Logic Discovered and Logic Imposed
(a Purim story)

CURTIS FRANKS

1. Bridges to nowhere

During the second Persian invasion of Greece, Xerxes organized the construction of pontoon bridges spanning the Hellespont to provide his foot soldiers a shortcut into Thrace. Herodotus described the event as a significant feat, involving state of the art craftmanship with Phoenician flax and Egyptian papyrus, traversing incredible lengths. But before the Persian army could cross, "a great storm swept down, breaking and scattering everything." Herodotus wrote:

> When Xerxes heard of this, he was very angry and commanded that the Hellespont be whipped with three hundred lashes, and a pair of fetters be thrown into the sea. I have even heard that he sent branders with them to brand the Hellespont. (*Godley 1920*, Book VII.33.1)

This image of Xerxes's troops whipping and "shackling" the sea became for Jean-Yves Girard a metaphor of a farcical, but prevailing, conception of logic. In "Between logic and quantic: a tract" he wrote: "To some extent, this is what logicians wanted to do to quantum physics, to punish it for being 'against common-sense.' . . . our beautiful minds were harboured by the wrong world and this was a mere accident" (§1.1.1). In *The Blind Spot*, he elaborated:

> Among the magisterial mistakes of logic, one will first mention quantum logic, whose ridiculousness can only be ascribed to a feeling of superiority of the language—and ideas, even bad, as soon as they take a written form—over the physical world. Quantum logic is indeed a sort of punishment inflicted on nature, guilty of not yielding to the prejudices of logicians . . . just like Xerxes had the Hellespont—which had destroyed a boat bridge—whipped. (p. xii)

Xerxes's bridges to nowhere raise a question: If it is wrong-headed to expect nature to conform to our preconceptions, if we should be sanguine even in the face of observed violations of the laws of logic, then what, exactly, is the status of those laws? In Wittgenstein's memorable phrase, "This seems to abolish logic" (*1953*, §242). If logic is beholden to the contingencies of the world as Girard invited us to think and as Wittgenstein's own critical investigation seems to suggest, "Its rigor seems to be giving way here.—But in that case doesn't logic altogether disappear?—For how can it lose its rigor?" (§108).

It is evident that with this question Wittgenstein sets up one of the central agendas of his later philosophy. In the first place, the preconception that he wanted to expose as fraudulent is one that he associated with his earlier view in the *Tractatus*:

> Thought is surrounded by a halo.—Its essence, logic, presents an order, in fact the *a priori* order of the world: that is, the order of possibilities, which must be common to both world and thought. But this order, it seems, must be utterly simple. It is prior to all experience, must run through all experience; no empirical cloudiness or uncertainty can be allowed to affect it.—It must rather be of the purest crystal. But this crystal does not appear as an abstraction; but as something concrete, indeed, as the most concrete—as it were the hardest thing there is (*Tractatus Logico-Philosophicus* No. 5.5563). (*1953*, §97)

Understanding how Wittgenstein meant to move beyond his earlier view will then depend on understanding the force of his reply to this Tractarian image, how "On the one hand it is clear that every sentence in our language is in order as it is. That is to say, we are not striving after an ideal, as if our ordinary vague sentences had not yet got a quite unexceptionable sense, and a perfect language awaited construction by us" (§98).

Additionally, Wittgenstein's famous mantra about the relationship of his investigations to traditional philosophical inquiry is expressed just in terms of his abandonment of this preconception about logic: "The preconceived idea of crystalline purity can only be removed by turning our whole examination around" (§108). So if we want to understand Wittgenstein's departure from his earlier account of the nature of logic, and if we want to understand how his later method is a reversal of traditional philosophy, we need to understand what logic is supposed to be for Wittgenstein instead of "the *a priori* order of the world." What is the alternative to trying to shackle and whip the sea?

The key to answering this question, I think, can be found in another of Wittgenstein's remarks. "The more narrowly we examine actual language," he wrote, "the sharper becomes the conflict between it and our requirement. (For the crystalline purity of logic was, of course, not a result of investigation: it was a requirement.)" It is just this "conflict" between actual language and "our requirement" that seems to threaten the very idea of anything deserving of the name "logic": "The conflict becomes intolerable," he continued; "the requirement is now in danger of becoming empty" (§107). So "turning our whole investigation around" is going to involve reconceiving logic in some way so as to avoid this conflict.

I think Wittgenstein is commonly understood to be asking us to drop the image of logic as a preconception altogether. Require nothing; read logic out of an empirical investigation of the world. Whatever nature has to offer in terms of its most general features just are the laws of logic. They don't differ from other laws of nature in kind, but in degree: They are more general than the laws of mechanics but not necessities of a higher order. But, according to this way of understanding Wittgenstein, if we fasten onto the image of logic as a halo surrounding thought, as *a priori* and a feature, not of what the world is like in its most general features, but of what the world, any world, *must* be like, then we will both be disappointed by nature's failure to satisfy our demands and blind to the logical contours that nature offers as a matter of fact. Logic, he said, should be "the result of an investigation," rather than "a requirement" that we bring to the table in all our investigations.[1]

But Wittgenstein doesn't seem to me to be saying this for two reasons.

Notice, first, that Wittgenstein doesn't say that logic should be the result of an investigation, just that it's wrong to think that it is when in fact it isn't. He seems to be content just to remind us that logic is something we are requiring, that its principles are our stipulations and our expectations. If this is all he is saying, then, far from saying that logic shouldn't be something we impose, Wittgenstein is warning against forgetting that it is. Then the "conflict" between logic and the empirical investigation of the world or actual language is "intolerable" only because we so often lose sight of logic's conventional status.

Perhaps more significantly, Wittgenstein doesn't say that "logic" itself, the laws of logic or results of logical research or anything like that, is the thing that is a requirement rather than a discovery. He says that logic's "crystalline purity" is. "In what sense is logic something sublime?" he asks at §89. Maybe it is just fine to stipulate logical principles up front, just so long as you don't pretend that their necessity is somehow discovered rather than imposed. Necessity discovered as a feature that the world could not but exhibit or a precondition of our ability to think about the word would be other-worldly. But if necessity is is nothing more than the fact that we impose certain laws, this sublimity disappears. The "requirement" that an honest appraisal of the conflict between the logic that we bring to bear on the world and the world itself exposes as "empty" is not logic but its alleged sublimity.

