# NATURAL FORMALIZATION: DERIVING THE CANTOR-BERNSTEIN THEOREM IN ZF

WILFRIED SIEG AND PATRICK WALSH

**Abstract.** The Cantor-Bernstein Theorem (CBT) is a classical result of general set theory. We have formalized a number of its proofs in the system ZF of Zermelo-Fraenkel. The formalizations are carried out via the system AProS, expanded to a convenient logical and set theoretic inference mechanism. AProS serves as a *proof assistant* and allows the direct construction of formal proofs that are humanly intelligible. Indeed, it is designed to emulate human proof constructions in a conceptually organized deductive frame.

The proof construction itself proceeds, importantly, by considering *partial proofs with gaps* that allow natural, systematic forward and backward moves. The development of such *human-centered* theorem proving takes steps to investigate "the concept of the specifically mathematical proof". It sharpens mathematical practice, facilitates the comparison of proofs, and uncovers their "rational structure", in particular, their motivating principal ideas.

In the case of CBT, the multitude of its "different" proofs has been reduced to exactly one. It comes in two types due to Dedekind and Zermelo, respectively. This seems to be a beautiful confirmation of the expectation expressed in Hilbert's 24th problem that, "under a given set of conditions, there can be but one simplest proof". We indicate directions for research at the intersection of (philosophy of) mathematics, cognitive psychology, and computer science.

> We must — that is my conviction — take the concept of the specifically mathematical proof as an object of investigation.
>
> Hilbert 1918

**§1. Proofs: the objects of proof theory.** The objects of proof theory are proofs, of course. This assertion is however deeply ambiguous. Are proofs to be viewed as formal derivations in particular calculi? Or are they to be viewed as the informal arguments given in mathematics? — The contemporary practice of proof theory suggests the first perspective, whereas the programmatic ambitions of the subject's pioneers suggest the second. We will later mention remarks by Hilbert (in sections 5 and 7) that clearly point in that direction. Now we refer to Gentzen who inspired modern proof theoretic work; his investigations and insights concern *prima facie* only formal proofs. However, the detailed discussion

of the proof of the infinity of primes in his [Gentzen, 1936, pp. 506–511] makes clear that he is very deeply concerned with *formalizing* mathematical practice. The crucial problem is finding the atomic inference steps involved in informal arguments. The inference steps Gentzen brings to light are, perhaps not surprisingly, the introduction and elimination rules for logical connectives, including quantifiers.

Gentzen specifies in [Gentzen, 1936, p. 513] the concept of a *deduction* and adds in parentheses *formal image of a proof*; i.e., deductions are viewed as formal images of mathematical proofs and are obtained by formalizing the latter. The process of formalization is explained as follows: "The words of ordinary language are replaced by particular *signs*, the logical inference steps [are replaced by] rules that form new formally presented statements from already proved ones." Only in this way, he claims, is it possible to obtain a "rigorous treatment of proofs". However, and that is strongly emphasized, "The objects of proof theory shall be the *proofs* carried out in mathematics proper." [Gentzen, 1936, p. 499] For us, the formalization of proofs is the quasi-empirical starting point for uncovering *proof methods in mathematics*: formal rigor is not to be considered a foe of simplicity or understanding.

When extending the effort from logical to mathematical reasoning one is led to the task of devising additional tools for the *natural formalization of proofs*. Such tools should serve to directly reflect standard mathematical practice and preserve two central aspects of that practice, namely, (1) the axiomatic and conceptual organization in support of proofs and (2) the inferential mechanisms for logically structuring them. Thus, the natural formalization in a deductive framework *verifies* theorems relative to that very framework, but it also deepens our understanding and isolates core ideas; the latter lend themselves often, certainly in our case, to a direct diagrammatic depiction of a proof's conceptual structure.[1]

We chose as the deductive framework Zermelo-Fraenkel set theory ZF. One can clearly choose different ones, for example, Higher Order Logic, Martin Löf's Type Theory or Feferman's Explicit Mathematics. The language of set theory is, however, the *lingua franca* of contemporary mathematics and ZF its foundation. So it seems both important and expedient to use ZF for the project of formalizing proofs naturally. As to the inferential mechanism, we implemented it in AProS, which is based on the *intercalation calculus* investigated in [Sieg and Byrnes, 1998]; that calculus modifies the *natural deduction calculus* in three deeply connected

