his account of the ordinary treatment of identity in mathematics. Finally, I will offer a brief overview of some typical examples of such phenomenological awareness in his mathematical practice. I would like to talk about the notion of cryptomorphism and take as *eidê* some examples of mathematical identity.

§ 1. — The discovery of phenomenology by Rota.

The variations alluded to in the title do not refer to any change in Rota's views about identity as such, except perhaps the great change that was his conversion to phenomenology.

In a short article from 2000, *Ten Remarks on Husserl and Phenomenology*, he recounts how, after seven years of intense reading, Husserl's work suddenly began to become clear, to make sense, when for the "first time" he "managed to decipher [Husserl's] writing". This "happened one morning in March 1964". He was reading Husserl's *Ideas* in the back of the car, while his wife was driving.

> "As I was dejectedly rereading the conclusion of a long argument, *it suddenly made sense*. The *experience reminded me of one of those mixed multicolored engravings that hide an image, which you can only see if you stare at the engraving in a certain way*. My first reaction was 'So that's really it!', followed by 'Finally, I found the key'. I was young and brash. It took me a few more years to begin to understand. Every time I pick up one of Husserl's writings and

$M\times\Phi$ **vol. I.2** © 2023

> start reading, I *get the same fuzzy feeling, and I've learned to wait for the hidden image to emerge*."[4]

As a statistical confirmation, let me add a personal note. What struck me, in reading this account for the first time, is that, while working on a PhD on Husserl under the direction of Derrida, and *independently* of any knowledge of Rota's work, I experienced the *same* feeling. And even more: the *same* analogy came to me, that between eidetic intuition based on a reading of Husserl's writings and the vision of stereoscopic image.[5] This analogy fully illustrates and captures how the phenomenological description and the eidetic variation work, and, consequently, it helps us to understand the concept of identity, here at stake.

Like the figure resulting from a tangle of colored spots thanks to a specific ocular adaptation (a certain *parallelization,* let us note), the *eidos* emerges from the arbitrarily changing aspects of ordinary perceptions, in an act of eidetic vision. But in a more general way, we discover retrospectively that any act of identification implies an ideation. We realize, says Rota, that "strictly speaking, we can never see the chair itself" but only "*examples of it*". And to avoid usual misunderstandings, we must insist on the fact that this "exemplary structure" or exemplarity, in Rota or Husserl's sense, is *not* an exemplification, and, consequently, it does *not* entail that perception should be infused by linguistic or conceptual structures. This eidetic exemplarity is prior to any expression or discourse, and underpins any act of identification, be it vague or acute, linguistic or prelinguistic, etc. Therefore: to perceive "*this*" chair, and recognize it *as* a chair, as well as to recognize the face, presupposes an eidetic apperception, which remains implicit and potential as long as we are not turning toward the specific or eidetic identity as such. But this holds also for higher objectifying activities: to *understand* a concept, to *discover* an unsuspected isomorphism between two distinct domains of mathematics, or to see and grasp an intentional essence from the reading and understanding of a series of propositions, of arguments, of a conceptual distinction, etc. — all these multiple and varied activities of identification that we call *knowledge*, which all tend towards a form of general objectivity, can be described in the same terms. Subsequently, as Rota concludes, "the

[4] *Ten remarks on Husserl and phenomenology*. O.K. Wiegand et al (eds.), *Phenomenology on Kant, German Idealism, Hermeneutics and Logic*, 89-97 2000 Kluwer Academic Publishers, p. 89. (Hereafter: Rota, 2000).

[5] C. Lobo, *Le phénoménologue et ses exemples*, 2000, Paris, Kimé, p. 16.

$M\times\Phi$ vol. I.2 © 2023

commonsense distinction between concrete and abstract collapses and becomes doubtful".[6] Moreover, before philosophy posed explicitly the problem of this eidetic identification, a kind of spontaneous eidetic variation is permanently and tacitly implemented at different scales and in different fields, under different objectifying attitudes, in the field of technics as Socrates remarked as well as in that of material eidetic, such as geometry, and the various branches of mechanics (statics, hydrostatics, etc.).[7]

It is true that an important difference remains in the phenomenological reflection, between the *performing of* an experience of identification and its thematic and phenomenological *description*. For this reason, I have compared the eidetic vision realized by phenomenology to the vision of a stereoscopic image and not to the mere focusing on an image or a perceptible object. The stereoscopic image presupposes a type of vision that is *neither* spontaneous nor natural. It is a vision or a focusing at two levels: a *vision at infinity*, which requires a parallax of our binocular vision, and at the same time a vision through the surface of the tangle of spots which begin to cover each other along a horizontal axis. After some time, a depth is produced, and the "phantom thing" appears. Similarly, a clear and explicit eidetic perception is quite difficult, and it takes a long time before it emerges, i.e., the time for the *adaptation of the view*, especially when it takes the form of a second order perception, a reflection on former eidetic formations. But how long did it take between the historical appearance of geometry and the first reflection on its hypotheses, with Riemann among others, which revealed the hidden metric or the specific topology at the basis of all Euclid's theorems? Or how long did it take for physics or astronomy to realize that the description of motions was meaningless if one did not assign coordinate systems and if one did not manage to explain the rules of their transformations? In addition, we must insist with Rota and Husserl, that the insight at stake in the eidetic intuition or ideation did not require any extraordinary faculty or mental power, but only a special capacity and a methodical exercise to push intentionality and attention in a particular direction. One cannot therefore accuse us here of resorting to a mystical faculty,

[6] G.-C. Rota, *The End of Objectivity*, Lectures at the M. I. T., by G.-C. Rota, 1974-1991, Second Preliminary edition. Drafted in Collaboration with Sean Murphy and Jeff Thompson, p. 180. (Hereafter, *End of Objectivity*).

[7] Axiological and practical attitudes, which belongs in the sphere of non-objectifying intentionality, would require different analysis.

$M\times\Phi$ **vol. I.2** © 2023

as does the characteristic antiplatonism among philosophers. For I can imagine how Platonic this discourse on *intellectual vision* may seem and will provoke in many a form of skeptical reaction, if not outright hostility. Following Rota and Husserl, I will answer that Platonism as such is no objection. And I would go on to argue that since there are at least three of us who have experienced it, independently of each other, and reported on it in similar terms, that is an encouraging statistic for a start. Moreover, involuntarily, as many idioms in all the languages I know attest, we have a confused sense of such a vision. Rota expresses it as a spontaneous implementation of ideation.[8]

But again, these experiences as well as their spontaneous expression *are not* their methodological thematization; and even the manifestation of an act of ideation in a so-to-speak spontaneous phenomenological reflection is not methodologically sufficient, unless it is performed itself as an eidetic variation under a transcendental bracketing, that of transcendental *épochè.* Only then the identity becomes a *pole* of a multifarious thetic or positional activity, of perception, memory, imagination, symbolization... as well as for higher categorial objectifications as those at work in mathematics. This difficult but central point will help below to understand better the relation that Rota reveals on a formal perspective between syntax and semantics. From the point of view of phenomenology, which is that of the *a priori* of correlation[9] between the noesis and noema, — i.e., between the intending and the intended, the act and the content of act –, the theme is not only the *identity* on which one concentrates in each case, but the whole activity of thematization, the whole thetic or positional activity. We insist especially on the thetic or positional characters of acts and their correlative character of content (the posited, the theme or thesis, in the intentional and not only in the logical sense), because, as Husserl insists repeatedly, the spere of "positionality" constitutes the core of intentionality,[10] and explains subsequently in which deep and peculiar sense, that Husserl exposes in the lessons on *Passive Synthesis,* any cognitive activity is "logical", which does not mean discursive or linguistic.

© 2023 $M\times\Phi$ **vol. I.2**

[8] Rota et alii, 1992, p. 172-173.

[9] Husserl, 1950, see. §46, *Das universale Korrelationsapriori*, p. 161 passim.

[10] See for instance, Husserl, 1976, p. 333. For a brief overview of Husserl's reformation of the theory of "eidos" in its relation to the enlargement of the concept of "proposition", and "thesis", see m'y comment in « L'idée platonicienne d'eidos selon Husserl", op. cit. pp. 175-182, and more particularly p. 180.

In the epilogue to his conversion to phenomenology, Rota sent a message to Gödel praising Husserl as the greatest philosopher of all times. Gödel replied, adding that he was indeed the "true Kant", and the greatest philosopher since the time of Leibniz. And as Leibniz distinguished two great metaphysical labyrinths, that of *freedom* and that of the *continuum*, so Rota distinguishes two fundamental phenomenological problems: that of *time* and that of *identity*. Like the problem of the continuum, the problem of identity, although less vaunted and more technical, is surely the deeper one. "The problem of identity is equally important but less dramatic", says Rota, and at any rate much more difficult to "dramatize", although not deprive of some dramatic aspects.

§ 2. — Philosophical account of identity as eidos, and the phenomenological performance of eidetic variation.

The problem of identity, in its most schematic form, asks *how a* thing, let us call it x, can occur in different instances, be presented in different forms or modes. This implies conversely that something appears as being *an x* or *that x*, insofar as it is an *instance* of x, and therefore appears *as* something that constitutes its identity, and that goes beyond its factual being. This concentrates the essential features of the type of *vision* that we mentioned initially, by comparison with the perception of the stereoscopic image.

This form of identity is the *eidos*, the similitude that flows from one form to the other, while the transitory carrier can be called either *an* instance of the *eidos* or more generally its facticity. The things that appear in the naive attitude do not appear for themselves, but *for* (or as) something else. We only come to this understanding of the *what* because we seek to keep an eye always on the *how*. To observe *how* this reference is obtained requires us to depart, so to speak, from our ordinary attitude. This is the shift to the phenomenological transcendental attitude, through a universal bracketing. On this basis alone, can start the analysis of the structures of acts, of their composition and foundation (*Fundierung*), since higher order cognitive activities take the form of an articulated series of founded acts. Following Rota's terminology, the founding act gives the support or *facticity* to the higher and founded act, which is *in function* (*fungierend*, in Husserl's terminology); and

© 2023 $M\times\Phi$ vol. I.2

as I say the former functionality of the founding act is neutralized, disactivated, put outer function. Or in other terms, there is no irreducible or absolute facticity. Once clearly distinguished facticity and function, it is essential at the same time to keep in mind the relativity of *function* to *facticity*, and vice versa, in order to understand why and how the eidetic or specific identity is itself contextual.

> "To say that two marbles are identical is to say that the two marbles have identical *eidos*, not that they have identical facticities. The identity is the identity of the *eidos*. They can be identical because both can have the same *eidos*. In other words, there is confusion about identity only if we confuse facticity with function, as happens with physical reductionism. If we stay away from reductionism, there is no confusion about identity, and thus there is no confusion about how a decontextualized *eidos* can be contextually particular 'this'."(11)

Now, all identity is given through a multitude of modes of giving, and exploring these modes is crucial to avoid any form of reductionism. But because we are usually tended toward the object whatever it may be, in a process of identification, we focus mostly on the *what* and neglect the *how*. As a result, since the *what* is contextual, we have a confused idea of the *what* itself, and a rigid and decontextualized notion of identity. Reductionism is the typical reaction, the compulsion towards such a rigid identification, expressing anxiety about the fact that *all* identities are *context dependent*, functionally based on multiple changing facticities and therefore never completely fixed. Or, what amounts to the same thing, to be concerned with the how, is to *look at* the reference *in its context*, *as context dependent*, *in its "fringe" of given modes of being*. As Rota insists:

> "One of the characteristics of phenomenological description [...] is that, in the description of contextual phenomena, we give equal importance to the 'fringe'. In fact, one of the great differences between a purely objectivist description and a phenomenological description is that, in the latter, all kinds of marginal phenomena are given equal importance."(12)

(11) *End of Objectivity*, p. 184.
(12) *End of Objectivity*, p. 75.

$M\times\Phi$ vol. I.2 © 2023

Because he neglects the how, the objectivist — or reductionist — has neither an idea of the "things themselves", nor even a clear idea of the *what* itself[13]. Any kind of objectivist or positivist is reductionist by restricting his field of action to what he has first posited as the real object, forgetting the positional activity, and the way in which this position presupposes the prior position of the world. Against this restriction, the bracketing of the phenomenological description enlarges the field to the widest horizon, and at each stage of the investigation includes in its perspective the *how* of *what* is at stake. For this reason, Rota diagnoses in the very process of objectivation that leads to an objectivist attitude a resistance to this widening, an inner and unconscious protection against the phenomenological bracketing.

There is an affective and axiological side to this process too,[14] which Rota calls "reductionist anxiety". This anxiety is at the background of any epistemological attitude affirming that any "thing" that cannot be reduced to a region (psychic, physical) supposedly already and rigidly identified is considered as nothing. This anxiety reaches its paroxysm when one is confronted with identity *as such*, when one is about to recognize the identity of a thing, of everything that escapes us and that does not fit into the framework of a ready-made and fixed ontology. This is exactly what happens when we are first exposed to phenomenology, which focuses on the eidetic intentional structure of each region of being, and that, as we become familiar with this kind of approach, a feeling of instability sets in.[15] Reductionist anxiety stems from it.

As we have said, there is a huge difference between the implicit, tacit and spontaneous eidetic variation, with the specific parentheses that accompany it, and the explicit, and methodic variation. But does this mean that all knowledge is implicitly phenomenological knowledge? If the *what* of knowledge is the *eidos*. What about the *how*? Rota answers:

> "We know in many ways. At first sight, it seems that the scientist knows by examining tables of data and by observing scientific laws, that mathematicians know by giving proofs, and that the everyday man knows by simple calculation. [...] What is primordial in all these phenomena that can be singled out as the philosophical

[13] Against reductionism, identity, *End of Objectivity*, p. 34-35, p. 43
[14] On axiological predicates, see *End of Objectivity*, p. 5, p. 10-11.
[15] *End of Objectivity*, p. 60.

$M\times\Phi$ **vol. I.2** © 2023

> meaning of knowledge? Each of these ways of knowing is a *process of bracketing by eidetic variation*. In other words, phenomenologically, we assert that even through various instances, there are many ways of knowing, primordially there is only one, which is the bracketing of an *eidos* through eidetic variation."(16)

In the standard presentation of the phenomenological method, bracketing is the transcendental reduction and *epochè*, which transforms the world into a world-theme, where the naïve belief in its existence is suspended, and, as a consequence and correlatively, the *world-thesis* becomes a theme for a phenomenological description. This is another aspect of Rota's understanding and appropriation of Husserl's phenomenology with which I fully agree, especially concerning the inter-implication between eidetic variation and *epochè*. Rota's thesis is that "every phenomenal vision" implies an eidetic vision, and an eidetic variation.(17) And one of the philosophical goals of phenomenology is precisely to describe this *eidetic vision* in its essential characteristics. Rota enumerates three of them: 1) "we look to familiarize ourselves" with the issues, to disclose an intentional meaning. 2) "We see" insofar as we have a "project", i.e., an intention with its horizon-structure. 3) "We look" insofar as we look beyond what we are looking at. Such a "beyond"(18) motivates the deceptive analogy between sensual appearances and signs (that before Galileo was already suggested by Democritus).(19)

$M\times\Phi$ vol. I.2 © 2023

§ 3. — Rota's explanation of the ordinary treatment of identity in mathematics.

Among Rota's most surprising statements about mathematical identities are certainly the following two, which seem to contradict each other to some extent.

S1: Mathematical identities are ultimately (abstract) individuals.

S2: They are given, through a series of profiles and under many aspects, as any real and worldly individual.

(16) *End of Objectivity*, p. 182.

(17) *End of Objectivity*, p. 174.

(18) *Ibid*.

(19) Husserl's critique of this analogy and that of the distinction between primary and secondary qualities, in *Ideas* I, § 40, (see Husserl, 1976, 82 passim), has triggered many misunderstandings on the relation of phenomenology with modern sciences, and consequently on the diagnose of crisis of European sciences (Husserl, 1950) and the role that transcendental phenomenology should play in it.

S1. The mathematical objects (idealities) are (abstract) individuals. Contrary to the (concrete) individual objects, they are not identities; but the *eidos* is confused with it. The real line is not an instance of anything. "Most mathematical objects have [the] property that there is only one. For example, there is only one real line, even though it has many properties", and can be formalized in different axiomatic perspectives. "No mathematician says there are many [real lines]- it is the same line, loaded with the properties of the real line."[20] The same analysis is repeated and developed elsewhere:

> "The objects of mathematics, objects such as the real line, the triangle, sets, and natural numbers, share the property of retaining their identity while receiving *axiomatic presentations* that can vary radically. Mathematicians have axiomatized the real line as a one-dimensional continuum, as a complete Archimedean ordered field, as a real closed field, or as a system of binary decimals on which arithmetic operations are performed in a certain way. Each of these axiomatizations is *tacitly understood* by mathematicians as an axiomatization of the *same real line*. In other words, the mathematical object thus axiomatized is presumed to be the same in each case, and this identity is not questioned. In this paper, we wish to analyze the logical conditions that make it possible for *the same mathematical object to be designated by a variety of axiomatic presentations*." [21]

S2. However, mathematical objects are given through different (abstract) profiles, through different (abstract) perspectives. Like any scientific object: say the earth, the universe, but also matter, motion, life, DNA, etc., the real line, the number three etc. are the poles of endless processes of objectification (determination, identification, clarification, etc.). The similarity or identity in both cases is given by a series of profiles that provide the facticity for endless presentations of the *eidos*. These profiles are not constituent parts of the thing itself, nor are they even aspects or faces of it.

More, as any facticity can become the object at the core or the process of identification, it can also become the pole given through profiles. For instance, there is only one north face from a determinate

[20] *End of Objectivity*, p. 181
[21] Rota, 1988, p. 376.

$M\times\Phi$ vol. I.2 © 2023

perspective of the Duomo (perspective including determinate distance, height, focus etc.), and an infinity of profiles of it, which stems from the variations in the context (changing light, changing color, etc.). Of course, this "face" becomes itself, in this case, a context-dependent ideality. Similarly, the real line can be accounted for from many different angles. But since the real line is not a being existing in the world or in *a* world, what can be the profiles and what about its faces or aspects? Rota's answer is surprising but straight to the point. The profiles of the real line are *formal presentations*, and more precisely *syntactic or semantic presentations*. This implies that axiomatic systems are not a theoretical goal *in themselves,* but mere formal presentations, i.e., approaches, in an endless process, to mathematical eidetic identities.

Rota's understanding of Husserl's analyses of mathematical idealities is very acute and profound, and it departs from the current philosophical discussions on "mathematical existence", because, insofar as the individual identity merges with the *eidos*, its existence is equivalent to its possibility.[22] Rota thus takes the side of a "synthetic" (which does not mean necessarily constructive) philosophy of mathematics, a philosophy attentive to the real work of mathematicians:

> "By saying this, we recognize the real practice of mathematics, and we do not plead for Platonism: we are not in the least concerned by the *problem of the existence of mathematical objects*, any more than the grammarian is concerned by the problem of the existence of the verb or the adjective. Our argument aims at showing that the problem of the identity of mathematical objects is not solved, but rather exacerbated, by a doctrinaire appeal to axiomatics."[23]

In passing, this would shed new light on the problem of the *imaginary* in mathematics which was at the center of Husserl's philosophical reflections on the concept of number. An impossible concept within the frame of a given theory, such as $\sqrt{-1}$ can be syntactically required, for a coherent axiomatic presentation of an eidetic identity whatever it maybe[24].

[22] On "*Wesenhaftigkeit*" (essentiality) which is tantamount to "*Geltung*" (validity), see Husserl's *Prolegomena*. (Husserl, 1975, 241)

[23] Rota, 1988, p. 382.

[24] See Husserl's solution to the problem of the introduction of imaginary numbers in, *Philosophie der Arithmetik, mit Ergänzenden Texten (1890-1901)*, Hsg. Lothar

$M\times\Phi$ vol. I.2 © 2023

This entails a series of consequences**.**

1. The mathematical *ideation* is given in a perspective where an axiomatic system thus becomes a sketch or a formal profile of an object. This means that the eidetic vision is structured like the empirical vision. In both cases, the identification is a never-ending process, never completed. It never reaches a complete and ultimate certainty and evidence. Shocking as it may seem, any axiomatic presentation of a mathematical identity must be consistent but can never be complete: the mathematician wants to find out what else the real line may be[(25)] — *and there is always something else to find*. The "real line, or any other mathematical element, is never completely given in an axiomatic system", or in a finite series of axiomatic systems.

> "It is impossible to state all the axiom systems for a given theory. Think for example of the theory of groups. One has to introduce several axiomatizations of the group concept, which may even use ternary or more general operations. When a mathematician designs a new system of axioms for a known theory, for example group theory, he is already guided by an already intuited notion of group."[(26)]

2. Although mathematical approximation never gives rise to any *rectification* (refutation, "falsification", etc.) as in the empirical sciences or in ordinary knowledge, there is nevertheless an open horizon of possible determinations anticipatively determined, conjectured, or demonstrated. The limit concept of the set of all axiomatic presentations of the same object, let us say the "real line, can neither be explained nor foreseen in an exhaustive way". Consequently, the idea of a "totality of possible axiom systems for the real line cannot be exhaustively stated or anticipated". The ideal of completeness is thus defeated: "Every mathematical object allows for an open sequence of presentations by ever-new systems of axioms; that is, the successive development of ever-new systems of axioms for what is perceived as one and the same object."[(27)]

Eley, Husserliana 12, M. Nijhoff, 1970, pp. 441 and 444. Two conditions must be satisfied: 1. The imaginary number must be axiomatically consistently defined; 2. All propositions within this axiomatic system must be decidable. In order to avoid introducing an extrinsic completeness axiom, the axiomatic system must allow an enlargement, that is, the introduction of a new axiom, if it stumbles upon an undecidable proposition.

[(25)] Rota, 1988, p. 383.

[(26)] Rota, 1988, p. 381.

[(27)] Rota, 1988, p. 381.

$M\times\Phi$ vol. I.2 © 2023

We find a similar analysis in *The End of Objectivity*, with a critique of a narrow conception of formal logic, the "myth of a standard logic" which would answer the requirement of the absolutist thinker, as if objects, including so-called abstract objects, say mathematical objects, did not depend on context. But formalization remains necessary and plays itself a heuristic role, by making evident the gap between the actual practice of mathematics and its formal account. Axiomatic systems reveal aspects of the mathematical object $(= X)$ that were not manifest *as such*. This is the case of the formalization of the notion of group to classify an enormous amount of pre-existing special groups, which led to an *a posteriori* discovery: "the most significant examples of groups were discovered after the abstract concept of group had been isolated by an axiomatization. Even if mathematicians consider explicitly classifying all models of groups, for example all finite groups, the practice of mathematics nevertheless belies such explicit programs. Examples of mathematical structures for which all models are found are rare (the most successful in recent memory being Elie Cartan's classification of simple Lie groups)."[28]

3. As a corollary, formal logic must itself be extended to take account of the operations of *searching* for an axiomatic system themselves and of the deliberation leading to prefer one axiom system over the other. The general motive being that a certain axiomatic system gives a *better presentation* of a new property of the mathematical object under consideration: "each of these systems is supposed to reveal new characteristics of the mathematical object."[29] Thus every formal approach is itself dependent on a *seeing*, on an informal anticipation of an eidetic content, i.e., a "*pre-axiomatic grasp*" of the mathematical identity that functions as a thread. Once a formal presentation of an aspect of the object is grasped and trivialized, once we have familiarized ourselves with the object through a certain perspective, the intuition can be put aside. The same thing happens with the grammatical rules of a language, once we understand those rules and become a fluent user of that language, the rule is set aside, or in the background of our attention and concern. Yet, without these rules, our speech would lose its meaning, just as without this "guiding intuition, even if not verbalized"[30], any formal system would lose its meaning and could not be motivated and designed. This happens

[28] Rota, 1988, p. 383.
[29] Rota, 1988, pp. 383-384.
[30] Rota, 1988, p. 384.