Of course, pinning any interpretation on Wittgenstein's gnomic remarks is a task one should be wary of trying to carry out convincingly. I myself am overburdening his words by saying *both* that flagging logic as a requirement is not a call to start recognizing it instead as a discovery *and* by pointing out that the requirement he thinks will turn up empty under scrutiny is not logic *per se* but a certain conception of it. I wouldn't expect a textual exegesis either to verify or to refute this reading and so don't plan to offer one. What I think is a better approach, both to clarifying my interpretation and to indicating its aptness, is to follow another lead from Wittgenstein: "Don't think, but look" (§66). I will illustrate logic as discovery and logic as requirement with two historical episodes. I think this illustration can facilitate an appreciation of how logic has functioned well, precisely when its practitioners have resisted neither the place of requirement nor the place of discovery in its development.

[1] Penelope Maddy's *2012* "second-philosophy of logic" is a forceful development of this interpretation of Wittgenstein. See also *Maddy 2014*.

I offer the image that emerges as an alternative to the dual extremes of whipping the sea in the name of the absolute and submitting to its tide in the name of empiricism. Here my lead is Mark Steiner's own example. Mark is well known not only to have pursued an interest in Jewish philosophy but to have woven that interest seamlessly into his interpretations of Wittgenstein and his study of logic and mathematics. As it happens, Xerxes's narrative is already subsumed into traditional Jewish writing. My conclusion that in logic's actual history one can find a more dramatic, if subtler, depiction of what it might mean to flip the script on Xerxes leans on a reading of *Megillat Esther*. And as I hope Mark would appreciate, this reading provides the key that disambiguates Wittgenstein's own remarks about logic.

2. Frege on deduction

Gottlob Frege is celebrated among philosophers for his groundbreaking work on quantification theory and deduction and for his conceptually motivated development of formal arithmetic. He is possibly even better situated in infamy, though. Most notoriously, his latter achievement is marred by its accommodation of a paradoxical construction that renders his whole system, and with it his very philosophy of mathematics, irreparably inconsistent. His former achievement exhibits a less dramatic but similarly problematic oddity, preventing Frege from appreciating important advances in the modernization of mathematics in his own time. This second Fregean quirk sets the stage for our first exhibit of logic as requirement and logic as discovery. But it is rarely thoroughly understood, and not uncommonly misunderstood. So before reading from it any moral about the conceptualization of logic, we will review the alleged problem itself.

The modern development in mathematics that Frege so notoriously failed to appreciate was the evaluation of systems of axioms for consistency. In the modern scheme, gradually introduced by algebraists and geometers in the 19th Century but made famous and rigorous primarily by David Hilbert, one demonstrates the consistency of a collection of axioms by reinterpreting them so that they can be read as simultaneously true, and one demonstrates the inconsistency of a collection of axioms by formally deriving from them a contradiction. Frege repeatedly and forcefully declared that these activities are incoherent. As for Hilbert's consistency proofs, Frege claimed that by reinterpreting a sentence you are left with a different sentence, so that you learn nothing in the process about the sentences you began with (*Frege 1980*, pp. 39–40). Turning to formal proofs of inconsistency, Frege maintained that because "When we infer, we recognize a truth on the basis of other previously recognized truths according to a logical law" it could never happen that one could formally infer a contradiction (*Frege 1917*). If you are inferring at all, you are necessarily working with true axioms, and truths do not conflict with one another.

Frege's attitude about logical inference is peculiar from our modern point of view, and it might not be possible for us to know today that we've understood him correctly. Contemporary readers have proposed various conflicting reconstructions of Frege's thought on this topic, typically with an eye towards minimizing the apparent gap between his views and our prevailing modern attitude. I survey some of those interpretations in *Franks 2018* but conclude that Frege is as foreign to our way of thinking as he sounds. What I think should be uncontroversial is that Frege considered the validity of an inference to depend *both* on the nature of the rule according to which one infers *and* on the nature of the premises, if any, from which one is inferring. The rule has to be what we would call a deductively valid pattern of reasoning, and the premises have to be logical truths in their own right.

While we are inclined to accept Frege's first constraint—if the rule is not deductively valid, one is not inferring but just pattern-matching—the second constraint is unfamiliar. But it figures centrally in Frege's thought. Not only is the violation of this second constraint the culprit in Frege's rejection of formal consistency proofs, it is in a way the more basic of the two. For Frege's verification of the deductive validity of his rules of inference is not (and could not be) conducted in terms of their ability to "preserve truth" from arbitrary premises but instead in terms of their reliability to lead from actual truths to other actual truths in his logical system. For Frege, what makes a rule of inference valid is the fact that the set of logical truths is closed under its application.

In modern terminology, Frege's concept of deductive validity is that of "admissibility": the closure of the set of theorems under the operation of a rule. This is opposed to the stricter concept of "derivability"—the inability to simultaneously interpret a rule's premises as truths and its conclusion as a falsity—that today is the more common way of explicating the idea of a rule's validity. What we have noted is that Frege doesn't just prefer the formulation of validity in terms of admissibility. It is forced on him. The concept of derivability, involving reinterpretation of the language of his system, is simply incoherent from his perspective.