[1] That, in turn, is the starting point of the *automated search for humanly intelligible proofs* or, what Gowers calls, *human-centered automatic theorem proving*; see his interview [Diaz-Lopez, 2016]. This topic is discussed in greater detail in section 7 and Appendix B.

ways. *First*, it has introduction and elimination rules for logical connectives and mathematical definitions. *Second*, it facilitates the "top down" and "bottom up" construction of proofs in a bi-directional, strategically motivated way. This requires the representation of *partial proofs* or *proofs with gaps* and turns it into a multiple-conclusion calculus. Third, we have a general procedure for using *lemmas-as-rules* to connect the conceptual organization of the deductive frame with the stepwise construction of proofs.

Together, these tools allow us to prove the Cantor-Bernstein theorem (CBT) formally and intelligibly in ZF.[2] CBT is a classical result; when writing his Notes for Cantor's *Collected Works* [Cantor, 1932], Zermelo characterized it as "one of the most important and most elementary *Hauptsätze* of general set theory" (p. 351). The theorem states for all sets $a$ and $b$, if there is an injection $f : a \to b$ and an injection $g : b \to a$, then there is a bijection $h : a \to b$. In section 2 we present a variety of proofs and analyze them from a perspective that is mathematically conceptual, not logically formal. In this way we separate the "foundational work" in set theory from the "mathematical work" that is driven by leading ideas and heuristics.

The *framework* for the formalization, implemented as an expansion of the proof search system AProS, is presented in section 3. It is a definitional extension of ZF with an inference mechanism whose principal features we sketched above and will discuss in detail below. However, formalizing in this frame even a well-organized argument is tedious and leads to unintelligibly long proofs. So we complement the inference mechanism with a careful organization of the material leading up to the CBT.

The material is layered into *conceptual levels* and the associated hierarchy is described in section 4. Section 5 presents the top-level proof of the CBT together with the proof of Dedekind's equivalent *Fundamental Lemma*. In addition to presenting these proofs we list the lower level lemmas to which they appeal in Appendix A. It is our contention that

[2]The interest in set theory as a framework for the formalization and verification of mathematics is currently rather limited. The significant earlier interest is described in [Paulson, 1993] and [Paulson, 1995]; indeed, Paulson thought of his own papers as providing "a productive environment for verification." There is the ongoing use of Mizar (see, `mizar.org`) and the work reflected in [Pastre, 2002, 2011], as well as in [Windsteiger, 2006]. The Mizar approach is discussed sympathetically in [Wiedijk, 2011] in precisely the context in which the AProS project is located, namely, the direct formal reflection of mathematical practice. Paulson verified in his [Paulson, 1995] the Cantor-Bernstein Theorem. As a matter of fact, the proof of the theorem had been formalized already with the EXCHECK system; see [Blaine, 1981], [McDonald and Suppes, 1984]. The latter paper presents on pp. 338–349 heuristic guidance for students to find a proof of the theorem when interacting with EXCHECK, explicitly developed for axiomatic set theory.

these lemmas are mathematically meaningful. They can be labeled "exercises" for readers who understand the developments at the lower level, or they can be seen as steps that should be taken automatically, we hope very soon. Finally, the modularity of our development allows us to informatively compare proofs of the Cantor-Bernstein theorem and make observations on the "identity" of proofs.

In section 6 we describe the need to formalize (parts of) mathematics in the service of foundational investigations during the late 19th and early 20th century; that later required the clarification of the very concept of "formal system" in the work of Gödel, Church, and Turing. Section 7 reemphasizes *programmatic reflections* on our approach to human-centered proof construction. As *context* for our work, we point to contemporary connections with [Barendregt and Wiedijk, 2005] on the logical side and with [Ganesalingam and Gowers, 2013] on the mathematical side. We discuss there also future directions, i.e., steps towards an automated *strategic proof search* with heuristic guidance that reflects human cognitive capacities. The connection to [Ganesalingam and Gowers, 2013] is detailed in Appendix B, where we solve, strictly following our strategic advice, a proof problem they consider as paradigmatic.