$M\times\Phi$ vol. I.2 © 2023

not only with seemingly geometric objects, but also with structural and highly algebraized objects, such as a group.

> "We will call 'pre-axiomatic grasp' the understanding of the notion of group that has been freed from the necessity of choosing a system of axioms. Since there is no single group, but groups are of incredible variety, such a pre-axiomatic understanding of the notion of group cannot be attributed to familiarity with a single group. One could argue that such an understanding is derived from familiarity with more than one group and their 'common' properties, but this would be begging the question, since *familiarity with more than one group presupposes an unstated understanding of the concept of group that allows one to recognize multiple instances as instances of the same general mathematical structure*."(31)

As underdetermined or blurred this "eidetic identity" may appear, it is presupposed by each instance that fall under consideration.

4. Since "a complete description of the logical role of a mathematical object is beyond the reach of the axiomatic method as it is understood today", the phenomenological description calls for a new logic, extended to cognitive operations, such as logical individuation, modalization, foundation, or relations such as context dependence, etc.(32) Identification as the presentation of an *eidos* through formal profiles is one of these operations. This position is common to Rota, Husserl and Stanley Ulam, who was convinced that "logic formalizes very little of the processes by which we think" and that "the time has come to *enrich formal logic with other fundamental notions* [emphasis added]. What do you see when you see? You see an object *as* a key, you see a man in a car *as* a passenger, you see some sheets of paper *as* a book. *It is the word 'as' that has to be mathematically formalized, just as the connectors 'and', 'or', 'implies' and 'not' have already been accepted in formal logic*. Until you do that, you won't get very far with [the] A.I. problem."(33)

(31) Rota, 1988, pp. 381-382. Emphasis mine.

(32) On *equiprimordial* modes, see *End of Objectivity*, p. 13, p. 15; specifically on modalities, excluded by standard formal logic, *Ibid.*, p. 19; on context dependence, see 8ff. and on *Fundierung*, *Ibid*, see p. 41.

(33) Barriers of Meaning, in *Indiscrete Thoughts*, (Rota, 1997b, pp. 58-59).

$M\times\Phi$ **vol. I.2** © 2023

We find the same demand for a formal account of a larger number of operations that come into play in any cognitive process, including affective stances, which would in any case require and presuppose a careful phenomenological investigation. But the question remains: "Can a rigorous formal analysis supplant what was a psychological description of intelligently performed operations?"[(34)] For example, are operations such as *comparison*, *theory formation*, *example citation*, and *reference* etc. formally and eventually algorithmically representable?

5. This implies a new relation of the logical problem between syntax and semantics, between an axiomatic system and its interpretations. Models are classes of mathematical objects satisfying certain syntactic rules, where each model includes an additional structure. The "paradigmatic example" is given by set algebra[(35)], which has provided the ideal example of a successful formalization of a mathematical theory, (1) where a syntax is specified with intuitive inference rules, together with a complete description of all models for such a syntax, in this case sets, (2) "and, what is more, a table of semantic concepts and syntactic concepts, together with a clear translation algorithm between the two. Such an algorithm may, for the purposes of this discussion, be called a 'completeness theorem'."[(36)]

6. The conception of formalization itself must be modified accordingly, to *preserve* the structural incompleteness of the mathematical horizon. Informally, this amounts to relating any identity to its context, i.e., to the "bundle of intended models" surrounded by "an unclassifiable variety of unexpected, unwanted, 'non-standard' models" and yet sought after. Semi-formally, Rota expresses this situation as follows:

> "If an axiomatic system A is used to describe a model X, and another axiomatic system B is also used to describe the same model, the contribution of system B is desirable because it shows *what* X can be *other* than what it is shown to be by A. The contribution of B is not desirable simply because it confirms what we know about X by A,

[(34)] Rota, 1988, p. 385.

[(35)] "Multisets can be added and multiplied; however, a characterization by algebraic operations of the family of multi sets of a set S — an analog of what Boolean algebra is for sets — is not known at present. The problem is of more than academic interest. There is a deep duality between the algebra of sets and the algebra of multisets that a syntactical description may well elucidate. [19, Chap. XII. Syntax, semantics, and the problem of identity of mathematical items". Rota, 1988, p. 154-155]

[(36)] Rota, 1988, p. 379.

$M\times\Phi$ vol. I.2 © 2023

> but because it reveals something else that is true about X. It allows X to be presented more fully and to be re-identified in another context."(37)

7. This analysis and similar better explain Rota's and Husserl's common project of a reform of logic and, in Rota's case, the reformed model theory. This reform entails a deconstruction of the absolutist and objectivist norm of formalization and of the usual logical understanding of the relation between syntax and semantics. Of course, logical algebra from Boole to Halmos and beyond provides the standard for all semantics and syntax, but, as Rota notes, the discovery of a formal hidden analogy (a "cryptomorphism") usually comes afterwards, only once the mathematician becomes familiar with two models:

> "Although rigorous criteria have been stated for the *equivalence of two axiomatic descriptions of the same* object (e.g., in first-order logic), these criteria are often expressed as if the mathematician were suddenly faced with the problem of verifying that two previously given axiom systems are in fact *cryptomorphic*. This verification is sometimes requested, but rarely. In *reality, the mathematician is normally confronted with two axiom systems when he is already fully aware of the identity of their models*." (38)

In order to understand this ternary relation between syntax, semantics and identity as eidos, let us come to the analysis of some examples.

§ 4. — Overview of some typical examples of such practice in Rota's mathematical work.

1. *The Reynolds operator*

While working on the Reynolds operator, Rota realized that the averaging operator used in many fields was formally analogous to the Navier-Stokes equations, when properly simplified. This opened a new field of investigation and repeated attempts to connect parts of logical quantization and probability measures on sigma-algebras, where the Reynolds operator appears as a mathematical identity.

(37) Rota, 1988, p. 383.
(38) *Ibid.*

$M\times\Phi$ vol. I.2 © 2023

Pushing forward Halmos's program of algebraic logic, Rota mentions, as possible application and as "the most promising outcome", the *translation* of "the notion of quantifier on a Boolean algebra" into that of "linear averaging operator": "in this way, problems in first order logic can be translated into problems about commuting sets of averaging operators on commutative rings".[(39)] The conclusion of the paper we are following here relates this translation of predicate logic to that of probability, which gives the starting point of a contribution with D. P. Ellerman (*A Measure Theoretic Approach to Logical Quantification*) (Rota, 1978), including the cryptomorphic bridge between probability measure and quantification. In its presentation from 1973, the analogy between probability and predicate logic is mediated through the following steps: assignment of a probability measure m to the canonical idempotents; definition of a lattice space norm on the valuation ring by using a linear functional, complementation of the resulting normed linear space which can be seen as the space of all integrable functions, representation of the averaging operator as a conditional expectation operator, which, once relaxed from the restrictive condition that "every element of the range be finite-valued", can apply universally.[(40)] Rota Remarks: "An averaging operator on a valuation ring $V(L)$ is a linear operator A such that: (1) $Au = u$, $Az = z$. (2) $A(fAg) = AfAg$. (3) If [the function] f is in the monotonic cone, so is Af. Sometimes these operators go by the name of Reynolds operators. In probability, they are called *conditional expectations*." We have here, as it has been noticed by Jean Dhombres, a perfect example of Rota's synoptic ability and his quasi-polymath profile.

> "Rota was a master at perceiving connections among disparate subjects and in this instance, he took a combinatorial point of view. In particular, he used canonical idempotents instead of characteristic functions, borrowing the idea from Louis Solomon's work on Burnside algebras [...]. But something else was occurring Rota *was no longer analyzing the structure of averaging operators, but, constructing averaging operators to fit the application. For the study of logic, he constructed specific 'averaging' operators on valuation rings, which simulated the properties of existential quantifier*. In this way, 'Problems of first order logic, such

(39) Rota, 1973a, p. 576.
(40) Rota, 1973a, p. 622.

$M\times\Phi$ vol. I.2 © 2023

> as the decision problem, can be shown to be equivalent to algebraic problems for valuation rings with averaging operators.' *Unfortunately, it proved quite difficult to carry through this program rigorously, as one can see in the paper* [...]*, written with David Ellerman*"[41].

Following Birkhoff example, who always kept an interest in logic, Rota inverts what would be the current perspective for mathematicians. And the eventual disciple and commentator's surprise and deception echoes the reaction of Rota's teachers in mathematics, who did not understand his interest in formal logic, and could even less guess his philosophical motivations. As to the latter, they did not stem from the deceiving perspective of some philosophers misled by the "fake philosophical terminology of mathematical logic" "into believing" that it "deals with the truth in the philosophical sense". According to Rota's terminology, this variety of believers constitutes precisely what he names philosophers under "a bad influence of mathematics", which relies on a *misunderstanding* concerning mathematics as well as philosophy and falls into that typical artificial philosophizing whose symptoms, in certain schools of philosophy, are: "casual and self-satisfied symbol-dropping of mathematical logic" and contempt for history and evidence (or intuition), especially in the Husserlian sense of the term.[42] Abandoning the belief that mathematical logic had something to do with truth and scientific truth was the due price that logic had to pay in order to become "a successful and respected branch of mathematics".[43] By giving up any philosophical pretention means that logic can be treated as any other branch of mathematics, and that we should not inject into it any philosophical meaning. Yet that does not mean we dismissed any research concerning a philosophical logic. On the contrary, we should keep logical questions opened. More: since mathematical logic has either nothing to do with the way we think, (as some logicians think) or, in the better case, as already stated, "formalizes only very few of the processes by which we think", *philosophic logical* investigations should remain at the same time an open field as well as an open source for eventual enrichment of formal logic.[44] Whatever the success or failure of Rota's attempt to reconstruct algebraically

[41] Dhombres, 2003, p. 163.
[42] Rota, 1991, 134 & 135, respectively.
[43] Rota, 1997b, 97. Rota, 1991a, 133-136.
[44] Rota, 1997b, p. 92-93.

$M\times\Phi$ vol. I.2 © 2023

the quantifiers, and in order to decide if he succeed or failed, it is worthwhile seizing precisely what are his logical aims. While doing so, we gain a better understanding of the strange transference of an operator from a mathematical field (constructed in the context of fluid mechanics) to logic (predicate calculus), via algebra. I mean Reynold's operator.

Although he recognizes a wider scope to Rota's exploitation of "the idea of averaging operators"[45], Dhombres considers that the main mathematical goal that Rota had in mind in using the Reynolds equation (a generalized form of the Navier-Stoke equations)[46] was a broadening of the field of applications (models) and a unification of the theory of measurement.

The Reynolds operator is an averaging operator, i.e., a heuristic way to simplify the Navier-Stokes equations which are fundamental for hydrodynamics (description of gas and fluid flows, etc.) and beyond. Their form is quite complex, combining three quantities of material elements (mass, momentum, energy) whose density is measured; the velocity field and the energy density per unit mass. From the beginning, Rota is interested in representations (i.e., models) of the Reynolds operator, in ergodic theory, in spectral theory and in probability theory. The simplified form is written $\mathrm{R}(fg) = RfRg + \mathrm{R}\{(f - Rf)(g - Rg)\}$. (NB. The mean form is: $A[f(Ag)] = (Af)(Ag)$, with f, g two functions belonging to a commutative algebra). The transition from the Navier-Stoke operator to the Reynold's operator is smoothed, so to speak, by considering the functions of the Navier-Stokes equations as *random functions.* The Reynolds operator is thus seen as a *conditional expectation operator* of the form: $E^{\Sigma'}(uE^{\Sigma'}v) = E^{\Sigma'}uE^{\Sigma'}v$. [47]

Dhombres comments:

> "In Rota's mind, the mathematical task was to understand the proper role of the identity [equation], but *not in an axiomatic or formal way. He thought that studying a variety of situations in which Reynolds operators played a*

[45] "In 1973, Rota exploited the idea of operator averaging in the radically different context of valuation rings of a distributive lattice in [50] [i.e., here (Rota, 1973a)]. This paper is fascinating because, as Rota writes, 'the method of presentation is deliberately informal and discursive. Much of the reasoning is by analogy, but not in the structuralist or Bourbakist sense, and many of the results are merely heuristic'. The objective is grandiose, nothing less than the 'linearization of logic'." (Dhombres, 2003, 163).

[46] *For* more details, see: Rota, 2003: p 137-139.

[47] Rota, 1960, 138.

$M\times\Phi$ vol. I.2 © 2023

> *role in the real* (*or phenomenological*) *world would provide a better understanding of the meaning of the Reynolds identity.* Although Rota believed that a mathematician should be inspired by the physical world to do pure mathematics, he disagreed with the view of those mathematicians who insisted on the primacy of problems arising from technology that had to be solved, whether or not they led to pure mathematical results."[48]

It is true that Rota did not *only* understand applications in an empirical sense but *also* had in view formal models (and as we have seen, vice versa). Significantly (recalling Rota's words as supervisor of his PhD thesis, in 1970), Dhombres reports "how he insisted on concrete realizations of such operators on different function algebras *and* how he focused on how the Reynolds identity and related identities could simplify known results"[49] ; and even how the Reynolds operator gradually disappeared from his writings, after 1974, since it might seem too complicated or too rich. But, at the same time, even before that, from 1964 on, Rota no longer claimed to use it as a means of producing a unified theory of measurement and had to take it as a simple game that could generate "calculations that could be used or illustrated by examples in functional analysis". The main reason given by Dhombres was that there were alternative and equally effective "tools" for the same purpose, "such as derivations and Baxter operators"[50]. From such a change, we should assume either an excessively adaptive attitude or an inconsistency in Rota's position, moving from one project to another.

His position in and vis-à-vis the history of mathematics should be, in both senses of the term, discrete (discontinuous and anecdotal).

> "For Rota, *progress in mathematics could sometimes be made by recognizing a pattern or by working by analogy*. Reynolds' operators [sic], which had been a *kind of exercise* for the mind, remained *for him a source of inspiration*. This suggests that other authors, even if they do not work specifically on Reynolds operators, might also be *stimulated by Rota's methods*. A crude way to measure this is a citation search,

[48] Dhombres, 2003, 157.
[49] Dhombres, 2003, 157.
[50] See Rota (1969, 1972, 1995, 1998). Cartier (1972).

$M\times\Phi$ vol. I.2 © 2023

> but mathematical papers tend to cite direct technical links rather than indirect links of inspiration."[51]

Finally, as a side effect, Dhombres suggests that "the need to represent Reynolds operators in different concrete situations may have fostered the development of Rota's phenomenological viewpoint for mathematics. He first explained his view in 1973b in an article ["Edmund Husserl and the reform of logic", *Explorations in phenomenology*, p. 299-305. Published also in *Indiscrete Thoughts*, Rota, 1973b]. See [also Rota, 1997b]".

Contrary to what Dhombres suggests in his yet very enlightening article[52], Rota does not look to phenomenology for "simple" explanations of the possible applications in different contexts of the Reynolds operator. Rather, this minimalist interpretation explains Dhombres' severe judgment and underestimation of the importance of Rota's interests in logic[53]. (But this is an old story, as we learn from Rota's *memories* of Fine Hall, his years at Princeton). In a deep and explicit affinity with Husserl's logical investigations, Rota was looking for new resources, which could help a methodical understanding of an implicit but necessary moment in the history and practice of science (and mathematics in particular), and which could operate — as seems necessary — or induce a deeper transformation of formal logic.

We will recall the general formal framework within which Rota builds his positive and ideal historical perspectives. This framework consists of — or at least can be expressed as — a dynamic interpretation of the logical relationship between syntax and semantics. Without such a framework, Rota's specific critical

© 2023 $M\times\Phi$ **vol. I.2**

[51] Dhombres, 2003, 157 (emphasis added).

[52] "In 1973, Rota *exploited the idea of operator averaging* in the radically different context of evaluation rings of a distributive lattice in [50] [here, under (Rota, 1973a)]. This paper is fascinating because, as Rota writes, «the method of presentation is deliberately informal and discursive. Much of the reasoning is by analogy, but not in the structuralist or Bourbakian sense, and many of the *results are merely heuristic*. The *goal is grandiose, nothing less than the 'linearization of logic'*" (Dhombres, 2003, 163).

[53] "I was *too young and too shy to have a personal opinion about Church and mathematical logic. I was in love with the subject, and his course was my first graduate course*. I could feel the disapproval around me; only Roger Lyndon (the inventor of spectral sequences), who had been my freshman advisor, encouraged me. Soon after, he himself was encouraged to move to Michigan. Fortunately, I had met one of Church's most flamboyant former students, John Kemeny, who, having just finished his tenure as a professor of mathematics, was being introduced — by Lefschetz's gentle hand — to the philosophy department." (Rota, 1997b, p. 7).

stance and his statements on the past and present history of science would be difficult to understand. Following Rota, let us call it *ideal history* (*history as it should or could have been if...*). This counterfactual history is combined with a vision of history *as it really happened*, to obtain a sharper vision of the latter as well as a richer vision of what should and could happen in the future. Ideal history is the exact opposite of ideological history. From this point of view, Rota's repeated, sharp and witty criticisms of the prejudices of different kinds of academics (mathematicians, philosophers, logicians, A.I. researchers, etc.) express neither an indictment of intellectual sportsmanship nor a form of socio-psychological complex. If one had to attribute to it any psychological disposition, one would have to label it with its beautiful Platonic name: the concern for the *communication of forms*.

From this point of view, Rota's generalization of Reynold's identity is a perfect example of his strategy of radical, systematic and rigorous epistemological *decompartmentalization* (between probability theory, set theory, first order logic, etc. and philosophy and mathematics).

2. *Construction of the norm*

In a collaboration with David Sharp and Robert Sokolowski, *Syntax, Semantics, and the Problem of the Identity of Mathematical Objects* (Rota, 1988), adopting a framework similar to that of Husserl's eidetic, Rota extends the scope of cryptomorphisms and creates a tension between the two formal categories: that of semantics and syntax.

Rota's article that we will follow here traces the various stages of the algebraization of logic. Starting from Boolean algebra (the algebraic version of propositional calculus), it reviews the progress of the algebraization of logical semantics: from the algebra of sets, which constitutes the "paradigmatic example" for mathematical logic, to Birkhoff's lattice theory (1967), which represents, at least for a time, the model (*ideal example*) of a "*successful formalization of a mathematical theory*" as well as of a successful mathematical theory. Indeed, let us repeat with Rota the list of criteria that are satisfied: (1) "the *syntax of it is specified by intuitive rules of inference*"; (2) "it further includes a *complete description of all the models of such a syntax*, in this case sets", (3) which is more "*a table of semantic concepts and syntactic concepts*", and "a clear *algorithm of translation between the two*".

$M\times\Phi$ vol. I.2 © 2023

After the success of the series of attempts to algebraize logic, what was only a non-formalized program with a hegemonic aim has imposed itself as the norm of all semantics and all syntax. This normative claim presupposes, as the previous characterization suggests, (1) a criterion *of syntactic equivalence* of different axiomatics of the same theory *T,* which Rota calls "cryptomorphism"; (2) a *classification of the models* of this theory *T* independent of the choice of specific axiomatics; (3) a *completeness theorem* establishing a relation between "truth in all models" (semantic truth) and "truth as a demonstration from the axioms" ("syntactic truth"). If this program had succeeded, we could have immediately concluded that it is impossible to pose a mathematical object in excess of its axiomatizations and, consequently, that the project of a reform of logic[(54)] has failed, is useless or impossible.... The principle (3), called "completeness", is not what gives the familiarity with a model either, but rather it presupposes it. *Cryptomorphisms* are most often discovered after the fact, i.e., once *the mathematician has acquired a sufficient familiarity of models and a full understanding of the "eidetic identity" of models*. We come here to Rota's illuminating conclusion already quoted above (Rota, 1988, p. 383).

By "identity" or "object", Rota does not mean a particular equation or structure, but rather a *pole-idea*. Taking a stance which is opposed (and complementary) to the current understanding of model theory, as well as category theory, Rota asserts that a *single* mathematical "identity" ("object")[(55)] is itself the "*identity*" of an infinite number of objectivations (and equations), bearing various syntactic *and* semantic presentations. From a historical point of view, this pre-seizure is exposed to various series of modalizations (doubts, partial seizures, incomplete or wrong seizures, etc.). In this article, Rota focuses on the pole idea of mathematical constructs, which are neither *the*, nor *a domain of* reality, according the "mathematical realism" i.e., to the so-called "mathematical Platonism". Nevertheless, it is an essential moment (component) of reality and of every object. Avoiding crystallizing reality by stuffing it with mathematical constructs ("substructuring" it in Husserl's terms), the eidetic intentional structure of scientific experience, described by Husserl, is adapted by Rota to the sphere of

(54) For a more extensive comment on this aspect of Rota's thought, see. C. Lobo, 2017.

(55) With good reason, following in the footsteps of Husserl's critique of objectivism, Rota would later refer to them as «items» instead of "objects" or "identities", claiming the "end of objectivity", i.e. "the end of the objectivist conception of experience" (Rota, 1974-1991, 138).

© 2023 $M\times\Phi$ vol. I.2

mathematical experience, and he sets out an *a priori* framework for the history of mathematics, which does not exclude its openness or annihilate the freedom of mathematical thought. But in this particular case, as in the general case, to put it in Husserlian terms: However great this freedom of union and categorical formation, it always has its limits governed by law.[56]

This scope is fully and explicitly assumed, since the starting point of the article is to recognize "that this duality of presentation, syntactic and semantic, is shared by all mathematical theories and is not only found in mathematical logic."[57] From the way in which the interaction between identity, syntax and semantics is regulated, we can deduce that there is neither a single syntax, nor an ultimate syntax, and that the dual relationship explored in symbolic logic, through its transfer to mathematics, is transformed into a ternary structure.

> "The analysis of the similarity of mathematical objects, which we'll call mathematical constructs, immediately points to a closely related problem, which was first recognized in the philosophy of symbolic logic, but which has a much wider scope in the philosophy of mathematics, as we explain below. A presentation of a mathematical system leading to the definition of an object is necessarily syntactic, i.e., it is given by axioms and rules of inference. Nevertheless, axioms and rules of inference are intended to characterize a class of mathematical objects made up of sets endowed with an additional structure (such as groups, manifolds, etc.). Any set endowed with such additional structure that satisfies the axioms is considered a *model* for the axioms, and the description of all such models is usually regarded as the semantic interpretation of the theory." (*Ibid*.)

The problem is thus rephrased: "how can disparate syntaxes nevertheless have the same semantics, i.e., the same model"?

The third term, "identity", is merely an idea *at a distance.* For this reason, it is precisely the one that is most immediately given, albeit confusedly. The introduction of this third term implies a softening and opening of the whole logical frame and forces the "fundamental conclusion": "that a complete description of the logical role of

[56] Sixth Logical Investigation, § 62. Husserl, 1984, p. 717; see as well: Husserl, 1970, p. 309.

[57] Rota, 1988, p. 377.

$M\times\Phi$ vol. I.2 © 2023

the mathematical object is beyond the reach of the axiomatic method *as it is understood today.*"[58] Among the examples of "identities" or "similarities": the *real line* is probably the most paradigmatic of a mathematical "similarity" supporting an infinite variety of formal syntaxes and semantics, which cannot be anticipated, but must each time be constructed. For this reason, and for the sake of clarity, before turning to the case of the probability measure, let's move on to the presentation of the eidetic and phenomenological structure of the mathematical process of its determination, as sketched at the end of the article. The words "object" and "mathematical object" are to be understood here as synonyms for "similarity", "sameness", "identity" or "eidos".