Against this conceptual background, Frege says something truly unexpected. One of the primitive symbols in his systematization of logic stands for the binary conditional operator →, allowing the presentation of sentences of the form $\mathrm{A} \rightarrow \mathrm{B}$ with subsentences A and B. The conditional symbol features in all of his axioms, as well as in his sole inference rule, *modus ponens* (from A and $\mathrm{A} \rightarrow \mathrm{B}$, infer B), but of course there are other logical truths of the form $\mathrm{A} \rightarrow \mathrm{B}$ that Frege would like to verify, and he indicates two ways of doing this. First, as one expects, Frege describes using the inference rule *modus ponens* to build from the axioms of his system a formal derivation that concludes with $\mathrm{A} \rightarrow \mathrm{B}$. Let us follow Frege in denoting the existence of such a derivation by $\vdash_{\mathfrak{B}} \mathrm{A} \rightarrow \mathrm{B}$.[2] But also, surprisingly, Frege describes building a formal derivation from the axioms of his system *together with* the arbitrary sentence A, again

[2] The gothic $\mathfrak{B}$ index denotes the formal system of derivation from Frege's *Begriffsschrift*.

using *modus ponens* but now in a generalized way so that sentences derived from A and even A itself can figure into the premise position of its application—a derivation that ends not with $A \rightarrow B$ but with just B. (*Frege 1910*: "I can, indeed, investigate what consequences result from the supposition that A is true without having recognized the truth of A; but the result will then contain the condition 'if A is true.' "[3] A contemporary generalization of Frege's notation would denote the existence of such a formal derivation by $A \vdash_{\mathfrak{B}} B$. Frege's surprising claim is that merely observing $\Gamma, A \vdash_{\mathfrak{B}} B$ suffices to verify $\Gamma \vdash_{\mathfrak{B}} A \rightarrow B$. It is fair to pose several questions about Frege's entitlement to this claim.

First: How is Frege suddenly deriving things from a sentence that hasn't been judged to be true. Didn't Frege declare such reasoning incoherent?

Second: *Modus ponens* was validated in terms of its reliability over the space of true Fregean judgements. How does Frege know that it is reliable also when applied to arbitrary sentences? Frege here seems to be assuming that the admissibility of an inference rule entails its derivability. But as we pointed out, derivability is a stricter notion. Consider the rule of propositional substitution: Any substitution instance of a theorem is again a theorem, so this rule is admissible. But it certainly is not derivable, for it leads from satisfiable sentences like $A \wedge B$ to unsatisfiable sentences like $A \wedge \neg A$. It seems that for all Frege has committed to, *modus ponens* could similarly be unreliable when reasoning from arbitrary sentences, allowing fallacious derivations of the form $A \vdash_{\mathfrak{B}} B$. One would not want to infer $\vdash_{\mathfrak{B}} A \rightarrow B$ from that.

Third: Frege's claim is a version of the "deduction theorem." This is a metatheorem about the existence of a concrete object: a derivation in Frege's system of the sentence $A \rightarrow B$. The first published verification of this result is in *Herbrand 1930*, and although Tarski (*1956*, p. 32) claimed to have arrived at the proof earlier, even he dates his discovery at 1921. How can Frege be confident, in *1910* and *1917*, that this metatheorem is true without offering any proof?

Applying the dilemma "logic discovered vs. logic imposed" to this curiosity in Frege's thought can be illuminating. Consider the deduction theorem, the claim that $\Gamma, A \vdash_{\mathfrak{B}} B$ only if $\Gamma \vdash_{\mathfrak{B}} A \rightarrow B$. In the formalist conception of axiomatic systems, in which axioms can be shown to be consistent through reinterpretation, and according to which inference rules are just formal patterns of reasoning whose sole burden is the preservation of truth from premise to conclusion under every possible reinterpretation, this claim certainly stands in need of proof. Herbrand's 1930 verification therefore stands as a discovery, the discovery that derivations witnessing the claim $\Gamma, A \vdash_{\mathfrak{B}} B$ guarantee the existence of proofs witnessing the claim $\Gamma \vdash_{\mathfrak{B}}$

[3] A most remarkable example of Frege reasoning in this way occurs in the same *1917* letter to Hugo Dingler in which he declared the whole enterprise of inferring from false or uncertain premises incoherent. Frege claimed there that from the thought that 2 is less than 1 and the thought that if something is less than 1 then it is greater than 2, one can derive that 2 is greater than 2.

$\mathrm{A} \to \mathrm{B}$. Because derivations of the first sort are typically easier to generate than those of the second sort, this discovery is of practical interest as well.

But its status as a discovery depends on this particular conception. The conception itself is deceptive, for it deliberately ignores the fact that neither the axioms nor the inference rules of the system in question are arbitrary. It operates on a pretense that the system wasn't carefully designed so that the conditional symbol should be coordinated with the derivation turnstile in just the way the deduction theorem claims.

To see this, consider a contrasting ("inferentialist") conception of logic, according to which *modus ponens* is not justified in terms of its reliability when reasoning about, among other things, sentences involving the conditional operator but, instead and in exactly the opposite manner, the conditional operator is defined in terms of the *modus ponens* inference. On this conception, to know $\mathrm{A} \to \mathrm{B}$ is to know that further knowledge of A suffices for concluding B: $\mathrm{A} \to \mathrm{B}, \mathrm{A} \vdash_{\mathfrak{B}} \mathrm{B}$. But if *modus ponens* is supposed, in this way, to be not just an inference pattern that happens to be valid, but the very definition of the conditional, something more follows. Suppose, for example, that knowledge of C is such that further knowledge of A suffices for concluding B, that $\mathrm{C}, \mathrm{A} \vdash_{\mathfrak{B}} \mathrm{B}$. That is to say that C can occupy the same role that $\mathrm{A} \to \mathrm{B}$ occupies by definition. Then one can conclude just from one's knowledge of C that $\mathrm{A} \to \mathrm{B}$ is true, i.e. $\mathrm{C} \vdash_{\mathfrak{B}} \mathrm{A} \to \mathrm{B}$. But this is just the deduction theorem itself.

In this way the deduction theorem is an immediate consequence of the understanding of *modus ponens* as the definition of the conditional. One might be reminded here of the setting of natural deduction, where *modus ponens* is relabeled as *→-Elimination* and the deduction theorem is relabeled as *→-Introduction*, as well as Gerhardt Gentzen's famous but cryptic slogan (from *Gentzen 1934–35*) about introduction rules being consequences of their corresponding elimination rules.[4] Gentzen's comment has been hard for many readers to appreciate precisely because, on a formalist conception of logic, there is no sense in which any one of the rules of his natural deduction system follows from the others: the rules exhibit formal independence. But on the alternative conception of logic described here it is clear how *→-Introduction* is an immediate consequence, not literally of *→-Elimination* as a formal rule of inference, but of *→-Elimination* conceived of as a definition of →. A capsule statement of this observation is that on our alternative inferentialist conception of logic, the deduction theorem is not something to discover but a requirement built in to the very definition of the conditional.