**§2. Chains, approximations and fixed-points.** Cantor conjectured 'our' theorem in 1882, but could not prove it without assuming his comparability or well-ordering principle.[3] Dedekind in 1887 and Bernstein in 1897 found the earliest proofs that do not appeal to the well-ordering principle, but they did not publish their proofs. König and Zermelo, among others, gave proofs in the first decade of the 20th century and essentially used Dedekind's *chain* theory to make explicit elementary inductive definitions. In Bernstein's proof, natural numbers are presupposed and are used for *approximating* an inductively specified set from below. Knaster and Tarski, finally, found a different method for achieving the same goal through a *fixed-point* construction for monotone functions [Knaster and Tarski, 1928] and [Tarski, 1955, note 2]. The proofs without choice have a definite methodological interest: they seem to be *different* and challenge us to pinpoint the differences.

[3]For the complex and intriguing history of proofs of the theorem we refer to the accounts in [Hinkis, 2013], [Deiser, 2010, §3] and [Kanamori, 2004, §4]. The early history involves Cantor, Dedekind, [Schröder, 1898], and Bernstein, but also [Korselt, 1911], [Peano, 1906a], [König, 1906], and [Zermelo, 1908] as well as [Borel, 1921], [Poincaré, 1906], [Whitehead and Russell, 1912] and [Whitehead, 1902].

Dedekind's proofs, one from 1887 and one from 1899, were published in [Dedekind, 1932, pp. 447–449] and [Cantor, 1932, p. 449], respectively; Bernstein's proof was first published in [Borel, 1921, pp. 104–106]. On the way to our formalization, we verified Dedekind's proof together with the necessary development of his chain theory presented in [Dedekind, 1888]; the latter development was of course carried out in ZF.

The terms in the heading of this section are associated with these three methods of making inductive definitions explicit. Dedekind introduced the first method, and we prove with its help his *Fundamental Lemma*. The Lemma not only implies the Cantor-Bernstein theorem, but is actually equivalent to it.[4]

THEOREM 2.1 (Fundamental Lemma (Dedekind)). *Let $h$ be a bijection from $a$ to $e$ and let $d$ be a set with $e \subseteq d \subseteq a$; then there is a bijection $h^*$ from $a$ to $d$, i.e., $a \approx d$.*

The Cantor-Bernstein Theorem is an immediate consequence.

THEOREM 2.2 (Cantor-Bernstein Theorem). *Let $f$ be an injection from $a$ to $b$ and $g$ an injection from $b$ to $a$; then there is a bijection $h$ from $a$ to $b$, i.e., $a \approx b$.*

PROOF. We want to establish $a \approx b$. As $g$ is an injection from $b$ to $a$, we have $b \approx g[b]$. If we have also (*) $a \approx g[b]$, the claim of the theorem follows by transitivity of the $\approx$-relation. We obtain (*) from the Fundamental Lemma, since its assumptions can be directly verified: composing $f$ and $g$ yields a bijection from $a$ to $g \circ f[a]$ such that $g \circ f[a] \subseteq g[b]$; $g$ is a function from $b$ to $a$, thus, $g[b] \subseteq a$. ⊣

The Theorem implies the Lemma: observe that $h$ is an injection from $a$ to $d$ and that the identity on $d$ is an injection from $d$ to $a$. Thus, by the Theorem, there is a bijection from $a$ to $d$. Let us now prove the Fundamental Lemma. Figure 1 reflects the motivation for Dedekind's notion, namely, to characterize $a$'s subset $c$ whose elements are obtained from $a \setminus d$ by finite iteration of $h$. *The* central concept of Dedekind's *Was sind und was sollen die Zahlen?* (*WZ*)[Dedekind, 1888] is that of the *chain of a system $d$ given $h$*, where $h$ is a function from $s$ to $s$, and $d$ is a subset of $s$; we denote such a chain by $c(s, d, h)$. This notion corresponds to the now familiar concept of the inductive closure of $d$ under $h$. Dedekind defined $c(s, d, h)$ as the intersection of all $h$-closed sets containing $d$. Thus, $c(s, d, h)$ is the smallest such set.