> "If one accepts our discussion leading to the conclusion that two (or more) axiom systems for the real line (say) can be recognized to present the *same* real line, then one is forced to draw the following consequences:
>
> **(1)** The real line, or any other mathematical object, is not entirely given by a specific axiom system, or by any specification of a finite set of axiom systems.
>
> **(2)** The totality of possible axiom systems for the real line cannot be exhaustively explained or predicted. Every mathematical object allows for an open sequence of presentations by ever-new axiom systems, i.e., the successive development of ever-new axiom systems for what is perceived as one and the same object. Each of these systems is supposed to reveal new characteristics of the mathematical object.
>
> **(3)** Studying the real line is not a game to be played with axioms and develop the ability to draw consequences from them. On the contrary, the very choice of which properties to deduce and how the theory is to be organized is guided from the outset by a *pre-axiomatic* understanding of the real line. Without a guiding intuition, even if unverbalized, such an axiomatic theory cannot make *sense*.
>
> **(4)** Even if the concept can initially be learned by working assiduously on the axiomatic approach, this axiomatic approach will be abandoned once familiarity with it is achieved. Thus, by learning through a particular axiomatic system, one discovers a concept whose full understanding lies literally *beyond the* scope of that system." (Rota, 1988, 384-385)

[58] Rota, 1988, p. 384. Emphasis added.

$M\times\Phi$ vol. I.2 © 2023

The highlights of the structure are as follows:

1) *Identity* (*similarity*) pre-inscribed in a variety of syntaxes, although prescribed at a distance, this identity is neither a trivial identity nor an equation.

2) An infinite plurality of syntactic presentations, uncountable in advance, each new syntactic presentation revealing, in a very slow and progressive process, new formal characteristics of *the same* object, which is not a fixed and empty closed form, but rather an inexhaustible potential for ever-new forms.

3) These new forms, these new determinations motivate and lead to new constructions, and are partially conceived, i.e., partially, informally, intuitively, tacitly etc. anticipated; this is where the "tracks" come in, giving solidity and meaning to mathematical activity.

4) Although understanding the identical object goes beyond mastering the axiomatic game, a syntactic presentation provides an unavoidable framework for anyone wishing to learn something *concrete about the* mathematical object; in other words, axiomatics is a necessary but not sufficient condition.

The question arises: *what are the operations that make it possible to realize such a structure and to place axiomatics in the phenomenological role of* (*formal*) *presentations* (Darstellungen) *of an idea-at-a-distance, of an idea as a pole* in infinitum (for instance that of the "real line")? By way of comparison, let us recall what Husserl says about the objects of geometry: they are at the same time *material eidê* and *ideas in the Kantian sense,* that pole-ideas. From the outset, the real line is neither the Euclidean line (although historically presented by it), nor, of course, the line perceived by the senses (which is precisely "idealized" by the former). Both are, in various respects, merely what Husserl calls *Anhiebe-Ideen*, that is: "initial ideas" or better, *eidetic sketches* of the pole-idea. The central eidetic operation performed is the construction of a particular type of analogy, through morphic (or "anamorphic") projections in both directions, without giving preference to one form over another. These structural analogies are precisely what is named by Rota "cryptomorphism" and, in my opinion, should be compared to Husserl's use of structural analogies (among others, in his research on affective and volitional intentionality, and correlatively, on axiology and ethics in the broad sense of the term). The promotion and cultivation of this mode of circulation are underpinned by a set of presuppositions, or rather, they involve a (historically and rationally motivated) relaxation in relation to a set of presuppositions concerning the development of rationality.

$M\times\Phi$ vol. I.2 © 2023

Among the decisions underlying this mode of circulation are the followings.

1) The integration and correction of model theory: "Every field of mathematics has its zenith and its nadir. The zenith of logic is model theory (we dare not say what we think its nadir is). The sure sign that it is a zenith is that when we ignorant and stupid non-logicians try to read the material, we get the feeling that it should be rewritten for the benefit of a general audience.[59]" Consequently, the severe reform of the Tarskian notion of truth (motivated by Paul Cohen's forcing): "A new paradise was opened when Paul Cohen invented forcing, soon followed by the reform of the Tarskian notion of truth, which is the idea of Boolean-valued models. On certain topics, like this one, one senses that an unfathomable depth of applications is at hand, which will lead to a recasting of mathematics."[60]

2) The admission of category theory to its rightful place: "We were turned away from category theory by the excesses of the sixties when a noisy crowd claimed to be rewriting mathematics in the language of categories. Their pretensions were toned down, and *category theory* took its modest place alongside *network theory*, more pretentious than the latter, but enjoying solid support from both Western and Eastern masters. — One wonders why category theory has aroused such fanatical opposition. One reason may be that understanding *category theory requires an awareness of the analogies between disparate mathematical disciplines*, and mathematicians are not interested in stepping outside their narrow territory"[61].

3) The rehabilitation of the updated projects of algebraic logic and universal algebra: "Ever since theoretical computer scientists began to supplant traditional logicians; we've witnessed the resurgence of non-standard logics. These new logics return problems to universal algebra, with salutary effects. Anyone who believes that the theory of commutative rings is the central chapter of algebra will have to change their tune. The combination of logic and universal algebra will take over." (*Ibid*.)

4) The reasonable and fair evaluation of lattice theory despite the bad press of quantum logic with the hope of drying up the source of a perfect example of *hopeless philosophy*: the philosophy of quantum mechanics. Rota is rather severe on that point:

[59] Rota, 1997b, p. 218.
[60] Rota, 1997b, p. 218.
[61] *Ibid*. (Emphasis mine).

$M\times\Phi$ vol. I.2 © 2023

> "It has always been difficult to take quantum logic seriously. One malicious algebraist scornfully dubbed it 'the poor man's von Neumann algebras'. The background of lattice theory made people suspicious, given the bad press lattice theory has always had. [...] It's hard to conceive of a more accomplished example of the *philosophy of despair* than the philosophy of quantum mechanics. It was born of a marriage of misunderstandings: the myth that logic is linked to Boolean algebra and the claim that a generalization of Boolean algebra is the notion of modular lattice. Thousands of articles confirmed mathematicians' worst suspicions of philosophers. This philosophy came to an end when someone proved conclusively that these observables, which are the analogues of random variables in quantum mechanics, cannot be described by the structure of lattice theory alone, unlike random variables. — This debacle had the salutary effect of opening the field to an honest philosophy of quantum mechanics, at the same level of honesty as the philosophy of statistics (of which we'd like to see more) or the philosophy of relativity (of which we'd like to see less)."(62)

In order to demonstrate the fruitfulness of this softened syntax/semantics relationship, Rota proceeds with a kind of eidetic variation: he lists examples of various syntactic presentations of the same semantics; and vice versa; but also, cases of syntactically accomplished mathematical theories but without semantics and conversely examples of semantics devoid of syntax. Interestingly, as a paradigmatic example of a fully realized formalization of a mathematical theory, with "*an array of semantic concepts and syntactic concepts, as well as a translation algorithm between the two*", Rota once again cites the distributive lattice theory of Birkhoff (1967) and Stone (1937). He observes that examples of semantics still lacking a clear syntax abound in mathematics, because setting up a syntax requires reflection and because, unlike logicians, mathematicians are somewhat wary of logical reflection. What's more, one need to familiarize oneself with the mathematical structures involved, through exercises, to obtain a basis for such reflection, and an initial axiomatization(63).

(62) Rota, 1997b, p. 218-219.

(63) Rota/Baclawski give the following advice to their students: "The aim of this course is to learn to think probabilistically. Unfortunately, the *only way to learn to*

$M\times\Phi$ **vol. I.2** © 2023

Unsurprisingly, Rota gives three examples of mathematical theories that, directly or indirectly, refer to probability theory and show the need to reform its current axiomatization. Let us review them briefly.

3. *The testing of the norm on the ground of some mathematical examples.*

To test this ternary relationship between *syntax*, *semantics* and *eidetic identity*, Rota proposes two variations: **A)** examples of mathematical objects with syntax but without a clear semantic structure; **B)** examples of mathematical objects with semantics, but whose syntax remains obscure. Let us recall the major lines of these analyses that we have already developed elsewhere.

A) Among the examples of mathematics with syntax but without clear semantic structure. Rota mentions as a first example: The Hilbert operator ε. This is undoubtedly a *semantic operator*, but one that cannot be formalized. This notion is that of ε: the individual variable x such that $F(x)$ is true, if it exists", where $F(\varepsilon(F(x)))$ replaces the usual existential proposition $(\exists x)F(x)$[64]. On this point, there were attempts shortly afterwards, in 1993 and following years. The second example is that of the semantics of intuitionistic logic and modal logic, which until the work of Kripke were not understood. The discovery of their semantics led to a breakthrough; it was realized that intuitionistic logic was a *syntactic presentation* of the [semantic] idea of *forcing* due to Paul Cohen.[65]

B) Examples of mathematics with semantics, but whose syntax remains obscure. In general, this is the most common case in mathematics. Because syntax implies a logical reflection, and this reflection the mathematician does not like to do. Not to mention that one must first become familiar with the structure for such reflection to take place and have a chance of succeeding. The existence of a mathematical logic which develops a syntax before its semantics does not contradict this order, because the logic has been based on the familiar experience of the language, and the repeated reflections to grasp its syntactic structure.

Example 1 (transitional, as this is already a case of semantics without syntax): Closed subspaces of a Hilbert space that are the *analoga* of events in probability theory, or propositions in the classical

think probabilistically is to learn probability theorems. It is only later, after the theorems have been mastered, that the probabilistic point of view begins to emerge as the specific theorems fade from memory: rather like the Cheshire Cat's smile." (Rota/Baclawski, 1979: vii).

[64] Rota's notation instead of: $F(\varepsilon_x Fx)$.

[65] cf. Fitting, M. C. 1969, *Intuitionistic Logic. Model Theory and Forcing*, North-Holland, Amsterdam.

$M\times\Phi$ vol. I.2 © 2023

sense. These Hilbert spaces are, in standard approaches to quantum physics, like the *elementary semantics* of quantum mechanics. But it is a semantics without syntax. Rota recalls here the failures of "quantum logic": "*the numerous attempts to develop a 'logic of quantum mechanics' have failed because nobody has been able so far to develop a syntactically efficient presentation of the mathematical structure consisting of the events of quantum mechanics*"[(66)] — the attempts of Birkhoff and Von Neumann in 1936 (modular lattices) or that of Loomis (in 1955). Ortho-modular lattices have not been admitted because of the discrepancy between their semantics and the one suggested by the current practice of quantum mechanics.

Example 2: The theory of multisets, to which Rota will have recourse in his introductory lessons on probability in 1976.[(67)] A multiset of a set is a generalization of a subset of *S*, whose elements can occur several times (thus greater than 1). Hence the definition: *a multiset is a function* (*called "multiplicity"*) *of the set S in the non-negative integers.* We can add and multiply multisets. We have an analog of a Boolean algebra for sets. But the duality between the algebra of sets and the algebra of multisets remains to be elucidated syntactically.

Example 3: The theory of probability whose presentations are syntactically informal. For example, when a probability theorist or statistician speaks of a *sequence of random variables forming a Markov chain*, he reasons directly about them without worrying about the complex construction of a phase space or state space.

> "The statistician who computes with confidence intervals and significance levels seldom appeals to the measure-theoretic justification of his reasoning. It can be surmised that a syntactical presentation of probability will view joint probability distributions as playing a role similar to truth-values in predicate calculus. Kolmogorov's consistency theorem shows how to construct actual random variables in a sample space whose joint distributions are formalized to be a given family of consistent distributions. Thus, from this point of view, Kolmogorov's theorem *should turn* out to be the completeness theorem relating the syntactic and semantic presentations of probability theory".[(68)]

[(66)] *Op. cit.* p. 380.

[(67)] Rota (Gian-Carlo), Baclawski (Kenneth), 1979, An introduction to probability and Random Processes.

[(68)] Art. cit. p. 381. This example has been presented more thoroughly in Rota, 2001. See my comments in Lobo, Some Reasons to reopen the Question of

$M\times\Phi$ vol. I.2 © 2023

§ — Conclusion.

Rota's practice of a history of mathematics, sensitive to the eidetic structure of mathematical experience has manifest and declared affinities with Husserl's historical considerations developed in the *Krisis*. Both are opposed to ideological exploitations from artificially reduced or blind perspectives, whose various individual and sociological symptoms are for Rota as for Husserl: *reductionism* and *objectivism*.

Unless we consider the ternary relation between syntax, semantics and identity, and identity as a moment of an eidetic structure, with its opened horizon structure, mathematical experience and historicity remain unintelligible. But taking them into account requires adequate concepts and analytical resources. In particular, the codetermining modes belonging to the domain of modalizations are essential.

Of course, Rota himself insists on the "empiricism" of such an attitude and refers it back to Lakatos. But this "empiricism" in brackets is, like the epistemic "relativistic" sensibility, the expression of an epistemic sense of *contextual dependence* of all identity, of all *idealities*. Far from falling into some form of relativism, historicism, sociologism, this sensitivity of knowledge to context is a new and deeper vision of the *eidos*, and *eidos* that is at the heart of "the inner historicity of mathematics as a science."[(69)]

$M\times\Phi$ *vol. I.2* © 2023

§ — References.

Benis Sinaceur, Hourya. (2018). Scientific philosophy and Philosophical Science, in *The Philosophers and Mathematics*, *Festschrift for Roshdi Rashed,* Ed. Hassan Tahiri, Springer, 2018, 25-66.

Cartier, Pierre. (1972). On the structure of free Baxter algebras, in *Advances in Mathematics,* 9 (2): 253–265.

Dhombres, Jean. (2003). Reynolds and Averaging Operators, in Dhombres J., Kung J. S.P., Starr N., *Gian-Carlo Rota on Analysis and Probability, Selected Papers and Commentaries,* Birkhaüser. Boston, Basel, Berlin: 154-166.

Fitting, Melvin. (1969), *Intuitionistic Logic. Model Theory and Forcing*, North-Holland, Amsterdam.

Foundations of Probability Theory Following the Rota Way, *The Philosophers and Mathematicians, In honour of Prof. Roshdi Rashed*, Ed. Hassan Tahiri, Springer, 2018.

[(69)] Art. cit. p. 384.

Husserl, Edmund. (1950). *Die Krisis der europäischen Wissenschaften un, die transzendentale Phänomenologie, Eine Einleitung in die phanomenologische Philosophie*. Hsg. Walter Biemel. *Husserliana*, Band VI. The Hague: Martinus Nijhoff.

—. (1969). *Formal and Transcendental logic*, English tr. Dorion Cairn, Den Haag: Martinus Nijhoff.

—. (1970). *The Crisis of European Sciences and Transcendental Phenomenology, An Introduction to Phenomenology*, tr. David Carr. Evanston: Northwestern University Press.

—. (1975). *Logische Untersuchungen*, *Erster Band, Prolegomena zur reinen Logik*, Text der 1 und der 2 Auflage, Hsg. Elmar Hollenstein. Husserliana Band XVII, Martinus Nijhoff Pub., Kluwer, The Hague, Boston, Lancaster.

—. (1976). *Ideen zur einer reine Phänomenologie und phänomenologische Philosophie*. Hsg. K. Schuhmann. Husserliana, Band III/1. Den Haag: Martinus Nijhoff.

—. (1984). *Logische Untersuchungen*, *Zweiter Band, Zweiter Teil, Untersuchungen zur Phänomenologie und Theorie der Erkenntnis*, Text der 1 und der 2 Auflage, Hsg. Ursula Panzer, Husserliana Band XIX/2, The Hague, Boston, Lancaster. Martinus Nijhoff Pub./Kluwer.

—. (2001). *Logical Investigations*, edited by Dermot Moran and translated by J. N. Findlay from the second German edition, London, New York, Routledge.

Lobo, C. (2000). *Le phénoménologue et ses exemples*, 2000, Paris, Kimé.

—. (2012). « L'idée platonicienne d'eidos selon Husserl », *L'idée platonicienne dans la philosophie contemporaine*, Jalons. Dir. A. Mazzu, S. Delcomminette, Paris, Vrin.

—. Some Reasons to reopen the Question of Foundations of Probability Theory Following the Rota Way, *The Philosophers and Mathematicians, In honour of Prof. Roshdi Rashed*, Ed. Hassan Tahiri, Springer, 2018: 145-188.

—. Le projet husserlien de réforme de la logique et ses prolongements chez Gian-Carlo Rota, *Revue de synthèse* 138 7^{e} série N° 1-4 (2017) : 105-150.

$M\times\Phi$ *vol. I.2* © 2023

Rota, Gian-Carlo. (1960), On the passage from the Navier-Stokes Equations to the Reynolds Equations, in Dhombres J., Kung J. S.P., Starr N., *Gian-Carlo Rota on Analysis and Probability, Selected Papers and Commentaries,* Birkhaüser. Boston, Basel, Berlin, 2003.

—. (1969). Baxter algebras and combinatorial identities. I, II, *Bull. Amer. Math. Soc*. 75, 325

Rota, Gian-Carlo and Smith, D. (1972) Fluctuation theory and Baxter algebras, *Istituto Nazionale di Alta Matematica* IX, 179 (1972).

Rota, Gian-Carlo. (1973a). *The valuation ring of a distributive lattice*. Proceedings of the University of Houston, Lattice Theory Conference (Houston, Tex., 1973) Dept. Math., Univ. Houston, Houston, Tex.: 574-628.

—. (1973b). Edmund Husserl and the reform of logic in D. In D. Carr & E. S. Cas (Eds.), *Explorations in phenomenology* (pp. 299–305). The Hague: Nijhoff.

—. (1978). In N. Metropolis, Gian-Carlo Rota, Volker Strehl and Neil White, *Partitions into chains of a class of partially ordered sets*, Proceedings of the American Mathematical Society, Volume 71, Number 2, September (1978): 193-196.

—. (1978). Ellerman, David P. and Rota, Gian-Carlo (1978). *A measure-theoretic approach to logical quantification*. Rend. Sem. Mat. Univ. Padova 59 (1978): 227-246.

Rota, Gian-Carlo and Baclawski, Kenneth (1979). *An introduction to probability and Random Processes*. (Original lecture, accessible on David Ellerman's website)

Rota Gian-Carlo, David Sharp and Robert Sokolowski. (1988). *Syntax, Semantics, and the Problem of the Identity of Mathematical Objects*, in *Philosophy of Science,* 55 (1988), pp. 376-386.

Rota, Gian-Carlo. (1989). *The barrier of meaning*. *In memoriam: Stanislaw Ulam*. Notices Amer. Math. Soc. 36 (1989), no. 2: 141-143.

—. (1991a). *The pernicious influence of mathematics upon philosophy. New directions in the philosophy of mathematics* (New Orleans, LA, 1990). Synthese 88 (1991), no. 2: 165-178.

© 2023 $M\times\Phi$ vol. I.2

—. (1991b). Mathematics and the Task of Phenomenology, in T. Seebohm, D. Follesdal, J.N. Mohanty (eds.), *Phenomenology and The Formal Sciences*, Dordrecht: Kluwer.

—. (1974-1991). *The End of Objectivity, Lectures at the M. I. T.,* by Gian-Carlo Rota, 1974-1991

Rota, Gian-Carlo, Kac, Mark, and Schwartz, Jacob T. (1986). *Discrete thoughts. Essays on mathematics, science, and philosophy*. Scientists of Our Time. Birkhaüser, Boston, Mass., 1986. xii+264 pp.

—. (1992). *Discrete thoughts. Essays on mathematics, science, and philosophy*. Revised and corrected reprint of the first edition. With the assistance of Peter Renz. Boston, Inc., Boston, MA: Birkhaüser

Rota, Gian-Carlo. (1995). Baxter operators, an introduction, In J. P. S. Kung Ed.: *Gian-Carlo Rota on Combinatorics, Introductory papers and commentaries*, Contemp. Mathematicians, Boston: Birkhaüser.

—. (1997a). *Memory of Garrett Birkhoff: The many lives of Lattice Theory* Published in the Notices of the AMS, Volume 44, Number 11, (1997): 1440-1445.

—. (1997b). *Indiscrete thoughts*. Boston, Basel, Berlin: Birkhaüser.

—. (1998). *Ten mathematical problems I will never solve*, Invited address at the joint meeting of the American Mathematical Society and the Mexican Mathematical Society, Oaxaca, Mexico, Dec. 6, 1997. DMV Mitteilungen Heft 2, 45.

—. (2000). Ten Remarks on Husserl and Phenomenology, in O. Wiegand, R. J. Dostal, L. Embree, J. Kockelmans, J.N. Mohanty, *Phenomenology on Kant, German Idealism, Hermeneutics and Logic*, Dordrecht: Kluwer

—. (2001). Twelve problems in probability no one likes to bring up, Fubini Lectures, in *Algebraic Combinatorics and Computer Science*, Springer, 2001, pp. 25-96.

Carlos Lobo, ⋆
Centre Gilles-Gaston Granger, Aix-Marseille ⋆ ⋆

$M\times\Phi$ **vol. I.2** © 2023

Annals of Mathematics and Philosophy
***vol.* 1, *no.* 2**, 2023

$M \times \Phi$

Homage to Francis William Lawvere (1937–2023)

Gonzalo E. Reyes

© 2023

$M \times \Phi$ *vol. 1.2*

When Jean-Pierre Marquis, after the death of Francis William Lawvere or simply Bill, as we affectionately called him, asked me to contribute some remarks in the mathematical/philosophical journal of which he is a co-editor, I accepted with enthusiasm, forgetting about the difficulties involved in such a task. In fact, I owe a great deal to Bill, both mathematically and philosophically and I can present only a small part of my debt.

As I mentioned in my Review Article "Topos Theory in Montréal in the 1970s : My Personal Involvement" [1], I had come to Montréal in 1967 with a PhD in Logic from Berkeley. As far as I remember, category theory was never mentioned in my years as a graduate student.

Much later I discovered that Lawvere had been in Berkeley from 1961 to 1963 as an informal student, following lectures by Alfred Tarski and Dana Scott before completing his PhD at Columbia in 1963 under the supervision of Samuel Eilenberg. Furthermore, Lawvere attended the 1963 International Symposium on "The Theory of Models" meeting in Berkeley, where he presented his functorial development of general algebra, and announced that quantifiers are characterized as adjoints to substitution. His article : "Algebraic Theories, Algebraic Categories, and Algebraic Functors" was published in the 1963 Proceedings of "The International Symposium At Berkeley" [2].

My first and most important mathematical contact in my new city was André Joyal. This contact was to prove determinant for a great part of my career. At the time of my arrival, Joyal was a student at Montréal University, working alone his way through Grothendieck's "Éléments de Géométrie Algébrique" [3] and several volumes of the "Séminaire de Géométrie algébrique du Bois Marie" [4]. I believed that it is where he learned category theory

and toposes. I learned category theory mostly from him and he learned some model theory from me.

During these earlier years we studied Lawvere's and Benabou's articles whose aim was to do Universal Algebra categorically. And by 1970/1971, several algebraic structures such as groups and rings were "categorized". Joyal and I decided to extend their ideas to do first-order logic categorically, following the basic discovery of Lawvere that started categorical logic, namely that the categorical counterparts of existential and universal quantifiers are left and right adjoints to pull-backs (the counterpart of substitution). These are called images and dual-images respectively.