Returning to Frege's seemingly cavalier embrace of the deduction theorem, does the foregoing analysis suggest that he shared with Gentzen an inferentialist conception of logic? I suggest, more modestly, that Frege's thought is ambiguous between the formalist and infer-

[4] Or, more accurately, just the reverse. Gentzen described elimination rules as consequences of introduction rules: "The introductions represent . . . the 'definitions' of the symbols concerned, and the eliminations are no more, in the final analysis, than the consequences of these definitions" (§II 5.13).

entialist conceptions just described. Pinning either on him is likely anachronistic. He never claims, for instance, that *modus ponens* defines the conditional operator, nor that the deduction theorem is an immediate consequence of the *modus ponens* rule so conceived. What he did is *treat* the deduction theorem as a consequence of *modus ponens*, indicating that he felt no need to formally verify it either. What Herbrand would later show is that *modus ponens* and Frege's first two axioms are all the ingredients one needs for a formal verification of the deduction theorem. What Gentzen would show even later is that *modus ponens*, reconceived as a definition of the conditional operator, already contains in it the deduction theorem and, in exactly the opposite manner of Herbrand, that these two facts are all the ingredients one needs for a formal verification of Frege's first two axioms.[5] Frege seemed to be operating in the ambient space where this mutually supporting network of axiom, inference rule, and metatheorem did not invite clear demarcations between those principles to conceive of as requirements and those to conceive of as discoveries.

From this position, Frege's several seemingly incompatible remarks can be reconciled. In reverse order: His confidence that a direct proof of $\Gamma \vdash_{\mathfrak{B}} \mathrm{A} \rightarrow \mathrm{B}$ exists whenever a derivation witnessing $\Gamma, \mathrm{A} \vdash_{\mathfrak{B}} \mathrm{B}$ can be found is perhaps too bold. But it is not a separate matter from his well-known confidence that his system is complete in the sense that it suffices to derive all valid formulas: Frege saw that whenever $\mathrm{A} \vdash_{\mathfrak{B}} \mathrm{B}$, based just on the meaning that *modus ponens* lends to the conditional operator, $\mathrm{A} \rightarrow \mathrm{B}$ is certainly valid as well. Therefore an adequate set of axioms is sure to supply its proof.

His application of *modus ponens* to arbitrary formulas, despite having defended its reliability only over the space of logical truths, is not problematic because he wasn't committed to its deductive validity in such circumstances: He only ever applied *modus ponens* in this way in order to establish conditional statements, based not on the rule's reliability to preserve truth but on the fact that it supplies the conditional symbol's meaning.

For the same reason, Frege's technique of establishing conditionals by deriving their consequent formulas from their antecedent formulas is consistent with his skepticism about inference from arbitrary hypotheses: He thought neither that proceeding in this way counted as inference nor that it depended on *modus ponens*'s deductive validity. He recognized a second use of *modus ponens* to establish, based on the meaning of the conditional operator, that a conditional statement is in fact valid, from which he concluded that it could be inferred.

It has proven to be valuable, for various purposes, to treat the deduction theorem at times as a discovery and at other times as a requirement. What hasn't emerged in logic's development is any indication that it is essentially one or the other of these things. It could be fair to point out to Tarski or Herbrand, upon their announcement of a proof of the deduction theorem, that they have displayed as a discovery something that in fact was a requirement. After all,

[5] See §7 of *Franks 2021*.

their rules and axioms each individually were devised in accordance with the same intuitive meanings of the logical connectives that secure the theorem's truth. Equally fair would be their rejoinder that, for all that, the axiom system could have been too meager to provide the direct derivation of all the valid conditional statements. This is exactly the dual to the challenge to Frege that he has treated the deduction theorem as a requirement, when in fact it needs to be verified. A fair reply on Frege's part would be that the axioms used in that verification are just a residue of *modus ponens* and the deduction theorem itself.

This example illustrates the practical value of bearing in mind what principles are in fact requirements in one's logical verification. What about the other sense of "requirement" that Wittgenstein seemed to reference: the "requirement" that logic be sublime and that the world conform to it? Notably, Frege does not pretend that his conditional symbol adequately formalize natural language conditionals. The latter, he stresses, connote a necessary or causal connection that isn't captured by his system.[6] But he is untroubled by this "conflict": He is content instead to explore the consequences of imposing this conditional on his investigations, on coordinating it with his notion of formal derivation and discovering how much mathematics can be recreated and analyzed with its guidance.

3. Gödel on disjunction

Other than half a dozen brief review articles, Kurt Gödel's first publication after his celebrated incompleteness theorems was the single-page "Zum intuitionistischen Aussagenkalkül" (*1932*). The explicit goal of this paper was a demonstration that the logical connectives of the intuitionistic propositional calculus (IPC) are not truth-functional. In the course of establishing this claim, Gödel accomplished more: the introduction of the class of "Gödel logics," which have been the topic of ensuing research, possibly the first explicit use of the "pigeon hole principle" in combinatorics, and the first mention of the "disjunction property" for IPC. This last item is especially noteworthy: Gödel announced the disjunction property without offering any proof or justification. The property is that $\vdash_{\text{IPC}} \mathrm{A} \vee \mathrm{B}$ only if either $\vdash_{\text{IPC}} \mathrm{A}$ or $\vdash_{\text{IPC}} \mathrm{B}$. This has since been established by several logicians beginning with Gentzen in *1934–35*. However, all known demonstrations of the disjunction property rely on coordinating IPC-provability with elaborate technological innovations—e.g., Kripke frames *via* the Kripke completeness theorem, the Kleene/Aczel SLASH operator, and in Gentzen's case, the cut-free sequent calculus PJ—to which Gödel had no access. This raises the question: How did Gödel come to know of the disjunction property in 1932?