[4]The Fundamental Lemma is a simplified version of Proposition 63 in Dedekind's *Was sind und was sollen die Zahlen?* [Dedekind, 1888] The proposition can be stated as follows: Let $h$ be an injection from $a$ to $d$ and let $d \subseteq a$; then there is a bijection $h^*$ from $a$ to $d$. — Dedekind did not formulate the Cantor-Bernstein Theorem in *WZ*, though from his 1887 manuscript we know that he was perfectly well aware of that consequence and gave a direct proof of the theorem. As to the details of this peculiar situation see [Ferreirós, 1993]. — The *Fundamental Lemma* we formulate next is the key theorem in Dedekind's proof, but its formulation is due to Cantor in a letter to Dedekind of 5 November 1882; see [Ferreirós, 1993, p. 354] and [Kanamori, 2004, p. 508]. An English translation of the letter is found in [Ewald, 1996, pp. 874–878]; the Lemma is formulated there on p. 875 and again on p. 878.

PROOF. Let $c$ be $c(a, a \setminus d, h)$ and obtain the function $h^*$ from two bijections, namely, $h_1 : c \to h[c]$ with $h_1(x) = h(x)$ and $h_2 : a \setminus c \to a \setminus c$ with $h_2(x) = id(x)$.

We clearly have $d \setminus h[c] = (d \cup (a \setminus d)) \setminus (h[c] \cup (a \setminus d))$. With $d \cup (a \setminus d) = a$ and the important *structural identity*

$$c = (a \setminus d) \cup h[c] = h[c] \cup (a \setminus d)$$

we have $d \setminus h[c] = a \setminus c$. Thus, $h_2$ is a function $a \setminus c \to d \setminus h[c]$. Both $h_1$ and $h_2$ are bijections; $c$ and $a \setminus c$ partition $a$, whereas $h[c]$ and $d \setminus h[c]$ partition $d$. So, the union $h^*$ of $h_1$ and $h_2$,

$$h^*(x) = \begin{cases} h(x) & \text{if } x \in c \\ id(x) & \text{if } x \in a \setminus c \end{cases}$$

is a bijection from $a$ to $d$. ⊣

A modification of this proof establishes that there is a bijection $h^*$ from $d$ to $e$. Let $c^*$ be the chain of $d \setminus e$ given $h$, $c(a, d \setminus e, h)$, and define $h^* : d \to e$ as follows:

$$h^*(x) = \begin{cases} h(x) & \text{if } x \in c^* \\ id(x) & \text{if } x \in d \setminus c^* \end{cases}$$

The composition of $h$ with the inverse of $h^*$ is a bijection $h^{**}$ from $a$ to $d$. This bijection is directly defined by exploiting $c^*$ as follows: $h^{**}(x) = h(x)$ if $x$ is in $a \setminus c^*$ and $h^{**}(x) = id(x)$ otherwise. Zermelo's proof in his [Zermelo, 1908] proceeds this way; cf. [Kanamori, 2004, pp. 508–509]. The functions $h^*$ and $h^{**}$ are the canonical mappings that are obtained also in the other proofs. That is made clear in our presentation of König's proof joining Dedekind's and Zermelo's considerations. Adapted to our set up, König uses both c and c*. Let $r$ be $e \setminus (h[c] \cup h[c^*])$ and define

$$h_1^*(x) = \begin{cases} h(x) & \text{if } x \in c \\ id(x) & \text{if } x \in c^* \cup r \end{cases}$$

$$h_2^*(x) = \begin{cases} h(x) & \text{if } x \in c \cup r \\ id(x) & \text{if } x \in c^* \end{cases}$$

It is not difficult to verify that $h_1^*$ is the $h^*$ from Dedekind's proof and that $h_2^*$ is the $h^{**}$ from Zermelo's proof.[5]

The proofs of the Cantor-Bernstein Theorem we presented use in an essential way the idea of *inductive closure* or finite iteration of a function $h$ from $s$ to $s$, starting with a subset $d$ of $s$. That was made quite

[5]König's paper appeared in 1906. His informal argument is rigorously presented in [Deiser, 2010, p. 55].