As Marquis and Reyes remarked in "The History of Categorical Logic (1963-1970)" [5], this was a key observation that convinced many mathematicians that this was the right analysis of quantifiers. They arise naturally as adjoints to an elementary operation, namely substitution, which appears as the basic operation of first order logic, contrary to the classical view which defines this operation by recursion, as a derived one.

Recall that if $f : X \longrightarrow Y$ is a function between sets X and Y and B is a subset of Y, then the pull-back of B along f is the subset of X defined by $f^*(B) = \{x \in X : f(x) \in B\}$. By considering B as a predicate of Y, say $B(y)$, the pull-back may be considered as the predicate of X obtained from $B(y)$, by substituting $f(x)$ for y. Lawvere's discovery was that the pull-back formation

$$f^* : P(Y) \longrightarrow P(X)$$

which is a functor between posets has a left and a right adjoints, namely the existential and the universal quantifiers

$$\exists_f \dashv f^* \dashv \forall_f$$

Indeed, by defining

$$\begin{cases} \exists_f(A) = \{y \in B : \exists_x(f(x) = y \text{ and } x \in A\} \\ \forall_f(A) = \{y \in B : \forall_x(f(x) = y \text{ implies } x \in A\} \end{cases}$$

it is easy to check that

$$\exists_f(A) \subseteq B \iff A \subseteq f^*(B)$$

which is the meaning that $\exists_f \dashv f^*$. Similarly, one can check that $f^* \dashv \forall_f$ namely

$$f^*B \subseteq A \iff B \subseteq \forall_f(A)$$

$M\times\Phi$ **vol. I.2** © 2023

Lawvere's ideas motivated Volger to do first-order logic categorically in 1971 [6], but his attempt was not fully successful. Although we partly followed him, we realized that his mistake was to incorporate Ω, the sub-object classifier in the category associated to a theory, an object that does not belong to first-order logic. Once that we realized this, the road was open to finish the job. This was done by my student Jean Dionne in his Master thesis [7], not without solving some difficulties along the way.

Of all the results that Joyal and I obtained in the 1970s, the most important and most widely quoted was the existence of the classifying topos of an arbitrary coherent theory. It was discovered as an attempt to solve the main *contradiction* between Logic and Geometry in topos theory that Lawvere had formulated as the search of certain adjoint functors, see [8]. For a more complete exposition see [9].

$M\times\Phi$ **vol. I.2** © 2023

In the meanwhile, Lawvere had obtained in 1969 the prestigious Killam professorship at Dalhousie University in Halifax, and was in that context allowed to invite collaborators in the subject of category theory. This meant that during 1969/1971, Dalhousie became the most important center of category theory. In particular, the notion of elementary topos was developed there.

However, in 1971, the group of collaborators at Dalhousie was dispersed; the university administration refused to renew Bill's contract due to his political activities. In fact, Bill had protested vigorously against the War Measures Act proclaimed by Pierre Elliot Trudeau, the Prime Minister of Canada, in October 1970.

The War Measures Act was the response to the kidnapings of a British diplomat, James Cross and a Québec Minister, Pierre Laporte by "Le Front de Libération du Québec" (FLQ), a militant Québec independence movement. Pierre Laporte was then killed the 17th of October and James Cross released on the 3rd of December after 60 days of captivity.

The War Measures Act gave the Government powers of arrest, detention and censorship for the third time in Canadian History and the first time in peacetime, hence, suspending civil liberties under the pretext of danger of terrorism.

My first encounter with Bill took place at the International Congress of Mathematicians in 1970 in Nice. I was in Nice by chance, waiting for a visa for Warsaw where I was supposed to study that year with Mostowski. This visa never arrived and I became stranded in Europe with my wife, Marie. At one point of

our journey in Europe I met Adrian Mathias, whom I had known in Berkeley. He obtained an invitation for me to give some talks in Cambridge. This was the first time I gave talks on category theory.

The next time that I met Bill was in April 1973 in Montréal, where Bill spent that month, away from Perugia. At that moment in Québec, there was a lot of political activity from inside and outside influences. Small political movements invaded universities, labor unions and journals with discussions between Marxists-Leninists, Stalinists-Maoists, Mouvement Révolutionnaire des Étudiants du Québec, etc. From the beginning of the month discussions were suffused with dialectical materialism explained to us by Bill, with its emphasis on the universality of contradictions, main contradictions and so on. (See my paper Topos Theory in Montréal in the 1970's : My Personal Involvement [1]). This was an intense, informal gathering, not an organized, planned meeting.

One must not suppose that category theory evolved as a long, quiet, mathematical river. Bill had a very rich, complex and integrated mathematical, philosophical and political personality. His presence in a room was always felt, imposing but never arrogant or pretentious. He listened carefully to anybody who wanted to speak with him. He was very generous with his ideas and very serious in his beliefs that he defended with vigor.

The following year there was another meeting in Montréal which Bill attended. It took place in the summer 1974 and was organized by Shuichi Takahasi in the context of the "Séminaire de Mathématiques Supérieures de l'Université de Montréal". During this meeting, Bill carried out an incredible activity, giving talks, suggesting new approaches, encouraging some people and criticizing others. Some of these criticisms were carried out by means of "mass democracies" (a mass democracy is a collective accusation against a "reactionary" point of view of somebody). All these discussions and meetings left an unforgettable imprint on all the participants. The atmosphere was electric and sleep was very often missed.

Besides all this activity, Bill attended some meetings of the Communist Party of Canada (Marxist-Leninist). Once, he heard a long talk given by Hardial Baines, the First Secretary of the Party that made a great impression on him. Of course, all this activity took a toll and when Bill returned home to Buffalo, his wife Fatima told us that "she had to pick him up with a spoon". Lawvere had close relations with the Communist Party of Canada (Marxist-Leninist). I don't know whether he was a member, but the party wrote his eulogy.

$M\times\Phi$ **vol. I.2** © 2023

This can be found in TML Monthly Newspaper of the Communist Party of Canada (Marxist-Leninist) No.1 January 2023.

Thanks mainly to the efforts of Bill, the theory of categories became very popular in Europe, North America, especially in Montréal, where a weekly seminar of categories lasted many years and also in Latin America, where I was invited to give several talks and seminars. In Bogota, people from the Universidad Nacional de Colombia and I organized a category meeting with Bill and Anders Kock as the special guests.

In 1975-1976, I obtained a sabbatical year which I spent mainly in Aarhus, Denmark. This was the beginning of a long collaboration with Anders Kock, the first student of Bill. From there on, I went to Aarhus so often during the summer, that once a professor seeing me in the corridor said : "Reyes !... It must be summer !"

During this and the following years Bill came often to Aarhus. Many people came to listen and some to work with him. During these stays I had the opportunity to deepen my understanding of his ideas that were not always easy to grasp. I often had the impression that his mind was meandering around the subject and I was surprised to see him arrive to the heart of the matter. He had an originality of thought that kept on surprising me.

A subject, among others, that shows Bill's originality and his way of working is "Synthetic Differential Geometry", also called "Smooth Infinitesimal Analysis".

It was the discovery of Lawvere that a "Grothendieck topos" may be viewed as a universe of "variable set" and that consequently set theoretic language can be interpreted directly in a topos. Therefore working from the topos built from schemes, rather than with the schemes themselves, one obtains a model for this generalized set theory with nilpotent infinitesimals or "Synthetic Differential Geometry".

To understand the motivation, we notice that "Differential Calculus", as we know it today was developed in the 19th Century by people like Weierstrass and others on a basis that was rigorous but far from being intuitive. This was not the way that the creators of the infinitesimal calculus in the 17th century (Newton and Leibniz among others) thought about these notions.

For instance, to define a tangent to a curve, they thought of the curve as the assemblage of an infinite number of infinitely small straight lines ; or (what is the same thing) as a polygon with an infinite number of sides, each of an infinitely small length, which

$M\times\Phi$ vol. I.2 © 2023

determines the curvature of the line by the angles they make with each other. This is rather complicated. Lawvere and Kock, on the other hand, turned the tables and considered an infinitesimal curve as a unique straight line. In symbols,

$$\forall f \in R^D \; \exists! \, (a,b) \in R \times R \; \forall \, h \in D \; (f(h) = bh + a)$$

where R is the ring of reals (not a field) and D the set of infinitesimals, namely, $D = \{x \in R \mid x^2 = 0\}$.

Indeed, the equation of a line is $y = bh + a$, where b is the slope and a is the initial value $f(0)$. Then the previous formula can be re-written as the Kock-Lawvere axiom or K-L axiom

$$\forall f \in R^D \; \exists! \, b \; \forall d \in D \; (f(d) = bd + f(0))$$

This says that every infinitesimal curve f is uniquely a straight line with slope b, *i.e.*

$$f(d) = bd + f(0)$$

Starting from here, Kock in his book "Synthetic Differential Geometry" [10] shows how the usual rules of differentiation can be derived.

The Kock-Lawvere axiom, however has an unexpected consequence : it is contradictory! as shown by Schanuel and Lavendhomme. Schanuel argument is given in Kock's book "Synthetic Differential Geometry" and is an example of the "law of the excluded middle" : define a function $g : D \longrightarrow R$ as follows :

$$g(d) = \begin{cases} 1 & \text{si } d \neq 0 \\ 0 & \text{si } d = 0 \end{cases}$$

If the Kock-Lawvere holds, $D = \{0\}$ is impossible. Hence, by using the law of the excluded middle, we may assume the existence of some non-zero $d_0 \in D$. By the Kock-Lawvere axiom,

$$\forall d \in D \; (g(d) = bd + g(0))$$

Substituting d_0 for d we obtain $1 = g(d_0) = bd_0 + 0$, which, when squared yields

$$1 = 0.$$

Lavendhomme in his book "Basic Concepts of Synthetic Differential Geometry" [11] proceeds by showing that under the hypothesis of the Kock-Lawvere axiom, it follows that R=0, the zero ring.

$M\times\Phi$ vol. I.2 © 2023

As Kock says, the moral of this story is that the Kock-Lawvere axiom is incompatible with the law of the excluded middle and hence one or the other must leave the scene.

I believe that this dilemma forces the law of excluded middle to leave the scene and forces us to weaken the logic, using intuitionistic or constructive logic rather than the classical one.

Lawvere also discovered that by considering smooth versions of the toposes occurring in algebraic geometry (toposes built from rings of smooth functions rather than polynomials) one obtains models for ordinary differential geometry. In these models infinitesimal structures of the kind used by Cartan, for instance, can be interpreted directly and, in this context, Cartan's arguments are literally valid. These ideas, dating from 1967, remained unpublished, and were taken up only in the mid-seventies. This resulted in two main lines of development. On the one hand, there was the purely axiomatic development of differential geometry with nilpotent infinitesimals, or "synthetic differential geometry" as developed by Kock [10] and Lavendhomme [11]. On the other hand, smooth toposes were constructed which show not only the consistency of the axiomatic approach but also provided a direct connection with the classical theory of manifolds. This second line was developed in the book "Models for Smooth Infinitesimal Analysis" written by Ieke Moerdijk and myself [12]. This approach avoids the three limitations of the category of manifolds discussed in the preface of our book. The basic ideas of this approach are due to Lawvere and can be seen to originate from the work of C. Ehresmann, A.Weil and A. Grothendieck . The aim is to construct categories of spaces, the so called "smooth toposes" which contain the category of manifolds (or more precisely, there is a full and faithful embedding of the category of C^∞-manifolds into each of these smooth toposes). Moreover, in each of these smooth toposes, *inverse limits* of spaces and *function spaces* can be adequately constructed, in particular, *infinitesimal spaces* like the ones needed in our interpretation of Cartan's arguments, e.g., the space D of first-order infinitesimals.

Lawvere had many followers and he was a source of problems at different levels of difficulty that he distributed freely among them. One could always find a problem that one could handle. One day, while in Denmark, Kock arrived at the office that was assigned to me, with a problem proposed by Bill that looked very interesting and maybe solvable : show that second

© 2023 $M\times\Phi$ vol. I.2

order differential equations (ODE's) and many similar prolongation structures in a topos constitute another topos receiving an essential morphism from the first, provided that certain fractional exponents exist. We succeeded in proving this, partly following Bill's suggestions, and he announced our solution in Appendix 8 of "Outline of Synthetic Differential Geometry" [13] (see the list of "The collected works of F. W. Lawvere" : https ://github.com/mattearnshaw/lawvere/tree/master/pdfs).

Among the interests of Bill were also the history and teaching of mathematics.

In the list of "The collected works of F. W. Lawvere" there are several titles on the history of mathematics. One example : In "Grassmann's Dialectics and Category Theory", Bill says : "Looking more closely into Grassmann, Stephen Schanuel and I found numerous ways in which it could be justified to say that Grassmann was a pre-cursor of category theory. The general algebraic operations which he discussed have become the explicit object of a particular developed theory, and those general concepts, general operations, system of operations and systems of equation in invariant coordinate free form have been made into a part of category theory. More specifically, we find that in certain cases the famous distinction between analytic and synthetic operations can only be explained in terms of adjoint functors." (lawvere/pdfs/1996-grassmans-dialectics-and-category-theory.pdf p. 256)

Other examples can be found in the list of publications of Bill's works on the following authors :

Euler : "Euler's continuum Functorially vindicated"

Hegel : "Display of graphics and their applications, as exemplified by 2-categories and the Hegelian "taco".

Peano : "The relation of the so-called natural numbers to real mathematics".

Another aspect that I would like to mention is his passion for the teaching of Mathematics. In particular he wrote a book, and a collection of notes and papers.

The book, "Conceptual Mathematics (A First Introduction to Categories)", was written with his friend and colleague Stephen H. Schanuel, with the assistance of Emilio Faro and Fatima Fenaroli, Danilo Lawvere and the students of Mathematics 181 at the University at Buffalo. This book was translated into Spanish by Francisco Marmolejo with the help of Ivonne Pallares [14].

$M\times\Phi$ vol. I.2 © 2023

In this book, the authors said "Our goal is to explore the consequences of a new and fundamental insight about the nature of mathematics which has led to better methods for understanding and using mathematical concepts. While the insight and methods are simple, they are not as familiar as they should be ; they will require some efforts to master."

"Categorical Foundations of Set Theory and Logic" [15] is the title of a collection of notes and papers. In his foreword, Lawvere gives a preliminary description of the significance of each of the words in the title.

In particular, he explains "Foundations" as making explicit the essential general features, ingredients and operations of a science, its origins and its general laws of development. The purpose of making these explicit is to provide a guide to the learning, use, and further development of the science. A "pure" foundations which forgets this purpose and pursues a detached speculative "foundations" for its own sake is clearly a *non* foundations."

© 2023 $M\times\Phi$ **vol. I.2**

In 1985 I attended an intriguing lecture by John Macnamara, professor of Psychology at McGill University. He raised the question of giving a logical account of the phrase "Freddie is a dog" that children learn by the age of 18 months, claiming that logical theories were unable to account for it. I thought that he was mad : how can there be a problem with that ? But I found his talk so intriguing that I went to discuss with him at the end. This was the beginning of a collaboration between John Macnamara, Marie La Palme Reyes, Houman Zolfaghari and myself that lasted until John's death in 1996. I gave some talks about this subject in the category seminar in Montréal when Bill was present and, contrary to my expectations, he got interested in the subject and suggested several improvements. Furthermore, he accepted to participate in the Vancouver Meeting on "The logical foundations of Cognition" with a paper on the "Tools for the Advancement of Objective Logic, Closed categories and Toposes" that was published in the book "The logical foundations of cognition" [16]. His conclusion was that "Thus, I believe to have demonstrated the plausibility of my thesis that category theory will be a necessary tool in the construction of an adequately explicit science of knowing."

In 2007 I received an invitation to participate in the Boston Meeting on "Trends in the Mathematical Representation of Space : Philosophical and Historical Perspectives" to take place on November 31 and December 1, 2007. I was very surprised, since

I had never published anything on this subject, until I discovered that Bill was responsible for this invitation. In fact, I remembered that I had discussed with Bill the approach of R.K. Sachs and H. Wu to obtain Einstein field equations together with some simplifications that I obtained by using infinitesimals. My contribution, "A Derivation of Einstein's Vacuum Field Equations", was published in [17].

Looking back at my relation with Bill, I cannot but marvel at all what I learned and enjoyed through our conversations, our discussions and our common love of mathematics and philosophy. To accept that all this marvelous period has ended is very painful, however all the memories will live on in those who had the good fortune to know and appreciate him.

§ — References.

[1] Reyes, G. E. [2019] Topos Theory in Montréal in the 1970s : My Personal Involvement, Review Article, *History and Philosophy of Logic*, Taylor Francis Group.

[2] Lawvere, F. W. [1965] Algebraic Theories, Algebraic Categories, and Algebraic Functors, *The Theory of Models*, Proceedings of the 1963 International Symposium at Berkeley.

[3] Grothendieck, A. and J. Dieudonné [1960] *Éléments de Géométrie Algébrique*, Volume 1, Le Langage des Schémas, Paris : Publications mathématiques No 4, Institut des Hautes Études Scientifiques.

[4] Grothendieck, A. [1972] SGA 4 (Volumes 269, 270), Séminaire de Géométrie Algébrique de Bois Marie dirigé par M. Artin, A. Grothendieck, J. L. Verdier, Springer Verlag.

[5] Marquis, J. P. and G. E. Reyes [2010] The History of Categorical Logic 1963-1977, *Handbook of the history of Logic : Sets and Extensions in the Twentieth Century*, Volume editor : Akhihiro Kanamori, General editors : Dov M. Gabbay, Paul Thagard and John Woods, © Elsevier BV, All rights reserved.

[6] Volger, H. [1971] *Logical categories*, Thesis, Dahlousie University, Halifax, N.S., Unpublished manuscript.

[7] Dionne, J. [1973] *Des théories élémentaires aux catégories conceptuelles*, Thèse de maitrise, Université de Montréal.

$M\times\Phi$ vol. I.2 © 2023

[8] Reyes, G. E. [1974] From Sheaves to Logic, *Studies in Algebraic Logic*, Volume 9, Aubert Daigneault, editor,The Mathematical Association of America.

[9] Makkai, M. and G. E. Reyes [1977] *First Order Categorical Logic*, (LNM611), Springer-Verlag.

[10] Kock, A. [1981] *Synthetic Differential Geometry*, London Mathematical Society, Lecture Notes 51, Cambridge University Press.

[11] Lavendhomme, R. [1996] *Basic Concepts of Synthetic Differential Geometry*, Kluwer Academic Publishers.

[12] Moerdijk, I. and G. E. Reyes [1991] *Models for Smooth Infinitesimal Analysis*, Springer-Verlag.

[13] Lawvere, F. W. [1998] Outline of Synthetic Differential Geometry, Buffalo Geometry Seminar notes, February.

[14] Lawvere, F. W. and S. Schanuel [1991] *Conceptual Mathematics*, Buffalo Workshop Press.

[15] Lawvere, F. W. [1992] Categorical Foundations of Set Theory and Logic, Collection of notes and papers.

[16] Macnamara, J. and G. E. Reyes [1994] *The Logical Foundations of Cognition*, Edited by John Macnamara and Gonzalo E. Reyes, Oxford University Press.

[17] Reyes, G. E. [2011] A Derivation of Einstein's Vacuum Field Equations, Volume 53 of CRM Proceedings and Lecture Notes, Centre de Recherches Mathématiques de Montréal, *Models, Logics, and Higher-Dimensional Categories*, A Tribute to the work of Mihály Makkai, Edited by Bradd Hart, Thomas G. Kucera, Anand Pillay, Philip J. Scott, Robert A.G.Seely.

© 2023 $M \times \Phi$ **vol. I.2**

Gonzalo E. Reyes,
Université de Montréal

Annals of Mathematics and Philosophy
***vol.* I, *no.* 2**, 2023

$M \times \Phi$

Does Reason Have Limits?

Review Essay: Stephen Budiansky,
Journey to the Edge of Reason: The Life of Kurt Gödel

Palle Yourgrau

Abstract. Stephen Budiansky has produced the second major biography of Kurt Gödel, after John Dawson's classic study. A comparison is made here between the two books, the latter, a more scholarly exercise — emphasizing Gödel's work — the former intended for a wider audience — emphasizing Gödel's life, in particular his private life. Budiansky, after so much has now been written about Gödel and the intellectual climate of the twentieth century, had an opportunity to reflect on just who Gödel was, on what he represented as one of the towering intellectual figures of not just the twentieth but of any century. Unfortunately, it will be argued, this is an opportunity lost. An attempt will be made, however, to do, in this regard, what Budiansky failed to do — to explain why Gödel to this day remains strangely misunderstood and not fully appreciated for who he really was. His time, it seems, has still not come.

© 2023

§ — General Remarks.

Stephen Budiansky's new biography of Kurt Gödel complements John Dawson's classic, academically oriented study. As Bernd Buldt writes in a comprehensive review of Dawson, "[He] has resisted the temptation to reach too wide an audience and he doesn't even pretend to be popular ... [His is not] a popular but a scholarly account of Gödel's life and work ..." Exactly the reverse is true of Budiansky. Dawson's is "a very austere ... chronicle of Gödel's life ... [T]here are no statements concerning possible 'influences' on Gödel that made him think or act this way or another. ...

[For example,] Dawson raises 'an interesting question' ... regarding '[h]ow relevant [his social and educational] milieu] was to [Gödel's] own social or intellectual development,' but refuses to answer it ..."[1] Once again, the exact opposite is true of Budiansky. Again, Dawson frequently interrupts the historical narrative in order to expound, in some detail, on the mathematical background of Gödel's work, which makes sense, given the purpose of his book. By contrast, given the alternative goal of Budiansky's, it makes sense that he offers us a continuous, uninterrupted narrative, more likely to grip the general reader than Dawson's.

Budiansky's book will, no doubt, enjoy a wide appeal, in part because, unlike Dawson's, the salient feature of his study is the attention paid to Gödel's life — in particular, to Gödel's private life — as opposed to his work (as evidenced by the titles of the two books). Indeed, Budiansky's book could have been entitled: *The Private Life of Kurt Gödel*. Whatever one makes of this fact, it remains that it is, as one says, "a good read." It is well written, far from austere, unified, at times dramatic, historically informed, historically oriented. To describe it as "popular", however, can be misleading. Budiansky's narrative is, throughout, serious and thoughtful. The book is clearly written "in a good will" (to borrow from Wittgenstein).[2] This, I would stress, is no small virtue.[3]

Given the importance and prominence of Dawson's biography, however, and the obvious comparison that would inevitably be made with Budiansky new biography, there is, to my mind, too little attention paid by Budiansky to Dawson's study — a fact made disconcerting by his odd, seemingly depreciatory reference to Dawson as a "mathematics teacher at a community college".[4] Budiansky also makes too little use, I believe, of Hao Wang's two

[1] Buldt 1995, 386. Page numbers refer to a pre-print of this review.

[2] From the "Forward" to Wittgenstein 1975.

[3] A virtue sorely lacking in Goldstein 2005, which is popular precisely in the invidious sense of the term. See the devasting review in Feferman 2006, which ends with these words: "Those who are fascinated by Gödel's theorems ... should not be satisfied with Goldstein's 'vast and messy' goulash: hers is not a recipe for true understanding." The two biographies to be read about Gödel are by Dawson and Budiansky.

[4] Dawson is Professor Emeritus in the Mathematics Department of the Pennsylvania State College at York, which is how, except for Budiansky, he is always described. I have never seen him referred to elsewhere as a "math teacher at a community college." Would it be appropriate to refer to Gödel simply as a researcher without students at a small institute in a tiny town in New Jersey?