To approach this mystery, it is helpful to consider first what the disjunction property has

[6] In *Begriffsschrift* Frege says that a sentence of the form A → B which "denies the case in which B is denied but A is affirmed" can only be read as "If A, then B" if, further, "a causal connection is present" (§5)

to do with the explicit goal and main construction in Gödel's paper, so that it even makes sense for him to include mention of it there. Gödel provided little insight even on this question. He simply concluded with the sentence: "Besides, the following holds with full generality: a formula of the form $A \vee B$ can only be provable in [IPC] if either A or B is probably in [IPC]."[7] This suggests that the disjunction property might somehow generalize a specific observation from earlier in the paper.

Let us attend, then, to the paper's main focus, the proof that the IPC connectives are not truth-functional. The question about the proper interpretation of the IPC connectives was on many minds in the 1930s. Because the intuitive understanding of intuitionism is to replace truth with the notion of having been constructively verified, some authors suggested that IPC might be a 3-valued logic: one value indicating that a statement was constructively verified, another value indicating that it had been constructively refuted, and a third value indicating that neither of these is the case. This seems at least to accord with IPC's most famous feature: its rejection of the "tertium non datur" principle, the fact that $\not\vdash_{\text{IPC}} A \vee \neg A$. However, because there are only a small number of 3-valued semantic environments, others were able to rule out this possibility case-by case. This led to speculation that the right semantics might have valence 4 or possibly 5. Gödel's proof rules out all truth-functional semantics of any finite valence.[8]

To prove this, Gödel defined two sequences: a sequence $\{S_n\}$ of semantic environments with valence n and a sequence $\{\mathrm{F}_n\}$ of formulas. He then observed:

1. Given any finite-valued semantic environment *MV*, if IPC is "sound" with respect to MV ($\vdash_{\text{IPC}}$ A only if $\models_{MV}$ A), then for n larger than the valence of *MV*, $\models_{MV} \mathrm{F}_n$. (by the pigeon-hole principle).

2. IPC is "sound" with respect to each S_n ($\vdash_{\text{IPC}}$ A only if $\models_{S_n}$ A).

3. $\not\models_{S_{n+1}} \mathrm{F}_n$.

From these observations, the argument is straightforward: Suppose IPC is sound with respect to *MV*, and the valence of *MV* is n. By 1, $\models_{MV} \mathrm{F}_{n+1}$. But 2 and 3 together imply that $\not\vdash_{\text{IPC}} \mathrm{F}_{n+1}$. Therefore IPC is not complete with respect to *MV*. So for no *MV* do we have $\models_{MV}$ A

[7] "Es gilt übrigens ganz allgemein, daß eine Formel der Gestalt $A \vee B$ in *H* nur dann beweisbar sein kann, wenn entweder A oder B in *H* beweisbar ist"—here "*H*" is the label for a particular axiomatization of IPC from *Heyting 1930*.

[8] *Mancosu and van Stigt 1998* is an excellent source for this history of attempts to pin truth functional interpretations on IPC and to replace *tertium non datur* with quartum non datur. See also Itenary VII of *Mancosu, Zach, and Badesa 2009*. Gödel explicitly raised the question about whether IPC is susceptible to an analysis in terms of finite-valued truth-functionality in *1933* (a short note only published after the answer appeared in *Gödel 1932*).

if, and only if, $\vdash_{\text{IPC}}$ A.

What might the disjunction property have to do with this argument? One idea is to consider the sequence $\{\mathrm{F}_n\}$ and observe that these formulas are disjunctions. Using A $\leftrightarrow$ B as an abbreviation of $(\mathrm{A} \rightarrow \mathrm{B}) \wedge (\mathrm{B} \rightarrow \mathrm{A})$, F_n is defined as:

$$\begin{array}{r}
(\mathrm{P}_1 \leftrightarrow \mathrm{P}_2) \vee (\mathrm{P}_1 \leftrightarrow \mathrm{P}_3) \vee \ldots \vee (\mathrm{P}_1 \leftrightarrow \mathrm{P}_n) \vee \\
(\mathrm{P}_2 \leftrightarrow \mathrm{P}_3) \vee \ldots \vee (\mathrm{P}_2 \leftrightarrow P_n) \vee \\
\ldots \vee (\mathrm{P}_3 \leftrightarrow \mathrm{P}_n) \vee \\
\ldots \\
\vee (\mathrm{P}_{n-1} \leftrightarrow \mathrm{P}_n)
\end{array}$$

Because $\nvdash_{\text{IPC}}$ $\mathrm{P}_i \leftrightarrow \mathrm{P}_j$, the disjunction property guarantees that $\nvdash_{\text{IPC}}$ F_{n+1} (in my experience, most students who try to recreate Gödel's proof appeal in this way to the disjunction property). But Gödel did not appeal to the disjunction property in order to verify the unprovability of the $\{\mathrm{F}_n\}$ formulas. We saw that he inferred it directly from 2 and 3.

Another possibility is that Gödel saw a converse relationship. His explicit proof demonstrates the unprovability of each F_n, itself a disjunction none of whose disjuncts is IPC-provable. Could the same construction, or some variation of it, be used to demonstrate the unprovability of each such disjunction?

In order to consider this possibility, one must attend to the details of the sequence $\{S_n\}$. S_n is defined as:

$$\begin{aligned}
v(\neg\mathrm{A}) &= \begin{cases} 0 & v(\mathrm{A}) = n-1 \\ n-1 & \text{otherwise} \end{cases} \\
v(\mathrm{A} \wedge \mathrm{B}) &= \mathbf{max}\{v(\mathrm{A}), v(\mathrm{B})\} \\
v(\mathrm{A} \vee \mathrm{B}) &= \mathbf{min}\{v(\mathrm{A}), v(\mathrm{B})\} \\
v(\mathrm{A} \rightarrow \mathrm{B}) &= \begin{cases} v(\mathrm{B}) & v(\mathrm{A}) < v(\mathrm{B}) \\ 0 & \text{otherwise} \end{cases} \\
v(\bot) &= n-1
\end{aligned}$$

with $\models_{S_n}$ A if $v(\mathrm{A}) = 0$ for all assignments of values from $\{0, 1, \ldots, n-1\}$ to atomic formulas.