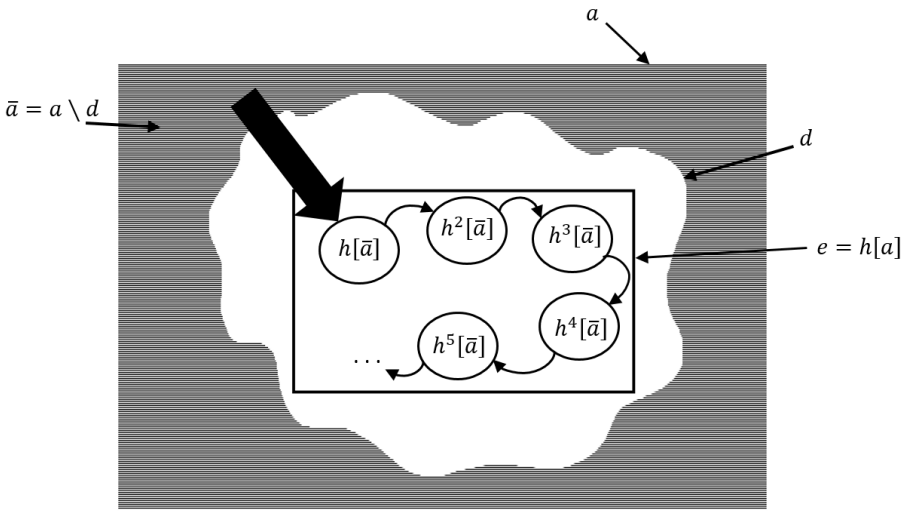



FIGURE 1. Inductive closure of $a \setminus d$ under $h$.

clear by Dedekind: having introduced natural numbers and definition by recursion, he shows in Theorem 131 of *WZ*:

$$c(s, d, h) = \bigcup\{d_i \mid i \in \mathbb{N}\},$$

where the elements of the sequence $\langle d_i \mid i \in \mathbb{N}\rangle$ are defined recursively by

$$d_0 = d \quad \text{and} \quad d_{i+1} = h[d_i]$$

Such an approximation of $c(s, d, h)$ from below underlies Bernstein's proof. Using Dedekind's theorem, we have:

$$c = \bigcup\{\overline{a_i} \mid i \in \mathbb{N}\} \quad \text{and} \quad \overline{a_0} = \overline{a} = a \setminus d \quad \text{and} \quad \overline{a_{i+1}} = h[\overline{a_i}]$$

and

$$c^* = \bigcup\{\overline{d_i} \mid i \in \mathbb{N}\} \quad \text{and} \quad \overline{d_0} = \overline{d} = d \setminus e \quad \text{and} \quad \overline{d_{i+1}} = h[\overline{d_i}].$$

The bijections we defined in Dedekind's and Zermelo's proofs can now be re-obtained. — Using the definition for $x \subseteq a, h^0[x] = x$, and $h^{n+1}[x] = h[h^n[x]]$, we can observe:

(i) for $n > 0, h^n[\overline{a}] \subseteq e = h[a]$,
(ii) for $m \neq n, h^m[\overline{a}] \cap h^n[\overline{a}] = \emptyset$;
(iii) for any $z \in c$, there is a unique $n$ such that $z \in h^n[\overline{a}]$.

Thus, the elements in $c$ have truly been obtained from $\overline{a}$ by finite iteration of $h$. The observations can be adapted to $c^*$ and $\overline{d}$, giving parallel lemmas (i*)–(iii*).

The proofs of the Fundamental Lemma we examined lead to one of two canonical bijections, namely, $h^*$ and $h^{**}$. Their definitions are clearly

based on particular partitions of the set $a$. Using the assumptions of the Cantor-Bernstein Theorem, Banach considers in his [Banach, 1924] defining partitions of the sets $a$ and $b$ as the central and primary task. His theorem[6] can be recaptured in our context as follows: let ${}^*c$ be the chain $c(a, a \setminus g[b], g \circ f)$, and ${}^*d$ be $f[{}^*c]$. Here ${}^*c$ and $a \setminus {}^*c$ obviously partition $a$, whereas ${}^*d$ and $b \setminus {}^*d$ partition $b$. It is easy to see that the parts of the partition of $a$ are mapped, respectively, to the corresponding parts of the partition of $b$ by $f$ and the inverse of $g$. The Cantor-Bernstein Theorem is then obtained as a direct consequence of this central theorem with the bijection from Dedekind's proof. (If ${}^*c$ and ${}^*d$ are obtained by approximations from below, then one obtains Bernstein's proof with its associated zigzag diagram.)