$M\times\Phi$ vol. I.2 © 2023

books[5] devoted to his conversations with, and thoughts about, Gödel's life and work. (More on this, below.) As for the title of Budiansky's book, I've suggested a possible alternative, but the actual title, *Journey to the Edge of Reason*, invites reflection. Does reason (Plato's and Aristotle's *noûs*) have an "edge"? Can reason be over-used? Can one be *too rational*? Consider Gödel's comments about reason or intellect in a letter to his mother: "I don't think it is unhealthy to apply the intellect [*Verstand*] to any area [whatever] (as you suggest). It would also be unjustified to say that in just this very area [religion] nothing can be accomplished with the intellect. For who would have believed 3,000 years ago that one would [now] be able know how big, how massive, how hot and how far away the most distant stars are ..."[6]

Is Wang, however, right when, with Gödel in mind, he contrasts the rational with the reasonable: "... [B]eing reasonable seems to include ... a flexibility ... associated with being sensible, and moderate, not excessive. ... [T]he rational should not depart conspicuously from the reasonable. ... [B]eing 'fanatically rational' is, for me, no longer rational."[7] Can one, however, be "fanatically rational"; can reason be overused? One might think so, remembering Aristotle's emphasis in the *Nicomachean Ethics* on the importance of "moderation", on the so-called "golden mean." If one looks more closely, however, at what Aristotle actually says, it emerges that, for Aristotle, one can't be too x, if x is a virtue, an excellence. Bravery, for example, is a virtue. One can't be too brave, though one can be too *rash*. (Similarly, mutatis mutandis, with vices. One can't be adulterous to the right degree, i.e. "moderately", as Aristotle humorously notes.) What is often described (especially with regard to someone like Gödel) as excessive rationality is thus more aptly characterized as irrationality — as misapplying rationality — just as what might be characterized

© 2023 $M \times \Phi$ vol. I.2

[5] Wang 1987 and Wang 1996.

[6] Wang 1996, 108.

[7] Wang 1987, 225. Yet Wang himself, it should be noted, says that, "Gödel exemplifies, I think, a way of life and work that inspires greater faith in reason ..." (Wang 1997, IX) Indeed, C. Anthony Anderson has claimed that Alonzo Church's contributions "place him high up among the most important philosophers of this century ... and make a persuasive case for the suggestion that Alonzo Church be taken for the denotation of the definite description, 'The most rational man since St. Thomas Aquinas' — a description which Church is said to have reserved for Kurt Gödel." (Anderson 168–69).

as a soldier's being too brave is more aptly described as being reckless or rash. Isn't Gödel right when he comments [in concert with Aristotle[8]] that, "Reason itself doesn't err ... Every error is due to extraneous factors (such as emotion and education)."[9]

Solomon Feferman's characterization of Gödel is, unfortunately, neglectful, or at best unclear, about this distinction. He says that Gödel "had a superb all-embracing rationality which could descend to a maddening attention to detail in matters of everyday life."[10] Grammatically, what "descends" is said here by Feferman to be Gödel's rationality. Moreover, to describe Gödel's rationality as "all-embracing" is itself ambiguous. Is the suggestion that Gödel extended his rationality beyond its proper place? Can rationality itself, however, "descend"? Is there a place where rationality doesn't belong? Wasn't Gödel right when he said to his mother, as we saw, that "I don't think it is unhealthy to apply the intellect [*Verstand*] to any area [whatever] (as you suggest)." More accurate would be to describe Gödel *himself* as sometimes descending from his rationality, as sometimes *misapplying* his intellect, rather than applying it where it doesn't belong. Wasn't Gödel also right when he said, "[i]n our time, *rationalism* is used in an absurdly narrow sense ..."[11] In the end, is there really anything surprising, paradoxical, or even noteworthy in the fact that someone can be maximally rational in their work, but at times quite irrational in their life? (Indeed, isn't it a bit of folk lore that logicians and mathematicians are famously known to have more than their share of eccentrics or unbalanced personalties?)[12]

[8] Aristotle, *De Anima*, Bk I, Ch. 4, 408b 25: "Thus thought and contemplation decay because something else within is destroyed, while thought in itself is unaffected." That Gödel was a close student of such passages in Aristotle is clear from Eva-Maria Engelmen 2016, "What is the Link Between Aristotle's Philosophy of Mind, the Iterative Concept of Set, Gödel's Incompleteness Theorems and God?".

[9] Dawson 1996, 241 (brackets added).

[10] Feferman 1986, 2.

[11] Wang 1996, 266.

[12] In the opposite direction, Feferman (Feferman 98), having stressed Gödel's extreme caution (especially, early in his career) in expressing his true thoughts, draws attention to what seems to be a parallel fact, that "Gödel was very careful in his personal habits, especially as concerns his health", and comments that "the caution discussed here [with regard to Gödel's work] is coherent with other characteristics of his personality." (163). He suggests that "only an in-depth biography could plumb the common sources and establish the interrelationships of the everyday and the scientific personality (for which, see now Dawson 1996)." In fact, Dawson doesn't go very deeply into the possibility of such interrelationships.

$M\times\Phi$ vol. I.2 © 2023

§ — Preliminary Concerns.

At the beginning of his Prologue, Budiansky refers to Gödel's "staggeringly brilliant and paradoxical proof that no formal mathematical system will ever capture every mathematical truth within its own bounds." Note first that Budiansky fails to point out that Gödel's incompleteness proofs concern only *consistent* formal systems. An inconsistent formal system, of course, can prove everything expressible in that system, including its own consistency. (A pedantic point, to be sure, but one that's always worth mentioning,) Second, what exactly is *paradoxical* about Gödel's result? Is there something paradoxical in the idea that formalizing an axiomatic system imposes constraints on it? Wang records Gödel being irritated by what he took to be Wittgenstein's interpreting his result as a paradox: "How can Wittgenstein consider it as a paradox if he had understood it?"[13] Note further that Gödel sharply distinguishes the idea of a formal system from that of an unformalized system of axioms, and emphasizes that the limitations proved by his theorems are a result of the constraints imposed by strict formalization. Wang writes in this regard that "[Gödel's] conception of an axiom system is more liberal than the precise concept of a formal system. ... [H]e would probably say that, although his theorem shows that mathematics is not a *formal* system, mathematics can be captured by an axiom system."[14] Further, Wang comments that "the contrast between formal systems and the axiomatic method as Gödel

$M\times\Phi$ vol. I.2 © 2023

Budiansky, by contrast, does. But just this fulfillment of Feferman's psychologistic suggestion or speculation I find suspect. Though clearly Feferman had no intention of demeaning Gödel, such speculations, developed by Budiansky, have the effect of demonstrating that one's hero, after all, had feet of clay, that, for example, Gödel "berated himself for his shortcomings: lying in bed too late in the morning, overestimating himself and underestimating others, 'laughing, talking in a high voice, making faces,' indecisiveness, 'nerves and insecurity in company.'" (Budiansky 2021, 103) Hegel has called out such historical psychologists. "These psychologists," he writes, "are particularly fond of latching on to the peculiarities of great historical figures as private persons. A man must eat and drink, he enters relations with friends and acquaintances, he has feelings and moments of anger [as his valet reveals]. ... Served by such psychological valets in historical writing, the historical personage comes off badly; he is brought down ... even a few degrees below the morality of these fine connoisseurs of humanity." (Hegel 1988, 34–35; brackets added) To be sure, to apply Hegel's comments to Budiansky is, admittedly, a bit too harsh; nevertheless, the general point, I believe, remains.

[13] Wang 1996, 84.

[14] Wang 1996, 334; emphasis in the original. It should be added that "Gödel does not take an axiomatic system as an 'implicit definition' of the concepts in it, because it is supposed to be a report of our (generally) incomplete intuitions of

appears to understand them may be seen as a fundamental aspect of what might be called Gödel's dialectic of the formal and the intuitive ... While Gödel is best known for his definitive results on the limitations of formalization, his primary interest is not in what has been formalized but in clarifying our larger intuitions."[15] (More, below, on Gödel's dialectic of the formal and the intuitive.)

On that same page, Budiansky repeats a famous description of Gödel's incompleteness theorem(s) as "the most significant mathematical truth of the century," without explaining who said this, and why. In fact, it was said by W.V. Quine on the occasion of Gödel receiving an honorary doctorate from Harvard University on June 17, 1952. As to what Quine meant by those words, Gödel provided an explanation to his mother. In Wang's words, "[Gödel] points out that the citation should not be taken to say that he is the greatest mathematician of the century, but, rather, that the phrase *most significant* means 'of the greatest *general* interest outside of mathematics."[16]

Whereas Budiansky provides too little context concerning Quine's remark, he provides, to my mind, too much historical context concerning Vienna, Brno, and the Austro-Hungarian monarchy, and even the Vienna Circle. Some of what Budiansky has to say about the Vienna Circle in relation to Gödel is valuable,

$M\times\Phi$ vol. I.2 © 2023

the concept, and can be revised and expanded." (Wang 1987, 247) This explains why, after the proofs of Gödel and Cohen had established the independence of the Continuum Hypothesis from the axioms of Zermelo Fraenkel set theory, Gödel, unlike most set theorists, believed that the next step was to find (intuitively justified) supplementary axioms that would decide the Continuum Hypothesis. To date, however, no such axioms have been found.

Note that "unformalized", or "informal", does not mean "*imprecise*". "The ideal of [Gödel's] philosophy," writes Wang, "is to be precise but not technical." (Wang 1987, 239).

[15] Wang 1987, 201.

[16] Wang 1996, 2. Quine, I believe, was right. In Yourgrau 2005, Ch. 4, I argued that the Leitmotif of the 20th century was "formalism", in the most general sense — the dominance of form over content, syntax over semantics, in art, architecture, music, physics, and, crucially, in logic and mathematics. Gödel's theorems (proved formally!) are surely the most important result concerning Hilbert's mathematical formalism, the purest form of formalism of the century. Those who have gone out of their way to point out that Gödel's theorems have not, after all, altered the course of mainstream mathematics may be right, but that's the wrong direction in which to look to assess the ultimate significance of those theorems. Indeed, "[even] among mathematicians who are not logicians," Dawson has commented," there are still many, including some of the highest stature, who have little awareness or understanding of what Gödel accomplished." (267).

but the amount of detail he provides constitutes, to my mind, overkill. And did we really need, for example, to read on page 21 the minute details of the annual Corpus Christi procession as presented in Joseph Roth's *The Radetzky March*, which include "the blood-red fezzes on the heads of the azure Bosnians burn[ing] in the sun like tiny bonfires", and so on, line after line, for nearly a full page? Indeed, I kept wondering, as I began reading the book, if we would ever get to Gödel.[17] Moreover, this topic has by now surely been over-mined. Besides classics like Carl Schorske's *Fin-de-Siècle Vienna*[18], there are the vivid accounts that have recently appeared by Karl Sigmund[19] and David Edmonds.[20] To be sure, Budiansky himself goes out of his way to draw attention to his indebtedness to Sigmund, but that doesn't keep him from repeating much of what Sigmund has already written. Of course, the two books don't completely overlap in this regard, but what overlap there is strikes me as excessive. This is especially disconcerting, since Budiansky notes, ironically, that Gödel himself said, "I don't consider my work a 'facet of the intellectual atmosphere of the early 20th century', but rather the opposite. It is true that my interest in the foundations of mathematics was aroused by the 'Vienna Circle', but the philosophical consequences of my results ... are anything but positivistic ... So my work points toward an entirely different world view ..."[21] By contrast, arguably, Wittgenstein is, to a much larger extent, a product of Viennese culture, as brought out by Allan Janik and Stephen Toulmin in *Wittgenstein's Vienna*.[22]

More disturbing are the detailed excerpts from Gödel's diaries and the notes of his psychiatrist which, though revealing, seem at times too personal, verging on the prurient. Gödel, after all, was an immensely private, reserved, even secretive person. There is something tasteless, now that he's dead and buried, in rifling through papers that he would have been horrified to see exposed to public

[17] To be fair, this is a constitutional problem of biographies. Like someone taking as long a run-up as possible to get a good jump off the diving board, biographers typically start their run as far back as possible into the pre-history of their main character, as if learning that his or her great great grandfather was a cloth-merchant who used to spend hours practicing the flute shines reflective light into the distant future.

[18] Schorske 1980.

[19] Sigmund 2017.

[20] Edmonds 2020.

[21] Budiansky 2021, 101.

[22] Janik and Toulmin 1973.

© 2023 $M\times\Phi$ vol. I.2

scrutiny. Indeed, Budiansky himself remarks on "[Gödel's] dread of revealing his inner self and failings to the scrutiny of others," but then proceeds, anyway, to reveal Gödel's inner self and failings to the scrutiny of others, as well as passages from Gödel's notebooks concerning details of his sexual life with his wife Adele.[23] There is something unclean about this. To be sure, by their very nature biographies tend to be voyeuristic, but Budiansky takes this too far.

By contrast, Budiansky is to be praised for his remarkably balanced portrayal of Adele, who has too often been the subject of mean-spirited caricature.[24] Budiansky points out, for example, that Georg Kreisel, Alfred Tarski, and Verena Huber-Dyson "all admired Adele's high spirits, wisecracking observations about Princeton life, and bantering style with her husband, which he seemed to enjoy." As Kreisel remarked, "she had little formal education, but a real flair for the *mot juste*."[25] Budiansky might have added something Dawson pointed out, that she was also "a talented seamstress" who "[was] said ... to have made a child's dress each day of the war for donation to the relief effort. Her contributions were so highly valued that after the war the Viennese gave her a bust of her father (who had died, out of touch with her, during the war years) as a token of appreciation."[26]

Budiansky is also to be commended for drawing attention to the fact that "Gödel's last years in Princeton as a reclusive loner have cast a long shadow over the received portrait of his personality. But that is not the Kurt Gödel his friends knew through his life,"[27] nor was that true of the years Gödel spent in Vienna before his move to Princeton. Indeed, nothing could be more unlike the cloistered intellectual atmosphere of the Institute for Advanced Study than the international cultural vibrancy of Gödel's Vienna, with its fertile coffeehouse culture. Budiansky also brings out the fact that the picture of a somber, unsmiling Gödel (as portrayed on the cover of Budiansky's book) should not overshadow the fact that Gödel may well have had a dry sense of humor, so dry that it could evaporate before his interlocutor realized what had happened. A good illustration of this is Budiansky's description of an episode the logician Gerald Sacks describes when he once talked to Gödel about his

[23] Budiansky 2021, 175
[24] As in Goldstein 2005.
[25] Budiansky 2021, 176.
[26] Dawson 1997, 162.
[27] Budiansky 2021, 108.

$M\times\Phi$ vol. I.2 © 2023

favorite philosopher, Leibniz. Sacks was taken aback when Gödel told him that "... Leibniz was wrong about everything," and then added, "[b]ut it's just as hard to be *wrong* about everything as to be *right* about everything." Was Gödel being serious, or was he pulling Sacks' leg? "He didn't crack a smile," said Sacks, "but I thought, this was a very amusing remark. And I thought, Wow, he cracked a joke. Or did he?"[(28)]

Another nice touch by Budiansky is the description he provides of Gödel's lecture style when teaching. We know from Carnap's "Intellectual Autobiography" that Gödel's great predecessor Gottlob Frege's back was usually turned to his students. "He seldom looked at the audience. Ordinarily we saw only his back, while he drew the strange diagrams of his symbolism on the blackboard and explained them."[(29)] Gödel's lecture style was similar, but with one important difference. Budiansky notes that a student at one of Gödel's lectures at Princeton said that "while the lecture was good, Gödel continued facing the blackboard the entire time, even though he did not write on it once."[(30)]

As for Gödel's general appearance, Budiansky notes that on arriving by ship in New York in 1933, Gödel was described on the list of arriving alien passengers as "height, five feet seven inches; hair, brown; eyes, brown; complexion, dark."[(31)] Curiously, however, in Dawson we're told that Gödel was described on arrival as being 5'7" tall, of light complexion.[(32)] Dark complexion or light? Who is right?

Concerning the appearance of Budiansky's book, I must say I found the cover quite striking, with a high resolution closeup of a famous, evocative photograph of Gödel.[(33)] By contrast, Dawson's cover, with an image of a "night-moth", is nothing less than bizarre. True, the name of the nightclub where Adele used to dance was "*Der Nachtfalter*" ("The Night-Moth"), but what has that to do with the subject of Dawson's book, Gödel's *Logical Dilemmas*? (Was Adele's dancing at that club one of the logical dilemmas Gödel had to resolve?) I do have, however, a reservation of another sort with Budiansky's cover. That same photographic portrait of Gödel

[(28)] Budiansky 2021, 105–6.
[(29)] Carnap 1963.
[(30)] Budiansky 2021, 159.
[(31)] Budiansky 2021, 146.
[(32)] Dawson 1997, 150.
[(33)] The resolution of photographs inside the book, however, is disappointing.

$M\times\Phi$ vol. I.2 © 2023

adorns the cover of Wang 1987 and is in the background of the cover of Wang 1996. Does not a new biography deserve a new cover? If Budiansky were going to borrow from Wang, it would have been better if he had instead borrowed more of the *content* of Wang's books, which remains, to this day, one of the richest sources of our knowledge of Gödel's philosophy. Moreover, there are striking quotations from Gödel provided by Wang that are cause for reflection. To take just one example, at one point Wang says, "Gödel was in favor of metaphysics and opposed positivism. When I asked him about the work of Saul Kripke[34], he said, 'Kripke is, though not a positivist, still doing linguistic philosophy.' "[35] This will, I suspect, surprise many people. After all, there was a few years ago an article in the Times Literary Supplement by Arif Ahmed, entitled, "Saul Kripke: The Return to Metaphysics". I think I know, however, what Gödel had in mind with that comment, and to an extent, I agree with him, but that's a discussion for another time.

To be fair, however, Dawson, too, fails to sufficiently avail himself of the rich material about Gödel provided by Wang (though he's far better in this regard than Budiansky). Indeed, there has been a tendency in recent years to be skeptical of the authenticity and/or depth of Wang's reflections on Gödel, and, to my mind, more generally, a lack of appreciation of Wang himself as a philosopher. A good corrective is to be found in a fair and moving portrait of Wang in Mar 2015.

§ — More Serious Issues.

Turning to the technical content of Budiansky's book, one finds several glitches. For example, he states that, "... Russell proposed that the way to think of the number 2 is simply as the definition of all sets that contain a pair of objects."[36] As it stands, this sentence makes no sense. Russell didn't propose that the number 2 is a *definition*. He proposed a definition *of* the number 2. And the definition he proposed is that the number 2 is a certain set of sets. Which set of sets? Not simply the set of all sets containing a pair of

[34] Kripke, by the way, was for a while a colleague of Wang's at The Rockefeller University, until the university — primarily an institute for scientific research — closed the department, which contained a who's who of distinguished philosophers, creating a diaspora that enriched a number of other universities.

[35] Wang 1996, 138.

[36] Budiansky 2021, 108–9.

$M\times\Phi$ vol. I.2 © 2023

objects, i.e. the set of 2-membered sets, since that would obviously be circular, defining 2 in terms of 2. Rather, Russell, following Frege, defined the notion of a 2-membered set using purely logical concepts. (After all, his program, like Frege's, is known as logicism.)

More. Concerning the inapplicability of the principle of excluded middle (*tertium non datur*) to infinite sets or sequences, Budiansky writes that, "[Brouwer's] favorite example, one that Wittgenstein later often used, was to ask how one could possibly prove or disprove the proposition that somewhere in the infinite decimal expansion of pi the sequence 7777 occurs. Not finding it could always be answered with the objection that one has just not looked far enough."[(37)] But of course, if the sequence 7777 *does* occur in the decimal expansion of pi, one could prove this simply by calculating the decimal expansion (for which there's an algorithm) and coming upon 7777 after finitely many steps. So, if the sequence 7777 does occur, it's *provable* (whether or not in fact proved at a given time.)[(38)] What Wittgenstein actually said was, " 'In the decimal expansion of pi either the group 7777 occurs or it does not — there is no third possibility.' That is to say: God sees, but we don't know.' But what does that mean?"[(39)] Since Wittgenstein, like Brouwer, is in effect a mathematical verificationist[(40)], truth is inseparable from, indeed, is constituted by proof, so for him there is no fact of the matter (even before God's eyes) as to what's "really" going on in the decimal expansion of pi, whether or not we know it. And at a given time, we may have no proof either that 7777 occurs or that it does not occur, but only a proof that it hasn't occurred so far as our calculations have proceeded.

Still more. Concerning the so-called Gödel formula, G, in Gödel's incompleteness proof, Budiansky says, "As Gödel, observed, this means that G is formally *undecidable*: it is impossible to construct with the system a proof of either G or of its negation,

© 2023 $M\times\Phi$ vol. I.2

[(37)] *Op. Cit.*, 114. It's not essential, but it should be pointed out 7777 was not the series Brouwer mentioned.

[(38)] Mark van Atten has pointed out to me that it has in fact been shown that 7777 occurs at position 1589 in the decimal expansion of pi. I thank him, more generally, for helping me to clarify Brouwer's views on this issue.

[(39)] Wittgenstein, 1973, 352.

[(40)] See the illuminating study of Brouwer, van Atten 2004, in which Brouwer's intuitionism is characterized as, in effect, a kind of mathematical verificationism. Van Atten brings out, clearly, how crucial the role of time is in Brouwer's conception of mathematics. Reflecting on this, I've drawn attention in Yourgrau 2019 to what I take to be Brouwer's "time problem."

not-G".[41] He forgot to add the crucial caveat: "assuming the system is consistent". An inconsistent, system, of course, *can* prove G, since it can prove everything expressible in the system.

And again. We read, "... [Gödel] was hard at work putting into final form his complete proof of the independence of the Continuum Hypothesis, which he already had published two short notices about in the Proceedings of the National Academy of Sciences."[42] Gödel, of course, never did construct a proof of the independence of the Continuum Hypothesis, but rather of its consistency with the axioms of Zermelo Fraenkel set theory.

Still more. Budiansky states,"If the Continuum Hypothesis is both consistent with the axioms of set theory (thus not *disprovable*) and independent of them (thus not *provable*) that would affirm the inadequacy of existing theory ..."[43] This should have read: "If the Continuum Hypothesis is both consistent with the axioms of set theory (thus not *disprovable* from them) and also not *provable* from them (thus independent of them), that would affirm the inadequacy of existing theory ..."

Also problematic is that Budiansky gives a strangely brief account of Gödel's writings on Einstein's theory of relativity and the reality of time.[44] Dawson's discussion, by contrast, is far more complete. And what Budiansky does say is unsatisfactory. He mentions, for example, what Gödel says about the analysis of time in special relativity (STR) without noting that Gödel is contrasting the intuitive, pre-relativistic notion of time with what is called "time" in the formalism of STR[45]. He goes on to say that "there remains in the standard cosmological solutions to Einstein's equations a privileged, and thus in a sense objective, 'cosmic time' ...", without noting that a) the discussion has now moved on from special to general relativity (GTR), and b) that now the contrast is between intuitive time and cosmic time. He continues: "In Gödel's [nonstandard cosmological] model, the universe is not expanding ...", without drawing attention to the fact that Gödel constructed several cosmological models, in some of which the universe is not

[41] *Op. cit.* 130.
[42] Op.Cit. 210.
[43] *Op. cit.* 214.
[44] *Op. cit.* 226–27.
[45] This is an important distinction, since, as I argue below, what Wang described as the dialectic of the formal and the intuitive is an overarching theme of Gödel's investigations.