To verify 2, notice that S_2 is just the classical bivalent theory of truth, with 0 = TRUE and 1 = FALSE. So the usual soundness argument for CPC applies to it. The same argument shows that IPC is sound for each S_n: consider for example a natural deduction presentation of propositional logic and observe that each inference rule other than the characteristically classical double negation rule $\frac{\neg\neg \mathrm{A}}{\mathrm{A}}$ remains sound as n increases. ($\frac{\neg\neg \mathrm{A}}{\mathrm{A}}$ fails to be valid already in S_3.)

To verify 3, simply observe that one can assign a different value to each atom in F_{n+1}. On any such valuation, each disjunct will then evaluate as non-zero. The value of the entire formula will then, being the minimum of its disjuncts' values, also be non-zero.

This is all one needs to observe in order to verify that IPC does not have a finite-valued truth functional interpretation. The formulas F_n serve as counterexamples to the completeness of IPC with respect to any truth-functional semantics for which it is sound. Can the same construction also demonstrate the unprovability, not just of the F_n, but of every disjunction whose disjuncts are each unprovable? If so, then we could understand Gödel's remark that the disjunction property is somehow a "fully general" version of the paper's main result.

No such direct verification of the disjunction property is available, though. Consider the formula $(\mathrm{A} \to \mathrm{B}) \vee (\mathrm{B} \to \mathrm{A})$. Clearly neither of its disjuncts is IPC-provable. Nevertheless, it is valid in each S_n: No matter what values are assigned to A and to B, one or the other disjunct will evaluate as 0, so that the disjunction, too, will evaluate as 0.

If Gödel's construction cannot be used to verify the general disjunction property of IPC, perhaps a modification of it will do. This idea does not appear promising, though. Any adjustment of the valuation function on S_n such that $\vee$ doesn't return the minimal value of its disjuncts or that $\to$ doesn't return 0 when its antecedent's value is less than or equal to its consequent's value will disrupt IPC-soundness.

Another place to look for clues for how Gödel knew about the disjunction property is his *1933a* paper "Eine Interpretation des intuitionistischen Aussagenkalküls." Here again Gödel stated the property, but this time he made reference to a related property of a different formal system (modal system S4) and described an embedding of IPC into this system. A verification of the related property of S4 would translate *via* this embedding as a verification of the disjunction property for IPC.

In order to assess the plausibility of this line of reasoning, we review the details of S4 and Gödel's embedding of IPC into it. S4 is CPC supplemented with

- a new primitive symbol: $\Box$,
- a new formula formation rule: If A is a formula, then $\Box$A is a formula,

a new rule of inference

- $\frac{\mathrm{A}}{\Box \mathrm{A}}$

and three new axioms

1. $\Box A \to A$
2. $\Box A \to (\Box(A \to B) \to \Box B)$
3. $\Box A \to \Box\Box A$

Gödel defined a translation of formulas from IPC into the language of S4 by:

$$\begin{aligned}\textbf{trans}(\neg A) &= \Box\neg\Box\textbf{trans}(A)\\ \textbf{trans}(A \to B) &= \Box\textbf{trans}(A) \to \Box\textbf{trans}(B)\\ \textbf{trans}(A \vee B) &= \Box\textbf{trans}(A) \vee \Box\textbf{trans}(B)\\ \textbf{trans}(A \wedge B) &= \Box\textbf{trans}(A) \wedge \Box\textbf{trans}(B)^{9}\end{aligned}$$

The main claim of the paper is

$$\vdash_{\text{IPC}} A \text{ if, and only if, } \vdash_{\text{S4}} \textbf{trans}(A). \qquad (*)$$

Gödel didn't supply a proof, submitting only that the claim is "*vermutlich*."[10] But without any such qualification, Gödel declared both that "the translation of A ∨ ¬A is not derivable in S4" and also that "neither in general is any formula of the form $\Box A \vee \Box B$ for which neither $\Box A$ nor $\Box B$ is already provable in S4":

$$\vdash_{\text{S4}} \Box A \vee \Box B \text{ only if either } \vdash_{\text{S4}} \Box A \text{ or } \vdash_{\text{S4}} \Box B \qquad (**)$$

(*) and (**) together obviously entail the disjunction property for IPC. Furthermore, the fact that he raised the question of the truth-functionality of IPC in the earlier but belated note *1933* makes it clear that the relationship between IPC and S4 already occurred to him before he discovered the result of *1932*. Could Gödel have been aware of the disjunction property because of prior insight into S4 and its relationship to IPC?

The obstacle to such a suggestion is that the first proof of (**) was published by McKinsey and Tarksi in *1948*, this time using algebraic methods that would be foreign to Gödel. If there is a question about how Gödel could have known about the disjunction property for IPC in 1932, there is to the same extent a question how he could have known about (**) in 1933.

Having reached this impasse, one might consider that the time has come again to "turn our whole examination around." We have been asking how Gödel could have discovered that IPC has the disjunction property given only the constructions and techniques known to him

[10] The first published proof is in *McKinsey and Tarksi 1948*.

in the early 1930s. This is a bad question if the disjunction property did not stand for Gödel as something to discover. What if, instead of the "result of an investigation," the disjunction property was for Gödel a "requirement?"

Other than the fact that Gödel neither provided a proof of the disjunction property nor suggested that one should be sought, the first hint that he might have taken the property as a starting point of inquiry is the fact that in *1933a* he read the modal operator $\Box$, not as a necessity operator (as is customary and as it is treated in the literature he cites (*Becker 1930*, *Parry 1933*, *Lewis 1932*)), but as a provability operator: Gödel opened his paper with the declaration that "$\Box$A" has the intuitive meaning "A is provable," and he closed with the observation that provability in this context cannot be understood as "provability in a certain formal system S" but instead as "provable by any correct means of proof."[11] It would require some clever argument to verify (**) based on a formal investigation of the theorems of S4. But it is not difficult at all to justify each axiom of S4 against Gödel's informal notion of provability. For example, axiom 1 says that if A is provable, then it is true, and axiom 2 says that if both A and A $\rightarrow$ B are provable, then B is, too. Both of these claims are obviously true when the notion of provability one has in mind is that of "provability by any correct means." Similar justifications apply to axiom 3 and the rule $\frac{\mathrm{A}}{\Box\mathrm{A}}$, bearing in mind that the latter is meant to apply only to theorems of S4, not to arbitrary formulas. But if S4 is sound with respect to the informal notion of provability, then (**) is immediate: If neither A nor B are provable in the informal sense, then it cannot be the case that either A is provable or B is provable.