Banach's perspective leads quite directly to the consideration of fixed-points in the proof of the Fundamental Lemma. We want to find a binary partition of $a$, such that $h^*$ is obtained as the union of $h$ restricted to one part of the partition and the identity on its second part, i.e., find a subset $x$ of $a$ that satisfies the constraints: (1) $h \restriction x : x \to h[x]$ is bijective, (2) $id : a \setminus x \to a \setminus x$ is bijective, (3) $\{x, a \setminus x\}$ partitions $a$, and (4) $\{h[x], a \setminus x\}$ partitions $d$. The first three conditions are satisfied by any (non-empty) subset $x$ of $a$. To verify condition (4), $x$ has to satisfy the equation

$$a \setminus x = d \setminus h[x]. \tag{i}$$

This equation can be rewritten as

$$x = a \setminus (d \setminus h[x]). \tag{ii}$$

We have from Equation (i) that $a = x \cup (d \setminus h[x])$. As $x$ and $d \setminus h[x]$ must be disjoint, we can subtract $d \setminus h[x]$ from both sides of this equation and obtain Equation (ii). The latter asks for a subset $x$ of $a$ that is a fixed-point of the function $m : \wp(a) \to \wp(a)$ given by $m(x) = a \setminus (d \setminus h[x])$.

The Knaster-Tarski *Fixed-point Theorem* guarantees the existence of fixed-points not only for this monotone $m$, but asserts in general: every monotone function $k$ on $\langle \wp(a), \subseteq \rangle$ has a least and greatest fixed-point. The fixed-point constructions can be related to Dedekind's and Zermelo's proofs as "joined" in our adaptation of König's proof. Given that

$$z \in (a \setminus (d \setminus h[x])) \leftrightarrow (z \in \overline{a} \vee z \in h[x]),$$

$m$'s definition can be reformulated as $m(x) = \overline{a} \cup h[x]$, where $x \subseteq a$ and $\overline{a} = a \setminus d$. Our version of König's proof involved also the chain $c^*$ of the system $d \setminus e$ and the set $r$ that was defined as $e \setminus (h[c] \cup h[c^*])$. Here

[6]Banach's central first theorem is this: If $f \in \mathrm{inj}(a, b), f[a] \subset b, g \in \mathrm{inj}(d, b), d \subset a, g[d] = b$, then there is a partition $\{a_1, a_2\}$ of $a$ and a partition $\{b_1, b_2\}$ of $b$, such that $f[a_1] = b_1$ and $g[a_2] = b_2$. Underlying the proof of this theorem is a precise version of König's argument for the Cantor-Bernstein Theorem.

is a crucial observation concerning $c$ and $(c \cup r)$: (1) $c$ is $m$'s smallest fixed-point, and (2) $(c \cup r)$ is $m$'s largest fixed-point.

How does the fixed-point construction affect the proof of the *Fundamental Lemma*? We adapt Dedekind's proof using the smallest fixed-point. Recall the formulation of the *Lemma*: Let $h$ be a bijection from $a$ to $e$ and let $d$ be a set such that $e \subseteq d \subseteq a$; then there is a bijection $h^*$ from $a$ to $d$. Here is the *second proof*: With the monotone function $m$ on $\langle \wp(a), \subseteq \rangle$, we define $c$ as

$$\bigcap \{z \subseteq a \mid m(z) \subseteq z\}.$$

Indeed, $c$ is a fixed-point of $m$, i.e., $c = m(c) = (a \setminus d) \cup h[c]$. In the first proof of the Fundamental Lemma, we used this structural identity to prove $a \setminus c = d \setminus h[c]$. That allows us also now to establish that $a$ is the disjoint union of $c$ and $(d \setminus h[c])$ and that $d$ is the disjoint union $h[c]$ and $(d \setminus h[c])$. Thus, we (re-) obtain $h^*$ as the union of the bijections $h_1(x) = h(x)$ for $x \in c$ and $h_2(x) = id(x)$ for $x \in a \setminus c = d \setminus h[c]$.

We take a final look at Zermelo's proof that yields the bijection $h^{**}$. We found $c$ and the associated partitions of $a$ and $d$ as the solution of the four constraints (1)–(4); we can similarly motivate $c^*$. Consider the constraints $(1')$ $h \restriction (a \setminus x) : a \setminus x \to h[a \setminus x]$ is bijective, $(2')$ $id : x \to x$ is bijective, $(3')$ $\{a \setminus x, x\}$ partitions $a$, and $(4')$ $\{h[a \setminus x], x\}$ partitions $d$. The first three conditions are again satisfied by any (non-empty) subset $x$ of $a$. To verify condition $(4')$, $x$ has to satisfy the equation $d \setminus x = e \setminus h[x]$, i.e., $x = d \setminus (e \setminus h[x]) = (d \setminus e) \cup h[x]$. As above, we can now define $m^* : \wp(a) \to \wp(a)$ by setting $m^*(x) = (d \setminus e) \cup h[x]$. We know $m^*$ is monotone and has $c^*$ as its smallest and $c^* \cup r$ as its largest fixed-point. Here is the *third proof* of the Fundamental Lemma. Let $c^*$ be