$M\times\Phi$ vol. I.2 © 2023

expanding, and some in which it is. Budiansky then records Gödel arguing that in the non-expanding model, time-travel is possible, which, even if the actual universe is expanding, (which it is) "renders meaningless the notion of the passage of time in any objective sense", without explaining a) just why time-travel, in itself, rules out the objective passage of time[46], and b) that at issue here is the intuitive concept of time (which, according to Gödel, implies the passage of time).[47]

Budiansky fails, also, to mention the fact that Gödel employs here a modal argument, from possibility to actuality[48], an argument form that he employs elsewhere, most famously in his construction of a classic ontological argument for the existence of God, moving from the possibility of God to God's actuality. As a mathematical Platonist, moreover, Gödel would naturally have endorsed a move from mathematical possibility to mathematical actuality.[49] Also disturbing is that Budiansky ends his discussion

© 2023 $M\times\Phi$ vol. I.2

[46] For an explanation of why, for Gödel, if there is time-travel, there is no time (in the intuitive sense), see Yourgrau 1991 and 2005.

[47] Budiansky's source for his account of Gödel's world models, William Bonnor's 1964 book, *The Mystery of the Expanding Universe: The Humorous, Caustic and Illuminating Story of What Men Think About Their Exploding World*, which contains a very brief account of Gödel's cosmology, is an odd choice, given the vast literature on this subject that exists today. Indeed, I've seen no reference to it in that vast literature. For some reason, Budiansky ignores the classic discussions of Gödel's cosmology in essays by Howard Stein, David Malament, and Lawrence Sklar, as well as the more recent discussion in Earman 1995, in which Earman responds in part to Yourgrau 1991, even though Dawson pointed out in 1997 that concerning "the ramifications of [Gödel's] cosmological work for the philosophy of time, [t]o date, only a single volume (Yourgrau *1991*) has examined the latter in any detail." (269) Indeed, even though Dawson remarked again, in 2016: "Surprisingly ... only one extended study of Gödel's contribution to the philosophy of time (as opposed to the technical papers in relativity theory) has so far appeared: Palle Yourgrau's *The Disappearance of Time*. (footnote: Yourgrau 1991. Later reissued in expanded from as *Gödel Meets Einstein: Time Travel in the Gödel Universe* ...)" (24–25) That this remains, even today, according to Dawson, the only book devoted to this subject is an indication, I think, of the low repute, or perhaps simply lack of interest, that still attaches to Gödel's philosophical reflections on time. Especially disconcerting, since Wang reports that for Gödel, in a sense, time is *the* philosophical problem.

[48] Gödel's modal argument for the ideality of time has proved elusive for many philosophers. Even as rigorous a philosopher of physics as Earman (*op. cit.*), for example, has, in my view, mis-formulated it (and thus falsely criticized it). See Yourgrau 1991, 92.

[49] See Yourgrau 1991 and 2005 and van Atten, "Gödel, Mathematics and Possible Worlds," in van Atten 2015.

by mentioning that Kreisel "thought he perceived 'between the lines' of Gödel's essay a hint of his fascination with the idea of ghosts that could reappear from the past." I found this reference to Kreisel's rather flippant comment demeaning of a subtle philosophical argument by Gödel, which is supported by the impressive formal contributions he made to relativistic cosmology. (See below.)

Also problematic is Budiansky's drawing on the mathematician Abraham Fraenkel's memoirs[(50)] to paint a one-sided portrait of the famous intuitionist, L.E.J. Brouwer — who had a complex relationship with Gödel[(51)] — as "a committed Aryan nationalist" who was "temporarily suspended from his university position as punishment for collaborating with the Nazis during their occupation of Holland" (80). Budiansky is apparently unaware of what Dirk van Dalen has written in this regard in his classic biography of Brouwer.[(52)] Based on van Dalen's account, van Atten writes that, "[In 1934–1935] Ludwig Bieberbach, a German member of the editorial board of *Compositio Mathematica* who in 1933 had joined the National Socialist movement by entering the SA, asks Brouwer to remove the Jewish members of the board. Brouwer refuses; Bierberbach resigns. ... During the Nazi occupation of the Netherlands, Brouwer assists the resistance by letting them use storage space in the Mathematical Institute, of which he was the director. He offers hiding places to Jews ... in small houses on his premises. .... In 1943, he advises ... students to sign the declaration of loyalty demanded by the Germans; when some students refuse and go into hiding, Brouwer and Heyting visit them to continue their education."[(53)] It would have been good if Budiansky had looked into this issue more closely and had drawn attention not only to Fraenkel but also to van Dalen and van Atten.

§ — Avenues Not Explored.

Budiansky mentions, at the end, briefly, that "[Gödel's] "relativity papers and philosophical insights have begun to receive

[(50)] Fraenkel 2016.
[(51)] For details, see van Atten 2015.
[(52)] Van Dalen 2005.
[(53)] van Atten 2020; brackes added.

$M\times\Phi$ vol. I.2 © 2023

considerably more respectful attention than they did from physicists and philosophers during his lifetime."[54] This mild statement grossly understates the situation. In point of fact, Gödel's relativity papers and the temporal idealism he derived from them were, and to a considerable extent still are, shamefully dismissed or derided, or both. Note, to begin with, that Gödel himself, in a bibliography of his writings, listed among his five most important published results, "rotating universes; its mathematical aspect, and its philosophical aspect."[55] Einstein agreed. According to Gödel's friend Oskar Morgenstern, "Einstein told me that Gödel's papers were the most important ones on relativity theory since his own [Einstein's] original paper appeared."[56] More recently, the distinguished cosmologist G.F.R. Ellis commented that "the issue of the beginning of the universe is still the subject of intense debate; the Hawking-Penrose theorems have set the parameters within which the discussion takes place. Those theorems owe much to Gödel's papers."[57]

Yet, as Budiansky himself points out, as late as 1978, at a memorial service for Gödel, Harry Woolf, then Director of the Institute for Advanced Study, in preparing notes for who should speak, wrote, "(minor: relativity — not worth a talk." Earlier, there had been a more dramatic dismissal. The noted astrophysicists S. Chandrasekhar and J.P. Wright published a paper in 1961 in the prestigious *Proceedings of the National Academy of Sciences*,[58] in which they claimed they had found an error in Gödel's results, since he alleged that "time-travel" in one of his rotating universes would take place along a geodesic. As Dawson comments, though Gödel was a member of the National Academy of Sciences, he wasn't consulted about this challenge, which, it turned out, was mistaken. Yet when the philosopher Howard Stein submitted an article refuting Chandresekhar and Wright, it was only accepted for

[54] Budiansky 2021, 278 (brackets added).

[55] Wang 1996, 93. Eric Audureau notes that Gödel "considered these ... papers as a whole. ... Thus, from the inner, relative (to Gödel), meaning of 'Gödel Philosopher', cosmology should belong to philosophy: his work on general relativity is an application of a philosophy suggested outside of general relativity and obtained on the occasion of thinking about it. ... [H]e believes that the usual cosmological notions of force, inertia, motion, time and so on, are philosophical matters." (Audureau 2016, 58–59)

[56] Moore 1990, 93.

[57] Ellis 1996, 46.

[58] Chandrasekhar and Wright 1961.

© 2023 $M\times\Phi$ vol. I.2

publication in 1970[59], after Gödel himself personally intervened. And the trend continues. As late as 2005, the noted Einstein scholar John Stachel, in a lengthy review of Yourgrau 2005 in *Notices of the American Mathematical Society*, dismissed Gödel's cosmological models as "... an example of that fetishism of mathematics to which some Platonists are so prone."[60] The distinguished philosopher of mathematics Charles Parsons, in turn, an expert on Gödel, went out of his way to express his approval of Stachel. "Stachel," he wrote, "presents a physically informed discussion whose point of view seems in harmony with what I'm trying to say here."[61]

As for the philosophical conclusions vis a vis the ideality of time Gödel believed were supported (NB, not proved) by his cosmological results, these have been accorded a no less unfriendly reception. Stachel, unsurprisingly, had no more patience with them than he had for Gödel's cosmological models. Meanwhile, the distinguished philosopher of science John Earman, after noting the "relative neglect of the philosophical moral Gödel ... himself wished to draw from his famous solution to EFE [Einstein's Field Equations]," concluded that "the neglect has been benign."[62] In Yourgrau 2005, I protested that claim by Earman, which prompted Stachel to affirm his solidarity with Earman: "I agree with John Earman's claim — which so horrifies Yourgrau — that the philosophers' neglect of Gödel's views is 'benign'."[63] I wasn't the only

[59] Stein 1970.

[60] The moral of Stachel's piece seems to have been, "When it comes to physics, Einstein good; Gödel bad." Yet, as we've seen, Stachel is here contradicting what Einstein himself said about Gödel and relativity. And as for "fetishizing mathematics", the charge might more plausibly have been made against Einstein, in his later years. As Howard Isaacson writes in his recent biography of Einstein, "Now, in [Einstein's] quest for a unified theory, there seemed to be a lot of mathematical formalism, but very few fundamental physical insights guiding him." As Einstein's collaborator Banesh Hoffmann commented, "…. the search was not so much a search as a groping in the gloom of a mathematical jungle inadequately lit by physical intuition." While Jeremy Bernstein said it was "like an all but random shuffling of mathematical formulas with no physics in view." (Isaacson 2007, 468). Yet Stachel chose to describe Gödel's physics, as opposed to Einstein's, as "a fetishism of mathematics."

[61] Parsons 2010, 182, footnote 32.

[62] Earman 1995, 194–5; brackets added.

[63] Stachel, *op. cit.*, 865. I should add that I was also "horrified" by Earman's dismissal of Gödel's attempt to grasp the essence of the intuitive concept of time: "This [Gödel's reasoning about the intuitive concept of time] is a pretty piece of ordinary language reasoning. But, like most of its ilk, it leaves one up in the air ... [O]ne can wonder how such intuitions can support such weighty philosophical

$M\times\Phi$ vol. I.2 © 2023

one, however, who was put off by Earman's attitude. (Criticism, of course, is one thing; dismissal, another.) The philosopher Mauro Dorato wrote that "[Gödel's] argument is important not only for the conclusion it — unsuccessfully, as we shall see — tries to support, but much more for the brilliant analysis of controversial philosophical notions that it provides; from this point of view, it has certainly *not* received the attention it deserves. Besides Yourgrau's pioneering work on Gödel's philosophy of time [1991, 1999], there are now *two* conflicting reconstructions of Gödel's argument for the ideality of time. Savitt's [1994], which endorses it, and Earman's [1995, pp. 194–200], which rejects it and *somehow considers it unworthy of much attention*."[(64)] And more recently yet, in his review of Yourgrau 2005 in the prestigious scientific journal *Nature*, the logician Martin Davis, an expert on Gödel, dismisses Gödel's argument with no more than a wave of the hand. He writes that "[Yourgrau] argues that because time fails to exist in these universes and because the differences between such universes and our own are accidental, time can't exist in our world either. I doubt many readers would be convinced by this argument."[(65)]

© 2023 $M \times \Phi$ vol. I.2

morals. Gödel's essentialist intuitions are not easy to fathom. ... This game of using an inner sense to perceive conceptual truths is a dangerous one. ... [In his response to Gödel,] Einstein brushes aside the question of the relation of GTR to idealistic philosophy. This seems to me to be the correct response to Gödel." (Earman 1995, 195–200; brackets added.)

[(64)] Dorato 2002, 257; brackets in the original; final emphasis added. Savitt, it should be noted, later reversed his position on Gödel's argument.

[(65)] Davis, 2005, Vol. 435, 19–20. As far as philosophical refutations go, "I doubt many readers would be convinced by this argument" does not, surely, rank very high. (Note, too, that Davis curiously attributes Gödel's argument to me.) His curt dismissal of Gödel's reasoning based on Gödel's analysis of the essence of the concept of time is surprising, given his important essay, "Why Gödel's Didn't Have Church's Thesis" (Davis 1982). There, Davis sang a different tune with regard to Gödel's search for the essence of an intuitive concept, in that case, the concept of effective calculability. Davis commented that Alonzo Church, in a letter to Stephen Kleene, related that Gödel told him [before the advent of the concept of a Turing machine] that "it might be possible, in terms of effective calculability as an undefined notion, to state a set of axioms which would embody the generally accepted properties of this notion ..." (9; brackets added) Davis went on to remark that "now, this is very much in line with what Turing accomplished. Although his treatment was not axiomatic in any formal sense, he did manage to show that 'generally accepted properties' of effective calculability lead inevitably to a definite class of functions (which turned out to be the same as lambda-definable or recursive functions)." (14)

Is all this, however, just a tempest in a teapot, which Budiansky had no need to take note of in such detail? Am I being prejudiced here, since, so to speak, I have a dog in this fight? I think not. It is indicative of a curious and disturbing phenomenon: the widespread failure to take seriously Gödel as a philosopher. (More on this below.) Saying only, as Budiansky does, that Gödel's philosophical insights have begun to receive considerably more respectful attention than they received previously, does not, I believe, begin to do justice to this disconcerting phenomenon. Why in the world would so many be so willing to dismiss Gödel, of all thinkers, to disrespect (vs. simply to criticize) him when he ventures to go beyond his strictly formal results, or beyond his primary field of research?

To be sure, Budiansky does mention, at the end of his book, the fact that today, "conferences and prestigious publication programs are being devoted to both topics [Gödel's relativity papers and philosophical insights]."[(66)] Once again, however, details matter, and once again, details are lacking. Given the dismissal of Gödel's contributions to philosophy that was a theme of the Boston University Conference in 1995 entitled "Gödel's General Philosophical Significance" that was drawn attention to in Yourgrau 2005 (162–164), Budiansky would have done well to mention a particular recent conference with a very different emphasis, entitled "Gödel's Legacy: Does Future Lie in the Past?", at the University of Vienna, sponsored by the Kurt Gödel Society of Vienna, from July 25 to July 27, 2019, celebrating the 70th anniversary of Gödel's contributions to relativity theory and the 100th anniversary of the confirmation of Einstein's GTR. Budiansky

Indeed, Turing himself made it clear that that he was similarly attempting an analysis of the essential nature of the concept of effective calculability: "In elementary arithmetic," he wrote, "the two-dimensional character of the paper is sometimes used. But ... I think it will be agreed that the two-dimensional character of paper is no essential of computation." (Turing 1936–7, 135). Obviously, Turing was maintaining that the other elements of his analysis *were* essential to the concept. In their attempt to discover the essence of the concept of effective computability, Gödel and Church (and, from his sympathetic account, Davis) were clearly not subscribing to what we saw earlier was Earman's position with regard to Gödel's analysis of intuitive concepts (in Earman's case, the intuitive concept of time): "This is a pretty piece of ordinary language reasoning. But, like most of its ilk, it leaves one up in the air ... [O]ne can wonder how such intuitions can support such weighty philosophical morals" — including such weighty philosophical morals as Church's thesis?

[(66)] *Op. cit.* 278, brackets added.

$M\times\Phi$ vol. I.2 © 2023

would also have been well advised to draw attention to the appearance in 2016 of a remarkable collection of essays entitled, *Kurt Gödel: Philosopher Scientist*[67] (a clear echo of the famous volume edited by P.A. Schilpp entitled *Albert Einstein: Philosopher-Scientist*[68]).

Budiansky would have done well, also, to document more fully Gödel's (in)famous conceptual realism and mathematical Platonism together with his attendant idea of mathematical intuition. The first question here is whether one should believe Gödel when he said that he had been a mathematical Platonist since 1925.[69] Skeptics about the accuracy of the date of 1925 include the logicians Solomon Feferman, Martin Davis, and Donald Martin. Budiansky refers to this controversy, but fails to mention that Dawson had already drawn attention to it. Dawson pointed out[70] that the question turns on the correct placement of a comma. The comma occurs in a sentence in a lecture Gödel gave in 1933, which asserted that the axioms "if interpreted as meaningful statements, necessarily presupposed a kind of Platonism, which cannot satisfy any critical mind." Dawson points out that crucial to the interpretation of that sentence is whether the placement of the comma before the word "which" reflected merely Gödel's unfamiliarity with the niceties of English grammar. In his biography, Dawson himself doesn't attempt to resolve the Question of the Comma. Nor does Budiansky.

Buldt, by contrast, in his review of Dawson[71], takes a stab at it. He develops an argument for the thesis that the placement of the comma was simply a grammatical mistake. His first point is based on a kind of principle of charity. This interpretation avoids the conflict that arises otherwise between Gödels's different claims. (And shouldn't one avoid, if possible, charging Gödel, of all people, with inconsistency?) His second point is that "the comma is required by German grammar in the corresponding German sentence" and that in 1933, Gödel, at the time, was probably not fully aware of the fine points of English grammar concerning the proper placement of

© 2023 $M\times\Phi$ vol. I.2

[67] Croco and Engelen eds. 2016. To be sure, this important collection is cited in Budiansky's Bibliography, but surely it should have been drawn attention to in the text.

[68] Schilpp, ed. 1949.

[69] Gödel 1933.

[70] Dawson 1997, 100.

[71] Buldt 1995.

a comma.[72] His third point, Buldt says, is the most important one: "[W]hen in his later years Gödel argued for Platonism the direction of the argument was one from intuition to axioms, justified by these intuitive insights: objects first — axioms second, so to speak. But the platonism he speaks of at the passage in question is of another kind; for in this context Gödel starts out with axioms as uninterpreted and meaningless formulae which are given meaning only afterwards: axioms first — objects second. This way of faking platonism must have been unsatisfying for Gödel. ... [S]uch an added ontology will never enjoy the richness of the 'original' one experienced by intuition and will never be able to be more than a mere and much poorer substitute."[73]

Buldt's reasoning, however, is not the last word. The plot thickens when one takes into account a draft of a passage written by Gödel in 1933, cited by van Atten, in which Gödel, having raised some issues concerning Heyting's intuitionist logical foundations, comments that they are "at least more satisfactory than the ordinary Platonistic interpretation ..."[74] Notice should also be taken of the defense by Martin Davis (Davis 1998) of his skeptical attitude toward Gödel's account of the history of his relationship with mathematical Platonism. At the same time, one should consider the reservations concerning Davis's defense found in Tait 2001, 100–102.

And there's much more Budiansky could have done to explore the question of Gödel's Platonism and his concomitant belief in mathematical intuition. For example, he could have cited the curious fact of the widespread disdain, amounting at times to contempt, for this aspect of Gödel's philosophy, as expressed, for example by Charles Chihara.[75] He could have pointed, too, to Michael Dummett, who, in his study of Frege's philosophy of mathematics, gave the back of his hand to Gödel's attempt to provide an epistemology for his mathematical Platonism.[76] He refers to Gödel's "postulation of a faculty of intuition of abstract objects, in analogy with the perception

[72] Over time, Gödel's grasp of the subtleties of English grammar increased greatly. When preparing translations of Gödel's essays to be included in his seminal collection (van Heijenoort 1967), van Heijenoort "... exchanged a total of seventy letters and met twice in Gödel's office in order to resolve questions concerning subtleties in the meanings and usage of German and English words." (Dawson, 1997, 217)

[73] Buldt 1995, 398.

[74] van Atten 2018, 7.

[75] Chihara 1973, Ch. II: "Gödel's Ontological Platonism", 60 -83.

[76] Dummett 1991, 301–302.

$M\times\Phi$ vol. I.2 © 2023

of material objects," and concludes his discussion by remarking on "the lameness of that analogy."[77] "By contrast," he says, "Frege, in virtue of his logicism, had none of these objections to face."[78] He forgets that, Frege, like Gödel, was a conceptual realist, and so he, too, needed to "postulate" such a faculty to account for our access to concepts and other abstract entities, like senses.[79]

More recently, the distinguished set theorist Donald Martin has also dismissed Gödel's notion of mathematical intuition as a kind of perception.[80] He cites a famous passage from Gödel in which he says that, "... despite their remoteness from sense experience we do have something like a perception also of the objects of set theory ... I don't see any reason why we should have less confidence in this kind of perception, i.e. in mathematical intuition, than in sense perception."[81] Martin objects to the reference, here, by Gödel to the perception of the objects of set theory, but he also objects to Gödel's idea that we have something like a perception of concepts. "[Concerning the] perception of concepts," he says, at the close of his essay, "... it is hard to see what such perception is supposed to yield that is not yielded by garden-variety understanding of the concept."[82] I'll comment, shortly, on Martin's criticism, but I note first that he curiously fails to remark on the fact that the distinguished logician and philosopher Alonzo Church made, in a well-known essay years before Gödel's remark, almost the exact same observation that Martin finds such fault with. "The preference," writes Church, "of (say) seeing over understanding as a method of observation seems to me capricious. For just as an opaque body may be seen, so a concept can be understood or grasped. And the parallel between the two is in fact rather close."[83] This remark by Church is especially striking given

© 2023 $M\times\Phi$ vol. I.2

[77] Dummett raises a good question about why, if there is such thing as mathematical intuition, proof is so dominant in mathematics. Gödel, however, has a lot to say about this, which Dummett neglects to point out. William Tait, too, in "Truth and Proof: The Platonism of Mathematics" (Tait 1986), has a lot to say about this. One fact, however, should be obvious. Proof ultimately rests on axioms, primitive concepts and rules of inference. These, of course cannot be proved. How, then, is their truth or reliability to be vouchsafed?

[78] *Op. Cit.* 302.

[79] See Yourgrau 1982 and 2012 and Kripke 2011.

[80] Martin 2005.

[81] Gödel 1990, 268.

[82] Op.Cit. 227; brackets added.

[83] Church 1951

Wang's comment that, "[Gödel] asserts, in analogy to our perception of physical objects, our ability to *perceive* concepts too. It is necessary to remind ourselves, however, that by this he means primarily our ability to understand and see that certain assertions about concepts are true. He also regards perceiving concepts ... as a kind of *observation*." (84) Indeed, in an essay on Church's contributions to philosophy and logic, C. Anthony Anderson has remarked that "it would be interesting to know if there was any influence, in either direction, between Church and Gödel on these matters."(85)

It is clear, then, that for Gödel, mathematical intuition involves the perception of concepts, and that by perceiving or grasping a concept he means understanding it. In Gödel's words, "[w]e perceive objects and understand concepts. Understanding is a different kind of perception."(86) Yet Dawson, commenting on this view of Gödel's, can still say, after noting Gödel's belief in the possibility of telepathy, that, "[w]hatever one may think of such beliefs [of Gödel's] in extrasensory perception, it is important to note how well they accord with Gödel's belief that mind is distinct from matter, and with his later advocacy of mathematical Platonism".(87) Does Gödel's sympathy with the idea of telepathy really "accord well" with his belief in our ability to understand mathematical concepts? According to Chihara, the answer is yes. "I think," he says, in his critique of Gödel, "we should look for some sort of natural explanation before we turn to postulating nonphysical processes and extra-sensory perception."(88)

$M\times\Phi$ **vol. I.2** © 2023

(84) Wang 1996, 13; emphasis in the original.

(85) Anderson 1998, 135, footnote 26.

(86) Wang 1996, 235.

(87) Dawson 1997, 30; brackets added.

(88) Chihara 1973, 79. Indeed, "Gödel's reasoning," says Chihara, "could lead to a massive population explosion of our ontology: if we use mathematical intuitions to postulate mathematical objects, it would seem as if we could use 'theological intuitions' to postulate theological objects like angels." (81). He derides the ontological argument for the existence of God, and suggests that Gödel might "express [an] insight by saying that *God is a necessary being* ... [adding] that this axiom forced itself on him as true." (78; emphasis in the original; brackets added). Ironic, since, Gödel, of course, did actually formalize an axiomatic presentation of the ontological argument, which has stimulated a great deal of discussion. Once again, I emphasize how dangerous it is to dismiss out of hand Gödel's philosophical explorations. One can disagree with Gödel's philosophy, of course, but then one can also disagree with Plato, Aristotle, Leibniz, Kant, or Wittgenstein, without thereby simply dismissing them or minimizing their importance as philosophers.