This is not to suggest that Gödel did, in fact, infer the disjunction property for IPC from (**). Again, reading Gödel that way is to get the context of discovery backwards and to underestimate the significance of the disjunction property's role as a requirement. Heyting and Kolmogorov had already justified the axioms of IPC against the informal notion of provability. Gödel knew that the intuitionistic disjunction A $\vee$ B was supposed to mean that there is either a proof of A or there is a proof of B. The soundness of Heyting's axioms with respect to the informal notion of provability therefore guarantees that such formula will only ever be provable when either A or B already are.

Gödel's point in *1933a* was that this property is reflected again, *via* the translation he described, in the theorem structure of S4 as (**). Rather than infer the disjunction property of IPC from S4, the disjunction property of IPC led Gödel to a reinterpretation of S4, and on this reinterpretation (**) can be observed as obviously true. This in turn specifies the precise translation scheme for the embedding of IPC in S4.

[11] The reason is that by applying the rule $\frac{\mathrm{A}}{\Box\mathrm{A}}$ to axiom 1, one can derive $\Box(\Box\mathrm{A} \rightarrow \mathrm{A})$. Substituting $0 = 1$ for A, one would then have $\Box\neg\Box(0 = 1)$, and were $\Box$ to mean "provable in S," this could say that the consistency of S is provable in S in conflict with Gödel's (*1931*) second incompleteness theorem.

Finally, what is one to make of Gödel's remark in *1932* that "besides" the explicit construction and result featured there, the disjunction property holds in "full generality?" In that paper Gödel showed that a certain type of disjunctive formula of the form F_n is unprovable in IPC. From this it follows that IPC-provability cannot be reduced, as many of his contemporaries had suggested it could be, to any method of analysis so simple as finite valued truth-functionality. But the failure of any finite truth-functional analysis of IPC-provability is directly evident already from the disjunction property itself: You describe a truth-functional semantics with finite valence. If IPC is sound with respect to this semantics, I should always be able to construct a disjunction, neither of whose disjuncts is a theorem of IPC, that is valid on that semantics, witnessing thereby the incompleteness of IPC with respect to it. This general phenomenon is that guiding idea behind the *1932* result, the one that led to the discovery of the specific sequence of F_n.

Rosalie Iemhoff's magnificent *2015* paper "On rules" concludes with an observation that admissibility in IPC currently lacks full explanation. She contrasts this with the situation regarding "the disjunction property" which she says "is perfectly explainable from the constructive point of view. That it really holds still requires a proof, but the meaning of disjunction in intuitionistic logic foretold us that it would hold." The known proofs of the disjunction property for IPC are valuable for many reasons. They provide insight into the structure of Kripke frames, highlight subtle structural features of the cut-free sequent calculus, and can inform our understanding of which features of intuitionistic inference account for its truth. But with Gödel I am not convinced that a proof ever was needed to show that it "really holds." For Gödel it wasn't a discovery but a requirement, and he treated it as such. As part of the very understanding of intuitionistic disjunction, it stood for Gödel not as something in need of verification but as a guiding principle leading to two discoveries: the irreducibility of intuitionistic proof to finite combinatorial calculation and intuitionism's translatability into a classical framework supplemented with a modal operator characterized by principles sound with respect to informal provability.

4. Megillat Esther

Xerxes, of sea-whipping infamy, was king of the Achaemenid empire in the 5th Century BCE. He is also portrayed in Jewish literature as the ambitious king in *Megillat Esther* whose scheme to annihilate the Jewish diaspora population in Persia is modified and actualized as an edict against the Jews' enemies. In *Megillat Esther*, his name is semitized as "Ahasuerus."

The story of *Megillat Esther* lays the foundation for the holiday of Purim, an especially festive occasion on which Jews celebrate the transformation of a genocidal edict into a guarantee of security and protection. Among the holiday's central observances is the public recitation of *Megillat Esther* in its entirety. This recitation is governed by several detailed laws, among which one is particularly peculiar. The Babylonian Talmud reads: “One who reads the Megilla backwards has not fulfilled one’s obligation” (tractate Megilla, 17a). Because this injunction against such a seemingly far-fetched error has puzzled rabbinical scholars, several commentators have proposed interpretations of it on a homiletical level. We wish to suggest still another, inspired by the historical episodes recounted above.

Megillat Esther is the ultimate celebration of situational irony: Haman is hanged on the gallows he erected to execute his chief rival, Mordechai. Mordechai occupies the royal promenade that Haman arranged for his own glorification. Esther becomes queen of Persia through her very efforts to subvert the royal edicts. The Purim holiday redoubles these inversions with its customs to dress up as one’s own enemy and to indulge in the same sort of festive banquet that exemplified the Persian antithesis to Jewish values. But amid it all comes the most peculiar injunction from Jewish Law: a prohibition on reciting the whole story backwards.

What could be the meaning of this prohibition? What, for that matter, could be the grounds for any suspicion that someone might attempt a backwards recitation? I suggest that the answer is that in the resistance against Ahasuerus’s/Xerxes’s narrative lies a temptation to fall into another trap. Here is a figure who would shackle the sea for defying his preconceptions, his ideas, the logical order that he cannot conceive a violation of. Here is a figure who sets a plan in motion and expects it to unfold according to a logical progression, carrying his original intention to its rational conclusion. If Ahasuerus wanted to sink everything in the tidy narrative of logic and reason, a reflexive opposition would be to subsume logic itself to the natural unfolding of the world, to give up all preconceptions. This would be to reverse the scroll’s narrative.