$$\bigcap \{z \subseteq a \mid m^*(z) \subseteq z\};$$

it is a subset of $a$ and satisfies

$$m^*(c^*) \subseteq c^*. \tag{iii}$$

Furthermore, $m^*(c^*)$ is not a proper subset of $c^*$, as the monotonicity of $m^*$ and (iii) imply that $m^*(m^*(c^*)) \subseteq m^*(c^*)$. So $m^*(c^*)$ is an element of $\{z \subseteq a \mid m^*(z) \subseteq z\}$ and $c^*$, as the intersection of all these sets, is a subset of $m^*(c^*)$. Thus, we have:

$$m^*(c^*) = c^*.$$

The remainder of the proof is now canonical.

These proofs give the high-level considerations of the different arguments for the Cantor-Bernstein Theorem; their crucial element is finding binary partitions of $a$ and $d$, such that the respective parts are equinumerous. Then the partition lemma yields a bijection from $a$ to $d$. Through our proofs we obtained exactly two such bijections, namely, Dedekind's

$h^*$ and Zermelo's $h^{**}$. In these considerations we have taken for granted a standard set theoretic frame. In *WZ*, the then very novel frame for the development leading up to the Fundamental Lemma is presented in great detail. We analyze this inconsistent frame in the next section. Our analysis is followed first by a brief description of the Zermelo axiom system and then by a more detailed discussion of our official deductive frame, the Zermelo-Fraenkel system ZF.

**§3. Formal frame and groundwork.** Dedekind used two principles concerning sets or, what he calls *systems*, namely, FULL COMPREHENSION and EXTENSIONALITY. The first principle is given by $(\exists y)(\forall x)(x \in y \leftrightarrow P(x, \dots))$. $P(x, \dots)$ is a property and allows us to view the objects falling under it "from a common perspective"; thus, Dedekind thought, they form a *system*. Two systems $a$ and $b$ are equal if they have the same elements, i.e., systems are considered as extensional.

$$(\forall x)(x \in a \leftrightarrow x \in b) \leftrightarrow a = b \qquad \text{(EXTENSIONALITY)}$$

These principles allowed Dedekind to define operations like arbitrary unions and intersections. He also introduced a second category of important mathematical entities, namely, *mappings* (Abbildungen) and articulated principles for them, but only implicitly.[7] For example, mappings can be composed, if their domains and co-domains fit appropriately; they can be inverted, if they witness the "similarity" of domain and co-domain or are, in modern terminology, bijections. He carefully investigated the set theoretic operations and their interaction with mappings; that is the basis for his chain theory. The core principle is of course inconsistent, as Russell's Paradox can be obtained in a most familiar way.

In his paper "Untersuchungen über die Grundlagen der Mengenlehre I" [Zermelo, 1908], Zermelo introduced an axiomatic theory in Dedekind and Hilbert's spirit. He replaced FULL COMPREHENSION by his SEPARATION PRINCIPLE, which *separates out* from a given set the elements satisfying a property expressed by a formula $\phi(x, \dots)$ of the language of first-order logic[8]:

$$(\exists y)(\forall x)(x \in y \leftrightarrow x \in a \ \&\ \phi(x, \dots)) \qquad \text{(AXIOM OF SEPARATION)}$$

Proper set existence principles supplement the separation principle, giving us particular sets to which SEPARATION can be applied. Zermelo's principles guarantee the existence of elementary sets (the empty set $\emptyset$, singletons $\{a\}$, and unordered pairs $\{a, b\}$), powersets and unions, but

[7] The important and very special role of mappings is brought out in [Sieg and Schlimm, 2014].