Budiansky, in turn, comments that "... [Gödel] was using the term ["intuition"] to suggest that the human mind can *literally* see mathematical realities through a kind of perception *no different from* the direct sensory perceptions the empiricists decreed to be the only valid basis of physical laws."[89] Of course, Gödel precisely did not say we can *literally* see mathematical realities. (Was Richard the Lion Heart *literally* a lion? Was he *no different from* a lion?) Gödel says: a is *analogous* to b. Budiansky says Gödel is claiming: $a = b$.

But what about Martin's protest about Gödel speaking of mathematical intuition and the perception of concepts — as opposed to, in Martin's words, merely "garden variety understanding of concepts?"[90] We need to keep in mind that Gödel was a conceptual realist, which accounts for his analogy with sense perception. "... [W]e begin with vague perceptions of a concept," Gödel said, "as we see an animal from far away or take two stars for one before using a telescope. For example, we had the precise concept of mechanical procedure in mind, but had not perceived it clearly before we knew of Turing's work."[91] Further, Wang reports that, "according to him [Gödel], it is not the concepts that change, but rather our perceptions of them."[92] Of course, if one isn't a conceptual realist, grasping a concept would be understood as mastering the use of a word.[93] For Gödel, however, " 'trying to see (i.e. understand) a concept more clearly' is the correct way of expressing the phenomenon vaguely described as 'examining what we mean by a word.' "[94] What Gödel says here should remind one of what another conceptual realist, Frege, had to say. "What is known," says Frege, "as the history of concepts is really a history of either our knowledge of concepts or of the meanings of words. Often it only after an immense intellectual effort, which may have continued over centuries, that humanity at last succeeds in achieving knowledge of a concept in its pure form, in stripping off the irrelevant accretions which veil it from the eyes of the mind."[95]

© 2023 $M\times\Phi$ vol. I.2

[89] *Op. cit.*, 266; emphasis added.

[90] I tried to answer this question in Yourgrau 2005, 173–4.

[91] Wang 1996, 235. Note that sometimes, as Gödel suggests, it turns out there was no one precise concept there in the first place, just as sometimes what we think of as one star is in fact two stars, seen dimly, from far away.

[92] Wang, 1987, 63.

[93] Is that what Martin has in mind by "garden variety understanding of concepts"? It certainly was what Wittgenstein and W.V. Quine had in mind. But, of course, Gödel disagrees with such accounts.

[94] Wang 1987, 63.

[95] Frege 1980, vii. Note that as a conceptual realist, Frege, like Gödel, means by knowing the meaning of a word, having knowledge of or grasping a concept.

The two different accounts of what it is to understand the meaning of a word correspond to two different roads leading from Frege[96]: one leads to Gödel and Church and conceptual realism, the other to Wittgenstein, Dummett, and the so-called linguistic turn in philosophy. Gödel, clearly, took the first road. Dummett, like most philosophers indebted to Frege, has taken the second road. (Indeed, I think it's fair to say that Dummett's "Wittgensteinian Frege" has dominated contemporary accounts.) As far as epistemology is concerned, there are arguments in Yourgrau 1982 and 2012 that Frege relied implicitly on something like Russell's notion of acquaintance. Frege, however, never developed that account or accounted for the role of the subject. It's not surprising, therefore, that, in later life, in his search for an epistemology to match his conceptual realism, Gödel turned from Frege to Edmund Husserl, a conceptual and mathematical realist who, in his phenomenology and transcendental idealism, attempted to account for the role of the subject.[97] Note that by turning toward Husserl's transcendental idealism, Gödel was accepting the idea that for a different kind of object to be grasped, a different kind of (or aspect of) mind was required.[98]

Along the same lines, Karl Popper, interestingly, engaged in what Julius Moravcsik called "Platonistic psychology", arguing from what the mind *knows* to what the mind *is*, as opposed to what could be called "empirical psychology", which argues (based on introspection and behavioral evidence) from what the mind (supposedly) *is* to what the mind is capable of *knowing*[99].

$M\times\Phi$ **vol. I.2** © 2023

[96] See Yourgrau 2005b.

[97] For an in-depth account and analysis of Gödel's turn to Husserl see van Atten and Kennedy 2003.

[98] See Tragesser 1977.

[99] Moravcsik 1981, 109. Similarly, the distinguished Plato scholar Harold Cherniss: "... there are two things in which Plato is more interested than in the theory of ideas itself, for that theory is, after all, only his way of satisfying these two requirements: first, there is such a thing as mind, which can apprehend reality, and second, this reality which is the object of knowledge has absolute and unqualified existence." (Cherniss 1962, 83) (Cited in Wang 1996, 11) And Frege, a conceptual realist, like Gödel, drew attention to the consonance between the (mathematical) mind and its objects: "In arithmetic we are not concerned with objects which we come to know as something alien from without through the medium of the senses, but with objects given directly to our reason, and, *as its nearest kin*, utterly transparent to it." (Frege 1980, 115; emphasis added).

Some have argued that this passage demonstrates, rather, the fact that Frege wasn't after all a true Platonist, a conceptual realist, but if Frege wasn't, neither was Plato: "When the soul investigates by itself," Socrates says in the *Phaedo*, "it

Popper[100] echoed Frege in postulating three worlds, the physical, the mental, and a third world of abstract intelligibles, all of which can be "grasped" by the mind. In effect, agreeing with those who are skeptical that the mind, as we know it, could grasp the third world, Popper suggested that we need to revise our conception of the mind. "I suggest," he says, "that one day we will have to revolutionize psychology by looking at the human mind as an organ for interacting with the objects of the third world."[101]

The tendency, thus, to dismiss, ignore, or downplay Gödel's philosophical explorations is ill advised — and this is true even when Gödel's thoughts are being expressed in a far from academic context. (So to speak, Gödel is always, as a thinker, "on the clock.") A particularly interesting example of this can be found in Dawson's biography. Dawson[102], mentions Gödel's letter to his mother in which he says that "had God created in our place beings who did not need to learn, these beings would not just be *we* ... That is, we would not exist at all." According to Dawson, this reasoning is akin to the controversial anthropic principle, and asserts that "in essence, [Gödel's] argument begs the question by asserting that God created us as we are because otherwise we wouldn't *be* what we are, and so *we* wouldn't exist at all. ... [T]he discussion seems rather naïve for one so philosophically sophisticated." Once again, however, we have someone being too hasty in dismissing Gödel's philosophizing. There is no hint in Gödel's reasoning to suggest that that is *why* God created us. His point is rather that we shouldn't complain that God could have created a better, indeed, a best possible world, since if God had done so, we — being the kind of beings we are — wouldn't have existed. Should we be rebuking God for creating us? Far from this being a question begging, naïve argument, it strikingly resembles what Robert Adams, one of the leading contemporary analytical philosophers of religion and a noted philosopher of modality, has written in his well-known essay, "Must God Create the Best?"[103] (And this in an informal letter by

$M \times \Phi$ **vol. I.2** © 2023

passes into the realm of what is pure, ever existing, immortal and unchanging, ... *being akin to this*. ... It is in touch then with *things of the same kind*." (79d; emphasis added) (See, further, Yourgrau 1997, 144, footnote 17).

[100] Popper 1962.

[101] *Op. Cit.* 127.

[102] Dawson 1997, 211. See also Wang 1996, 106–7.

[103] Adams 1972.

Gödel to his mother!) It is Dawson, it turns out, not Gödel, who is here being philosophically naïve.[104]

Skepticism of Gödel as philosopher has even, surprisingly, extended to doubts concerning his appreciation of and understanding of Kant, to the point of characterizing Gödel as "pre-critical". Yet, though Gödel most admired Leibniz, he took Kant very seriously and studied him closely. As Wang said: "... [Gödel] felt he needed to take Kant's critique of Leibniz seriously and find a way to meet Kant's objections to rationalism."[105] Indeed, "Gödel often compared his work with Kant's. He knew Kant's work very well, and spoke highly of his ideas ..."[106] Further, Gödel, as mentioned above, came to admire Husserl, and in Gödel 1995, he argued that in a sense Husserl's phenomenology represented the fulfillment of Kant's intentions. Note his closing remark: "But now, if the misunderstood Kant has already led to so much that is interesting in philosophy ... how much more can we expect from Kant understood correctly?"[107] Further, his papers on relativity were inspired by his reflections on Kant: "[my work on rotating universes]," he wrote, "came from my interest in Kant's views. In what was said about Kant and relativity theory, one only saw the difference, nobody saw the agreement of the two. ... Kant said that the ordinary notion [of time] was wrong and that real time is something quite different."[108]

[104] In fact, there is more than a little wisdom, and not infrequently something quite touching, in Gödel's letters to his mother. "What you say," for example, he writes, "about sadness is right. If there were a completely hopeless sadness, there would be nothing beautiful in it. But I believe there can be rationally no such thing. Since we understand neither why this world exists, not why it is constituted exactly as it is, nor why we are in it ...: why then should we presume to know exactly this is to be all ...? (Wang 1996, 43) And this: "You are completely right that mankind does not become better through the moon flight. ... If the progress in history, legal and political science, philosophy, psychology, literature, art, etc. were as great as that in physics, there would not be the danger of atomic war." (Wang 1996,44)

[105] Wang 1996, 164. Indeed, we find Gödel saying, "I have never achieved anything definite on the basis of reading Leibniz. Some theological and philosophical results have just been suggested [by his work]. One example is my ontological proof of the existence of God." (87; brackets added).

[106] Wang 1996, 171.

[107] Gödel 1995, 387.

[108] Wang 1996, 88. In fact, the longer, unpublished version of his philosophical essay on relativity is entitled, "Some Observations About the Relationship Between Theory of Relativity and Kantian Philosophy", some drafts of which have been published in Feferman, et al., eds., 1995. For an argument that Gödel's understanding of Kant, with respect to relativity theory and the ideality of time, was in

$M\times\Phi$ vol. I.2 © 2023

§ — The Big Picture.

In addition to not emphasizing, in detail, the curious and disturbing treatment that has often been accorded Gödel's philosophical reflections, Budiansky, more broadly speaking, never attempts to provide a largescale overview of the mainsprings of Gödel's thinking. Too many trees, too little forest. By contrast, to his credit, Dawson, in his biography, does, if only briefly, provide a big picture. Toward the end of his book , he suggests "four deeply held convictions" of Gödel's": 1) "the univ. is rationally organized and comprehensibly to the human mind"; 2) "it [the universe] is causally deterministic"; 3) "there is a conceptual and mental realm apart from the physical world"; 4) "conceptual understanding is to be sought through introspection."[109] One should add one more theme from Gödel that Dawson draws attention to: 5) Gödel's distinctions between internal and external perspectives, which "are crucial to the incompleteness theorems, in which the truth of the undecidable statements can be seen only metamathematically, and to Gödel's consistency proofs in set theory, which rest of the invariance of certain notions under internal relativization. Long before anyone else, Gödel saw and clearly understood the need for making such distinctions."[110]

Dawson's list, I think, is good, but I would supplement it by adding the following: 6) what I've called (by contrast with the Hilbert program) the Gödel program, the attempt to determine the limits of formal methods in capturing intuitive concepts,[111] something we see evidenced in his completeness theorem, his incompleteness theorems, and his writings on Einstein and relativity. Notably, in all but the first case, the result was negative. Another positive example, however, and, according to Gödel, a very surprising one, of the capturing of an intuitive concept by formal means was Alan Turing's analysis of mechanical computability.

© 2023 $M\times\Phi$ vol. I.2

fact subtle and sophisticated (contrary to what is suggested in Parsons 2010), see Yourgrau 1999, 115–122.

[109] Dawson 1997, 261.

[110] *Op. cit.*, 263–4.

[111] Yourgrau 2005a, 75. Van Atten has used the idea of the Gödel program to help locate Gödel with respect to Brouwer and Hilbert: "A particularly attractive topic is what has been called by Palle Yourgrau (2005, 75) 'the Gödel program'. While Hilbert focused on the formal aspects of mathematics, and Brouwer on intuitive aspects, Gödel focused on the interplay between these two aspects ..." (van Atten 2015, 83.) (Recall, in this regard, Wang's comment about Gödel's dialectic of the intuitive and the formal.)

Finally, I would add 7) the dominant role of conceptual vs. combinatorial thinking in Gödel's investigations. Note, for example, a remark Buldt makes in his review of Dawson: "These statements are not intended to dwarf Gödel's mathematical brilliance. Rather, they are intended to highlight that his genius was in conceptual issues."[112] Budiansky, too, I was glad to see, draws attention to this crucial aspect of Gödel's thinking. He provides a revealing quotation from a notebook of Gödel's: "I am apparently untalented nor interested in combinatorial thinking ... I am apparently talented and interested in conceptual thinking ... Therefore, I should dedicate myself to the foundations of the sciences (and philosophy)."[113] Budianksy, however, fails to expand on this theme, except to make it an occasion for a psychological speculation: "... in a world that would often fill Gödel with anxieties, the realm of pure mathematical logic beckoned with the enfolding reassurance of certainty, so noticeably absent everywhere else, except perhaps in the atavistic feelings of childhood security that he would at times desperately cling to ..."[114]

This was an opportunity lost for Budiansky to have provided a wide-angle view of Gödel as a thinker. He could, for example, have placed alongside that quotation from Gödel 's notebook what the distinguished mathematician Paul Cohen — whose proof, combined with Gödel's, established the independence of the Continuum Hypothesis — said in his essay, "My Interaction with Kurt Gödel: The Man and His Work".[115] Pondering Gödel's incompleteness theorem, he remarked that he once thought to himself that "... the proof was basically *simple*, when stripped to its essentials" (xiv)[116], though he added that "I felt that its discoverer was above me and other mere mortals in his ability to

[112] Buildt 1995, 373.

[113] Dawson 1997, 76.

[114] Ibid. Some will no doubt find such psychological speculations convincing. I'm not one of them. If memory serves, for example, during some of the saddest periods in Mozart's life, he wrote some of his "happiest" music, while in some of his happiest periods, he wrote some of his "saddest". In any case, in the matter of psychological speculations concerning Gödel and mathematical Platonism, Budiansky is not alone. Dawson suggests that Gödel's Platonism may be a sign of "arrested development." (Dawson 1997, 2) Surely, this is going too far.

[115] Which introduces his monograph, Cohen 2008.

[116] Gödel, by the way, agrees: "My work is technically not hard. One can see why my proofs work." (Wang 1996, 85) Interestingly, Feferman seems to disagree, at least In part, writing that "[c]oncerning Gödel's methods, one may say that many of the constructions and arguments were technically difficult for their time, or at any rate, too novel or unexpected to be readily absorbed ..." (Feferman 1986, 33)

$M\times\Phi$ vol. I.2 © 2023

understand what mathematics — and even human thought for that matter — really was." (xiv) He went on to comment that "... logic seemed to me to have little of the element of *combinatorial* thinking — or more crudely put, *cleverness* — that I found in analysis and number theory." (xiv). The contrast, here, between the philosophical logician and the pure mathematician with regard to *conceptual*, or *philosophical*, thinking vs. *combinatorial* thinking is something worth pondering. Including Cohen's remark that, in his youth, "I felt it [Gödel's Incompleteness Theorem] was probably primarily a theorem of *philosophical* interest, having little or nothing to do with the concrete questions of number theory that so infatuated me in those years." (xiii) (Later, however, he had wondered "[h]ow can someone thinking about logic in almost *philosophical* terms discover a result that had implications for Diophantine equations?" (xiv) (all emphasis above, added).

Indeed, it's striking how often Cohen mentions in that essay the crucial role played by philosophical thinking in Gödel's logical/mathematical work, in contrast to Cohen's own attitude to philosophy: "... by temperament, even to the present day, I could not really be involved in *philosophical* controversy." (xiv). And there's still more that's relevant from Cohen. "His [Gödel's] approach in discovering the notion of a constructible set seems to have been decidedly influenced by the *philosophical* discussions of the past ... I found the tone of his article difficult to understand, as it seemed to hover between *philosophy* and mathematics." (x) "His [Gödel's] ideas [about the Continuum Hypothesis] were grounded in a deep *philosophical* belief as to what the human mind could achieve." (xiii) "[T]here seemed no way for me to find an entry into the kind of deep *philosophical* thinking that I so admired in Gödel and translate that into concrete mathematics." (xv) "I read Gödel's article on the continuum problem (Gödel, 1947) ... [T]o a large extent it was not even comprehensible to me. For one thing, it assumed a certain *philosophical* point of view and the *philosophical* ramifications of various attitudes toward it. I had never been attracted to *philosophy* ..." (xv) (all emphases added)

Surely, the strikingly contrasting relationship between Gödel and Cohen in this regard is worth further exploring,[(117)] as is, by

[(117)] Recall the discussion above on Gödel's response to the question of whether to accept Church's Thesis concerning effective computability. As Feferman put it, "Gödel was unconvinced by Church's thesis, since the proposed identification of the effectively computable with the lamba-definable functions did not rest on

© 2023 $M \times \Phi$ vol. I.2

contrast, Gödel's kinship, here, with Frege, whose logical thought also centered around the consideration of (what today we would call) concepts.[(118)] Frege's seminal monograph that gave birth to modern formal logic, after all, was entitled, *Begriffsschrift*, or *Concept Writing* [or *Conceptual Notation*], and the subtitle of his monograph, *The Foundations of Arithmetic* (1980) (the *Grundlagen*), was: *A Logico-Mathematical Enquiry into the Concept of Number*. Notice how many

a *direct conceptual analysis* of the notion of finite algorithmic procedure. [Gödel] always referred to Turing's work as having provided the 'precise and unquestionably adequate definition of formal system' by his 'analysis of the *concept* of 'mechanical procedure' ..." ("Kurt Gödel: Conviction and Caution", Feferman 1998, 162; emphasis added) Again, one sees why Gödel became interested in Husserl's idea of categorial intuition.

In this regard, Feferman poses a good question: "... one may ask why Gödel did not pursue such an analysis *himself*. ... My guess is that he also feared that no such proposal could be made convincing to the mathematical public of his day, just as the concept of truth would not be taken seriously. If so, the subsequent development showed Gödel to have been mistaken. Though certainly there were controversies about both Tarski's analysis of truth and Turing's thesis, they eventually took their place as accepted cornerstones of mathematical logic." (162–3).

Yet precisely one of the controversies still surrounding Tarski's analysis of truth is its relevance to the *intuitive* concept of truth. As for what a reasonable assumption would have been about the reception during the early years of Gödel's work of an account of the concept of truth, see Dawson 1997 (59): "... in discussing Tarski's epochal paper on the concept of truth in formalized languages, Carnap commented that when he invited Tarski to speak on the subject at the September 1935 International Congress for Scientific Philosophy, 'Tarski was very skeptical. He thought that most philosophers ... would be not only indifferent, but hostile to the explication of the concept of truth.' And in fact, 'at the Congress it became clear from the reactions to the papers delivered by Tarski and myself that Tarski's skeptical predictions had been right ... There was vehement opposition even on the side of our philosophical friends.' "

That was then, however. How about now? I have been trying to demonstrate in this Review that with respect to Gödel's entire career, early and late, one should be skeptical of Feferman's claim that "... Gödel's spreading fame should have reassured him that he would not be laughed offstage if he were to go beyond the purely (logico-)mathematical formulation of his results." (Feferman 1998, 163) Similarly, we should think twice about Hao Wang's "mention[ing] a number of times some of the features of Gödel's temperament that led to [his] isolation ... [including] his tendency to overestimate the hostility that his own views would encounter." (Parsons 1998, 19–20; brackets added).

[(118)] Frege refers to functions, which he calls "unsaturated", to contrast them with objects (like numbers and sets). A function whose value is always a truth-value he calls a concept. As pointed out in Feferman 1998, concepts in this sense, are, for Frege extensional, in contrast with what Frege calls "senses", which are intensional. (For an investigation of the complexities of Gödel's conception of concepts, see Crocco 2006 and 2016, and Yourgrau 1997, 162–63, footnote 29).

$M\times\Phi$ vol. I.2 © 2023

times Frege, too, invokes concepts in the *Grundlagen*. The conclusion, so called, of his crucial "Relativity Argument"[(119)], which addresses the question: what do numbers (i.e. the natural numbers, the counting numbers) number?, is that numbers are assigned only to *concepts*, which are the units of counting. As for defining the numbers themselves, he employs the crucial notion of equinumerosity, which relates *concepts* in terms a one-one correspondence between the members of (in effect) their extensions. Finally, according to Frege, the number assigned to the *concept* c is itself defined as: the extension of the *concept*, equinumerous with the *concept* c, i.e. as a set of *concepts*.[(120)]

Returning to Gödel, keep in mind that his explorations into relativistic cosmology were at the service of his attempt to assess the limitations of the "formalism"[(121)] of relativity theory to capture the intuitive concept of time. Similarly, his approach to securing the foundations of mathematics in contrast to the Hilbert program's emphasis on syntax and consistency, focused on the clarification of meaning of *concepts*. Thus Dawson[(122)], quoting Gödel 1995: "the certainty of mathematics is to be secured not by ... manipulation of physical symbols, but rather by cultivating ... knowledge of the abstract concepts themselves." Which points, once again, to Gödel's interest in Husserl and categorial intuition[(123)] and to

$M\times\Phi$ **vol. I.2** © 2023

[(119)] For details see Yourgrau 1997.

[(120)] Ironic, since Frege insists that numbers are objects, and that the fundamental distinction in ontology is between objects and concepts (more generally, functions), yet defined a number as a set of *concepts*, i.e. as an object whose only members are non-objects! By contrast, Russell basically followed Frege's definition, but defined a number as a set of *sets*. (Gödel, it should be noted, agreed with Frege's ontology. "Every being is for Gödel," said Wang, "either a concept or an object ..." [Wang 1987, 309]).

[(121)] "Formalism" in a general, mathematical sense, not the strict sense involved in the notion of a formal system of logic.

[(122)] 1997, 219.

[(123)] Interestingly, Gödel's optimism about Husserl's methodology was countered by Wang's skepticism. "Gödel seems to suggest (as an ideal)," wrote Wang, "that we should aim to use the phenomenological method to discover the axioms for the primitive concepts of philosophy ... But I am not aware of any conspicuous example of definite axioms arrived at in this manner. Neither the axiom of choice, the axiom of replacement, the 'axiom' of constructability, the 'axiom' of determinacy, nor even Dedekind's axioms for arithmetic were obtained by going back to the ultimate acts and contents of our consciousness in the manner recommended by phenomenology." (Wang 1996, 159) No doubt Frege would have been sympathetic with Wang's skepticism. "Do the concepts," he wrote, "as we approach their supposed sources, reveal themselves in peculiar purity? Not at all. .... It is as

the radical difference between the Hilbert program and the Gödel program, as well as to the striking contrast between Gödel's understanding of mathematics and Wittgenstein's. "In mathematics," said Wittgenstein, "*everything* is algorithm and *nothing* is meaning."[124] Though physically their paths crossed only once[125], philosophically, Gödel and Wittgenstein collided. Gödel himself spent scant time on Wittgenstein, but Wittgenstein devoted a great deal of attention to what he saw as the necessary demythologizing of Gödel's incompleteness theorems. I agree with Juliet Floyd when she says that "... one of the most fascinating and extreme philosophical clashes of the century is that between Wittgenstein and Gödel."[126] Given that Wittgenstein is widely considered to be one of the greatest, if not the greatest, philosophers of the twentieth century, Floyd's remark suggests that Gödel, too, should be taken seriously as a philosopher, whereas, as I've been arguing throughout this Review, quite the opposite has in fact been the case.[127]

though everyone who wished to know about America were to try to put himself back in the position of Columbus, at the time when he caught the first dubious glimpse of his supposed India." (Frege 1980, vii–viii) At issue, here, it should be emphasized, is not whether ultimately our confidence in the fundamental concepts and axioms is based on some sort of mathematical intuition (though Frege would no doubt resist using the term, "intuition", here), but rather on what is the best method to help us arrive at these foundational intuitions — indeed, on whether there even is a "philosophical method" to be found here.