A subtler inversion, though, is to let the story play out in its logical order and just to observe how, despite the rational grid we impose on the world, be it through ordinances or scripted narrative, a super-logical scheme emerges. As Xerxes whipped the sea when it defied his intentions, one might find the conflict between our logical requirements and the observed behavior of the world to be intolerable. One antidote to this impulse is to reconcieve logic as the most general laws of nature itself, as something to be read out empirically from observation of the world. But this is to “read the Megilla backward,” an overreaction inspired, not by the failure of the logic for the purposes it really serves—reasoning, mathematics, etc.—but by its failure to live up to our expectation that it be inviolable.

The right antidote is instead not to lose sight on what we are requiring and what we are discovering, to be willing to impose logical structure provisionally in order to determine what one can discover thereby, and never to expect that the framework we have settled on is final—on the contrary, to be ever vigilant for hints of a hitherto hidden world revealing itself between the cracks of the mold we try to cast it in. When we are aware of and willing to acknowledge them, our “requirements,” like Frege’s and Gödel’s, are not in opposition to discoveries in logic and mathematics. They facilitate them.

Works cited

Becker, O. “Zum Logik der Modalitäten.” *Jahrbuch für Philosophie und phänomenologische Forschung* **11**, 495–548.

Bynum, T. W. (ed.) 1972. *Conceptual Notation and Related Articles*. New York: Oxford University Press.

Franks, C. 2018. “The context of inference.” *History and Philosophy of Logic*. **39**(4), 365–95.

Franks, C. 2021. “The deduction theorem (before and after Herbrand).” *History and Philosophy of Logic*. **42**(2), 129–59.

Frege, G. 1980. *Philosophical and Mathematical Correspondence*. G. Gabriel, et al. (eds.). Chicago: University of Chicago Press.

Frege, G. 1879. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*. Halle: L. Nebert. Translated by T. W. Bynum as *Conceptual Notation: a formula language of pure thought modeled upon the formula language of arithmetic* in *Bynum 1972*, 101–208.

Frege, G. 1910. “Letter to Jourdain,” translated and reprinted in *Frege 1980*.

Frege, G. 1917. “Letter to Dingler,” translated and reprinted in *Frege 1980*.

Gentzen, G. 1934–35. “Untersuchungen über das logische Schliessen.” Gentzen’s doctoral thesis at the University of Göttingen, translated as “Investigations into logical deduction” in *Szabo 1969*, 68–131.

Girard, J. Y. 2003. “Between logic and quantic: a tract,” *Mathematical Structures in Computer Science*.

Girard, J. Y. 2011. *The Blind Spot: Lectures on Logic*. European Mathematical Society.

Gödel, K. 1931. “Uber formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I.” *Monatshefte für Mathematik und Physik* **38**, 173–98.

Gödel, K. 1932. “Zum intuitionistischen Aussagenkalkül.” *Anzeiger der Akademie der Wissenschaften in Wien* **69**, 65–6. Translated as “On the intuitionistic propositional calculus,” in *Feferman et al. 1986*, 223–25.

Gödel, K. 1933. Untitled Remark following *Parry 1933*. *Ergebnisse eines mathematischen Kolloquiums* **4**, 4.

Gödel, K. 1933a. "Eine Interpretation des intuitionistischen Aussagenkalküls." *Ergebnisse eines mathematischen Kolloquiums* **4**, 39–40.

Goldfarb, W. 1971. *Jacques Herbrand: Logical Writings*. Cambridge: Harvard University Press.

Haaparanta, L. (ed.). 2009. *The Development of Modern Logic*. New York: Oxford University Press.

Herbrand, J. 1930. *Recherches sur la theorie de la démonstration*. Herbrand's doctoral thesis at the University of Paris. Translated by W. Goldfarb, except pp. 133–88 trans. by B. Dreben and J. van Heijenoort, as "Investigations in proof theory" in *Goldfarb 1971*, 44–202.

Heyting, A. 1930. "Die formalen Regeln der intuitionistischen Logik." *Sitzungsberichte der Preussischen Akademie der Wissenschaften, physikalisch-mathematische Klasse*, 42–56.

Iemhoff, R. 2015. "On Rules." *Journal of Philosophical Logic* **44**(6), 697–711.

Lewis, C. I. 1932. "An alternative system of logic." *The Monist* **42**, 481–507.

Maddy, P. 2012. "The philosophy of logic." *The Bulletin of Symbolic Logic* **18**(4), 481–504.

Maddy, P. 2014. *The Logical Must: Wittgenstein on logic*. New York: Oxford University Press.

Mancosu, P. 1998. *From Brouwer to Hilbert: the debate on the foundations of mathematics in the 1920s.* New York: Oxford University Press.

Mancosu, P. and W. P. van Stigt. 1998. "Intuitionistic logic," in *Mancosu 1998.*

Mancosu, P., R. Zach, and C. Badesa. 2009. "The development of mathematical logic from Russell to Tarski," in *Haaparanta 2009.*

McKinsey, J. and A. Tarski. 1948. "Some theorems about the sentential calculi of Lewish and Heyting." *The Journal of Symbolic Logic* **13**, 1–15.

Parry, W. T. 1933. "Zum Lewisschen Aussagenkalkül." *Ergebnisse eines mathematischen Kolloquiums* **4**, 15–17.

Godley, A. D. (trans.). 1920. *Herodotus, with an English Translation*. Cambridge: Harvard University Press.

Szabo, M. E. 1969. *The Collected Papers of Gerhard Gentzen*. London: North Holland.

Tarski, A. 1956. *Logic, Semantics, Metamathematics*. J. H. Woodger (trans.). New York: Oxford University Press

Wittgenstein, L. 1922. *Tractatus Logico-Philosophicus*. Edited by C. K. Ogden. Routledge and Kegan Paul.

Wittgenstein, L. 1953. *Philosophische Untersuchungen*. Translated by G. E. M. Anscombe as *Philosophical Investigations*. Second Edition (1958). Walden, Massachussetts: Blackwell.