[124] Wittgenstein 2005, 468, quoted in Juliet Floyd 2002.

[125] The occasion was a famous lecture that Brouwer gave in Vienna. "I was never introduced to Wittgenstein, and have never spoken a word to him," said Gödel. "I only saw him once in my life when he attended a lecture in Vienna. I think it was Brouwer's." (Wang 1987, 49).

[126] Floyd 2002, 287.

[127] Besides the explicit criticisms of Gödel as philosopher I've drawn attention to, there is also Gödel's failure to engage in debate with most of his philosophical contemporaries. Thus Parsons comments that "I argue elsewhere that Gödel, in his reliance on his notion of concept, does not really have an answer to the deeper Quinean criticism of the ideas about meaning that underlie the analytic-synthetic distinction." (Parsons 1998, 15) Indeed, according to Parsons, there is "a serious weakness in Gödel's engagement with philosophy in the latter part of his career: his largely self-imposed isolation after 1940 from the philosophical community around him ... [B]oth [Gödel and Wang] felt themselves outsiders in the contemporary philosophical world." (19; brackets added) To an extent this is all true, but Parsons neglects to mention Gödel's extensive engagement with the "philosophical community" that did not surround him, including, in particular Husserl. Indeed, until recently, there has been a widespread failure of the philosophical community that surrounded Gödel, including Quine, to engage with the philosophical community that included Husserl. One could add, further, that Gödel's colleague, Church, a fellow conceptual realist, did engage, in detail, famously,

$M\times\Phi$ vol. I.2 © 2023

I'm puzzled, therefore, by Floyd's remark that "... we ought not to defer to Gödel's authority as a philosopher simply in virtue of his greatness as a logician."[128] Who, exactly, is supposed to be deferring to Gödel's greatness as a philosopher simply in virtue of his greatness as a logician?

§ — Who Was Kurt Gödel?

In any case, it's clear that Gödel was a logician with a decidedly conceptual/philosophical orientation. Considered, however, purely formally, as a mathematical logician, how should Gödel be ranked? John von Neumann described him as "the greatest logician since Aristotle." True, Aristotle invented logic, but a better comparison would be to Frege, who developed logic into the mathematical science we know now. Not for nothing is van Heijenoort's seminal collection of original papers on mathematical logic entitled *From Frege to Gödel*.[129] Wang, however, puts it slightly differently: "Nobody can deny that [Gödel] is the greatest logician of the century."[130] Best, however, I believe, would be to describe Gödel, in this regard, simply as "the greatest logician of all time", as I did in Yourgrau 2005a. Stachel, however, considers that description over the top, complaining that "[in Yourgrau 2005a] both Gödel and Einstein are described in over-the-top superlatives".[131] Yet Saul Kripke, whose decision on this question is surely decisive, has described Gödel as "the greatest logician of all time".[132] That is strong praise, indeed. But does it capture who Gödel really was?[133]

$M\times\Phi$ *vol. I.2* © 2023

with Quine. (See Anderson 1998.) Perhaps Gödel believed that he had no need to add to what Church had accomplished in this regard.

[128] *Op. cit.*, 288.

[129] Van Heijenoort 1967. Similarly, best would be to compare Einstein to Newton. Gödel himself said that "the beginning of physics [as we know it] was Newton's work of 1687." (Wang 1996, 167; brackets add).

[130] Wang 1987, 168.

[131] Stachel 2005, 862, brackets added.

[132] Kripke 2013, 95, footnote 3.

[133] Compare: who really was the Canadian pianist, Glenn Gould? The violinist Yehudi Menuhin, himself a major artist, described Gould as "this great phenomenon ... [who] presented a great rift in the continuity of the expected and the ordinary." (McGreevy 1983, 303) In other words, Glenn Gould was not just a gifted pianist. He was a "historical discontinuity". Is not the same true of Kurt Gödel, who was not simply a gifted logician?

Was Gödel, for example, also one of the most significant *philosophers* of the 20th century?[134] Interestingly, in Wang 1987, one finds the following comment: "It is common to view Gödel as a mathematician. ... It is, however, more appropriate to view Gödel as a 'philosopher-scientist' [echoing the title of the famous Schilpp volume on Einstein[135]]."[136] In the years immediately following his book, however, few took up Wang's suggestion. Indeed, in Yourgrau 2005, I commented on a 1995 Boston University conference devoted to Gödel's contributions to philosophy, where prominent speakers were clearly irritated at Gödel's philosophical pretensions, which, the claim was, were taken seriously only because he was a great logician.[137] Apparently, my critique of that attitude succeeded in causing further irritation. In that book, as in Yourgrau 1999, I had been making a case for the idea suggested by Wang of taking Gödel to be, in essence, a philosopher-scientist, focusing on Gödel's philosophy of time — just one aspect of Gödel's wide philosophical interests, though an important one[138], and illustrative about what made Gödel's philosophy special. The winds, however, have finally, begun to shift, if only a little. In 2016 Wang's suggestion finally came to a kind of fruition with the publication of a volume entitled, *Kurt Gödel: Philosopher Scientist*[139], centered around Gödel's newly transcribed and translated Max. Phil.[140] Notebooks. Surely, Budiansky should have drawn attention, in some detail, to the changing climate with regard to Gödel's

[134] Buldt, it should be noted, admires Dawson's austerity in his biography and praises him for precisely *not* doing what I'm recommending here. In Dawson, says Buldt, "there are no sweeping claims about Gödel's greatness or exaggerations about the significance of his results." (Buldt 1995, 387) "Dawson," he continues, "avoids a growing 'Gödelmania' ... in the public." I, too, appreciate Dawson's austerity, his extreme carefulness, his fact-based approach. That is one of the virtues of his book. At the same time, at some point, I believe, it behooves one to try to determine who Gödel really was, what his place is in cultural history. And is not a biography the right place in which to attempt to determine this? Is not one of the main reasons why one reads a biography to find out who someone really is?

[135] Schilpp, ed. 1949.

[136] Wang 1987, 1; brackets added.

[137] I was, frankly, taken aback by this attitude toward Gödel. I wasn't alone. At the conference, I was seated next to the distinguished astrophysicist from the Institute for Advanced Study, Piet Hut, who whispered that he couldn't believe what he was hearing.

[138] Indeed, as mentioned earlier, according to Wang, Gödel believe time to be in a sense *the* philosophical question.

[139] Crocco and Engelen eds. 2016.

[140] *I.e.* Philosophical Maxims.

$M\times\Phi$ vol. I.2 © 2023

reputation as a philosopher-scientist. To be sure, as mentioned, at the end of his book, Budiansky does comment that Gödel's formal and philosophical writings on relativity are beginning to receive "respectful attention", but a) he gives no details about this, and b) he doesn't mention the respectful attention Gödel is beginning to receive for his more extensive philosophical reflections.

As for Gödel's incompleteness theorems, which, like Einstein's theory of relativity, enjoy a life of their own in cultural folklore, Budiansky notes, correctly, that "probably more wrong things have been said about [Gödel's] proof than any other mathematical theorem in history, a dubious distinction — if perhaps an inevitable consequence of popular fame in a field such as mathematics."[(141)] But, once again, he fails to provide details concerning where to look to find a good corrective to the abuse those theorems have suffered. I would strongly recommend, in this regard, the excellent study, written approachably but on a high level, by Torkel Franzén, *Gödel's Theorem: An Incomplete Guide to Its Use and Abuse*.[(142)] Surely, a serious biography such as Budiansky's should serve, among other things, as a guide to how to resist the abuse to which Gödel's theorems have been subjected, as well as how to find an illuminating and rigorous presentation of the theorems themselves.

In conclusion, I've said a lot here about what Budiansky might have done, or should have done, in his new biography of Gödel, especially in light of what Dawson has already accomplished in his now classic study, but it remains that what Budiansky has actually done is significant and commendable. He has produced an engaging historical narrative that will no doubt grip the imagination of the general reader and, at the same time, provide new insights into the private and professional life of Kurt Gödel from which both experts and neophytes will profit. And to repeat what I said at the beginning: the book is clearly written "in a good will", and this is no small virtue.[(143)]

[(141)] Budiansky 2021, 278. Indeed, notes Budiansky, "like Heisenberg's Uncertainty Principle and Einstein's Theory of Relativity, Gödel's Incompleteness Theorem has provided what Alan Sokal and Jean Bricmont ... called 'an inexhaustible source of intellectual abuse' ..." (278) In fact, those three epoch-making discoveries have something more fundamental in common. "Einstein, Gödel, Heisenberg [were] three men whose fundamental scientific results opened up new horizons, paradoxically, by setting limits to thought or reality. ... Mysteriously, each established an ontological conclusion about *reality* through the employment of an epistemic principle concerning *knowledge*." (Yourgrau 2005a, 3; brackets added).

[(142)] Franzén 2005.

[(143)] I thank Mark van Atten for his helpful comments on this Review.

$M\times\Phi$ vol. I.2 © 2023

§ — References.

Adams, R. "Must God Create the Best?", *The Philosophical Review*, Vol. 81, No. 3, July, 1972, 317–32.

Ahmed, A. "Saul Kripke: The Return to Metaphysics", *Times Literary Supplement*

Anderson, C. A. "Alonzo Church's Contributions to Philosophy and Intensional Logic", *The Bulletin of Symbolic Logic*, Vo. 4, Number 2, June 1998, 129–171.

Aristotle, *De Anima* (*On the Soul*), in J.L. Ackrill, ed., *A New Aristotle Reader*, Princeton, New Jersey, Princeton University Press, 1987, 161–205.

— *Nicomachean Ethics*, in J.L. Ackrill, ed., *A New Aristotle Reader*, Princeton, New Jersey, Princeton University Press, 1987, 363–478.

Audureau, E. "Gödel: From the Pure Theory of Gravitation to Newton's Absolute," in G. Crocco and E. Engelen, eds., 2016, 57–79.

Van Atten, M. *On Brouwe*r, Thomson, Wadsworth, 2004.

— *Essays on Gödel's Reception of Leibniz, Husserl, and Brouwer*, Cham, Springer, 2015.

— "Gödel, Mathematics, and Possible Worlds, in M. van Atten, 2015, 147–155.

— "Predicativity and Parametric Polymorphism of Brouwerian Implication", `https://arxiv.org/pdf/1710.07704.pdf`, 2018, 1- 41.

— "Luitzen Egbertus Jan Brouwer", in E. N. Zalta, ed., *Stanford Encyclopedia of Philosophy,* Spring 2020.

Van Atten, M. and Kennedy, J. "On the Philosophical Development of Kurt Gödel", *The Bulletin of Symbolic Logic*, Vol. 9, Number 4, Dec. 2003, 425–476.

Bonnor, W. *The Mystery of the Expanding Universe: The Humorous, Caustic and Illuminating Story of What Men Think About Their Exploding World*, MacMillan, New York, N.Y., 1964.

Budiansky, S. *Journey to the Edge of Reason: The Life of Kurt Gödel*, New York, N. Y., W. W. Norton & Company, 2021.

$M\times\Phi$ vol. I.2 © 2023

Buldt, B. "Stories of Genius: Recent Accounts of Kurt Gödel's Life", *Europolis* 5. *Kurt Gödel und Die mathematische Logik*, Linz: Linzer Universtätsverlage, 1995, 358–391.

Carnap, R. "Intellectual Autobiography", in P.A. Schilpp, ed., *The Philosophy of Rudolf Carnap*, Cambridge, Cambridge University Press, 1963.

Chandrasakar, S. and Wright, J. P. "The Geodesics in Gödel's Universe," *Proceedings of the National Academy of Sciences*, U.S.A., **47**: 241–47.

Cherniss, H. *The Riddle of the Early Academy*, New York, Russell and Russell, 1962.

Chihara, C. *Ontology and the Vicious Circle Principle*, Ithaca, New York, Cornell University Press, 1973.

Church, A. "The Need for Abstract Entities in Semantic Analysis", *American Academy of Arts and Sciences Proceedings*, vol. 80, 1951, 110–133.

Cohen, P. *Set Theory and the Continuum Hypothesis*, Mineola, New York, N.Y., Dover Publications, Inc., 2008.

Crocco, G. "Gödel on Concepts", *History and Philosophy of Logic*, Vol. 27, 2006.

— "*Sinn/Bedeutung* and *Intension/Extension* in Gödel's Max Phil IX", 127–152, in G. Crocco and E. Engelen, eds., 2016.

Crocco, G. and Engelen, E. eds. *Kurt Gödel: Philosopher-Scientist*, Presses Universitaires de Provence, Aix-en-Provence, 2016.

Van Dalen, D. *Mystic, Geometer, and Intuitionist: The Life of L.E.J. Brouwer*, Vol. 2: *Hope and Disillusion*, Oxford, Clarendon Press, 2005.

Davis, M. "Why Gödel Didn't Have Church's Thesis", *Information and Control*, **54**, 1982, 3–24.

— "Review: John W. Dawson, Jr. *Logical Dilemmas: The Life and Work of Kurt Gödel*", Philosophia Mathematica, (3) Vol. 6, 1998, 116–128.

— "Gödel's Universe: Review of R. Goldstein, *The Proof and Paradox of Kurt Gödel*, and P. Yourgrau, *A World Without Time: The Forgotten Legacy of Gödel and Einstein*," *Nature*, Vol. 435, 2005, 19–20.

$M \times \Phi$ *vol. I.2* © 2023

Dawson, J. *Logical Dilemmas*: *The Life and Work of Kurt Gödel*, A.K. Peters/CRS Press, 1997.

— "What Have We Learned from the Gödel *Nachlass*, and What More May it Have to Offer?", in G. Crocco and E. Engelen, eds., 2016, 15–31.

Dorato, M. "On Becoming, Cosmic Time, and Rotating Universes", in C. Callender, ed., *Time, Reality and Experience*, Cambridge, Cambridge University Press, 2002, 253–76.

Dummett, M. *Frege: Philosophy of Mathematics*, Cambridge, MA, Harvard University Press, 1991.

Earman, J. *Bangs, Crunches, Whimpers and Shrieks: Singularities and Acausalities in Relativistic Spacetimes*, New York, N.Y., Oxford University Press, 1995.

Edmonds, D. *The Murder of Professor Schlick: The Rise and Fall of the Vienna Circle*, Princeton, Princeton University Press, 2020.

Ellis., G.F.R. "Contributions of K. Gödel to Relativity and Cosmology", in P. Hajek, ed. *Gödel '96: Logical Foundations of Mathematics, Computer Science and Physics — Kurt Gödel's Legacy*, Berlin, Springer-Verlag, 1996, 34–39.

Engelen, E. "What is the Link Between Aristotle's Philosophy of Mind, the Iterative Conception of Set, Gödel's Incompleteness Theorems and God?", in G. Crocco, and E. Engelen, eds., 2016, 171–188.

Feferman, S. "Gödel's Life and Work", in S. Feferman et al., eds., 1986, 1–36.

— "Kurt Gödel: Conviction and Caution", in Feferman, *In the Light of Logic*, New York, Oxford, University Press, 1998.

— "Provably Unprovable: Review of R. Goldstein, *Incompleteness: The Proof and Paradox of Kurt Gödel*", *London Review of Books*, Vol. 28, No. 3, February 2006.

Feferman, S. *et al.* , eds. *The Collected Works of Kurt Gödel*, Vol. I, New York, N.Y., Oxford University Press, 1986.

— *The Collected Works of Kurt Gödel*, Vol. II, New York, N.Y., Oxford University Press, 1990.

— *The Collected Works of Kurt Gödel*, Vol. III, New York, N.Y., Oxford University Press, 1995.

$M\times\Phi$ vol. I.2 © 2023

Floyd, J. "Review of Mathieu Marion, *Wittgenstein, Finitism, and the Foundations of Mathematics*, (Oxford, Clarendon Press, 1998)," *Philosophia Mathematica* (3), Vol. 10, 2002, 67–88.

Fraenkel, A. *Recollections of a Jewish Mathematician in Germany*, Birkhäuser, 2016.

Frege, G. *Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought*, in J. van Heijenoort, ed., 1967, 1–82.

— *The Foundations of Arithmetic: A Logico-Mathematical Enquiry into the Concept of Number*, Transl. J.L. Austin, Evanston, Illinois, Northwestern University Press, 1980.

Franzén, T. *Gödel's Theorem: An Incomplete Guide to Its Use and Abuse*, Boca Raton, CRC Press, Taylor & Francis Group, 2005.

Gödel, K. "The Present Situation in the Foundations of Mathematics", an address delivered to the Mathematical Association of America, Dec. 1933.

— "What Is Cantor's Continuum Problem?", in S. Feferman et al. eds., 1990, 254–270.

— "The Modern Development of the Foundations of Mathematics in the Light of Philosophy", in S. Feferman et al. eds., 1995, 375–387.

Goldstein, R. *Incompleteness: The Proof and Paradox of Kurt Gödel*, New York, N. Y., W. W. Norton & Company, 2005.

Hegel, G. *Introduction to the Philosophy of History*, Indianapolis, Hackett Publishing Company, 1988.

Van Heijenoort, J., ed. *From Frege to Gödel: A Source Book in Mathematical Logic*, 1879–1931, Cambridge, MA, Harvard University Press, 1967.

Isaacson, W. *Einstein: His Life and Universe*, New York, Simon & Schuster, 2007.

Janik, A. and Toulmin, S. *Wittgenstein's Vienna,* Simon & Schuster, 1973.

Kripke, S. "Frege's Theory of Sense and Reference: Some Exegetical Notes", in S. Kripke, *Philosophical Troubles: Collected Papers*, Vol. 1, Oxford, Oxford University Press, 2011, 254–91.

— "The Church-Turing Thesis as a Special Corollary of Gödel's Completeness Theorem", in B. J. Copeland et al., eds., *Computability: Turing, Gödel, Church and Beyond*, Cambridge, MA, The MIT Press, 2013, 77–104.

© 2023 $M \times \Phi$ vol. I.2

Malament, D. "'Time-Travel' in the Gödel Universe", *Proceedings of the Philosophy of Science Association* 2, 1984, 91–100.

Mar, G. "Hao Wang's Logical Journey", *Journal of Chinese Philosophy*, 42:S1, Dec. 2015, 540 – 561.

Martin, D. "Gödel's Conceptual Realism", *The Bulletin of Symbolic Logic*, Vol.11, Number 2, June 2005, 207–224.

McGreevy, J. *Glenn Gould: Variations*, Toronto, Ontario, Doubleday Canada Limited, 1983.

Menger, K. *Reminiscences of the Vienna Circle and the Mathematical Colloquium*, Springer, 2013.

Moore, G. 1990. "Kurt Gödel", in L.H. Adams and F.H. Laves eds., *Dictionary of Scientific Biography*, Vol. 17, Supplement II, New York, Charles Scribner's Sons, 348–357.

Moravcsik, J. "Frege and Chomsky on Thought and Language", in P. French, et al., eds., *Midwest Studies in Philosophy* VI, 1981: *The Foundations of Analytic Philosophy*, Minneapolis, University of Minnesota Press, 1981, 105–123.

Parsons, C. "Hao Wang as Philosopher and Interpreter of Gödel", *Philosophia Mathematica*, (3) Vol. 6, 1998, 3–24.

— "Gödel and Philosophical Idealism", *Philosophia Mathematica* (III) **18**, 2010, 166 – 192.

Plato, *The Phaedo*, in *Plato: Five Dialogues*, transl. G.M.A. Grube, Indianapolis, Hackett Publishing Company, 1981, 93–155.

Popper, K. "On the Theory of the Objective Mind", *Akten des XIV. Internationalen Kongresses Für Philosophie*, vol. I, Vienna, 1968.

Schilpp, P. A., ed,. *Albert Einstein: Philosopher-Scientist*, "The Library of Living Philosophers", La Salle, Illinois, Open Court, 1949.

Schorske, C. *Fin-de-Siécle Vienna : Politics and Culture*, Vintage, 1980.

Sigmund, K. *Exact Thinking in Demented Times: The Vienna Circle and the Epic Quest for the Foundations of Science*, New York, N.Y., Basic Books, 2017.

$M\times\Phi$ vol. I.2 © 2023

Stachel, J. "Review: *A World Without Time: The Forgotten Legacy of Gödel and Einstein*," *Notices of the American Mathematical Society*, 2007, 861–868.

Stein, H. "On the Paradoxical Time-Structures of Kurt Gödel", *Philosophy of Science* 37: 1970, 589–601.

Tait, W. W. "Truth and Proof: The Platonism of Mathematics", *Synthese* **69**, 1986, 341 -70.

— "Critical Study/Book Review: Gödel's Unpublished Papers on Foundations of Mathematics", *Philosophia Mathematica* (3) Vol. 9, 2001, 87–126.

Tragesser, R. *Phenomenology and Logic*, Ithaca, Cornell University Press, 1977.

Turing, A. "On Computable Numbers, with an Application to the *Entscheidungsproblem*", in M. Davis, ed., *The Undecidable: Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*, New York, Raven Press, 1965, 230–265.

Wang, H. *Reflections on Kurt Gödel*, Cambridge, MA, A Bradford Book, The MIT Press, 1987.

— *A Logical Journey: From Gödel to Philosophy*, Cambridge, MA, The MIT Press, 1996.

Wittgenstein, L. *Philosophical Investigations*, Pearson, 1973.

— *Philosophical Remarks*, Chicago, The University of Chicago Press, 1975.

— *Philosophical Grammar*, University of California Press, 2005.

Yourgrau, P. *The Disappearance of Time: Kurt Gödel and the Idealistic Tradition in Philosophy*, Cambridge, Cambridge University Press, 1991

— "What is Frege's Relativity Argument?", *The Canadian Journal of Philosophy*, Vol. 27, Number 2, June 1997, 137–171.

— *Gödel Meets Einstein: Time Travel in the Gödel Universe*, 1999, Chicago and La Salle Illinois, Open Court, 1999.

— *A World Without Time: The Forgotten Legacy of Gödel and Einstein*, New York, N.Y., Basic Books, 2005a.

$M \times \Phi$ vol. I.2 © 2023

— "Gödel, Kurt", in P. Borchert, ed., *The Encyclopedia of Philosophy: Supplement*, 2nd edition, Macmillan/Collier, 2005b, 117–119.

— "Kripke's Frege", *Thought* **1**, 2012, 100–107.

— "Proof vs Provability: On Brouwer's Time Problem," *History and Philosophy of Logic*, 2019, 1–14.

© 2023 $M\times\Phi$ *vol. I.2*

Palle Yourgrau, Department of Philosophy,
Brandeis University.

⋆
⋆ ⋆

N° d'édition : 413597845-1
Dépôt Légal : décembre 2023
Achevé d'imprimer : décembre 2023

Imprimé en France par

AQUIPRINT

33520 Bruges
www.impression-edition-gironde.com