[8] Zermelo did not use formulae of the language of first-order logic to specify the separation principle, but rather a more informal concept of "definite property".

also the existence of an infinite set. He bypassed Dedekind's second category of mathematical entities by defining mappings as sets of *unordered* pairs. Such mappings are difficult to work with. So we are taking *functions* to be sets of *ordered* pairs, as is standard. The frame for our set theoretic development will be Zermelo-Fraenkel set theory ZF. Its axioms formulate the principles underlying the construction of the *cumulative hierarchy* of sets that was introduced and investigated in [Zermelo, 1930]. The axioms are described next.

The language of ZF is that of first-order logic with identity and the binary relation $\in$ for membership. Zermelo, following Dedekind, took sets as *extensional*. In ZF there is an additional axiom that specifies the general character of sets: they are not only extensional, but also *well-founded*. That is expressed by the AXIOM OF FOUNDATION:

$$(\exists y)\phi(y) \rightarrow (\exists y)(\phi(y) \ \& \ (\forall x \in y)\neg\phi(x)).$$

This axiom is easily seen to be equivalent to the principle of $\in$-INDUCTION:

$$(\forall y)((\forall x \in y)\phi(x) \rightarrow \phi(y)) \rightarrow (\forall y)\phi(y).$$

For the formulation of the first set existence principle of ZF it is convenient to use the defined notion of subset: $a \subseteq b \leftrightarrow (\forall x \in a)x \in b$. In the defining formula we used the *bounded universal* quantifier, where $(\forall x \in a)x \in b$ abbreviates $(\forall x)(x \in a \rightarrow x \in b)$.

$$(\exists y)(\forall x)(x \in y \leftrightarrow x \subseteq a). \qquad \text{(AXIOM OF POWERSET)}$$

The *bounded existentially* quantified formula $(\exists z \in a)x \in z$ abbreviates $(\exists z)(z \in a \ \& \ x \in z)$. That allows us to formulate the next axiom as follows:

$$(\exists y)(\forall x)(x \in y \leftrightarrow (\exists z \in a)x \in z)). \qquad \text{(AXIOM OF UNION)}$$

The AXIOM OF REPLACEMENT guarantees the existence of Zermelo's elementary sets and of all sets given by (instances of) separation. To formulate it easily we express "a relation $\phi(x, y)$ is single-valued on $a$" by $(\forall x \in a)(\forall y, z)((\phi(x, y)\&\phi(x, z)) \rightarrow y = z))$. This formula is abbreviated by $SV(\phi(x, y), a)$.

$$SV(\phi(x, y), a) \rightarrow (\exists z)(\forall y)(y \in z \leftrightarrow (\exists x \in a)\phi(x, y)) \qquad \text{(AXIOM OF REPLACEMENT)}$$

In sum, ZF has these axioms: extensionality, foundation, powerset, union, and replacement together with the axiom of infinity formulated below. The sets that exist according to the axiom of replacement as well as the powerset and union axioms are uniquely determined. So we introduce in a *definitional extension* of ZF *constants* and *set-theoretic operations*

through "defining axioms"; here are the axioms for the powerset and union operations:

$$(\forall x)(x \in \wp(a) \leftrightarrow x \subseteq a) \qquad \text{(POWERSET)}$$

$$(\forall x)(x \in \bigcup a \leftrightarrow (\exists z \in a)x \in z) \qquad \text{(UNION)}$$

We remarked that all instances of the separation principle are provable from replacement, so we introduce the most familiar set theoretic operation associated with the separation principle as follows:

$$(\forall x)(x \in \{y \in a \mid \phi(y, \dots)\} \leftrightarrow x \in a \;\&\; \phi(x, \dots)) \qquad \text{(SEPARATION)}$$

This principle can be generalized to allow separation from "constructed" sets, denoted by terms in the expanded language. Using these operations one obtains the BIG INTERSECTION of all the elements of a non-empty set a as well as the standard *Boolean operations*. Finally, we come to the AXIOM OF INFINITY; it postulates the existence of a set that contains the empty set and is closed under a successor operation; Zermelo considered $\{a\}$ as the successor of $a$, whereas von Neumann used $a \cup \{a\}$ as its successor.

$$(\exists v)(\emptyset \in v \;\&\; (\forall x \in v)\{x\} \in v) \qquad \text{(AXIOM OF INFINITY: ZERMELO)}$$

$$(\exists v)(\emptyset \in v \;\&\; (\forall x \in v)x \cup \{x\} \in v) \qquad \text{(AXIOM OF INFINITY: VON NEUMANN)}$$

We use von Neumann's formulation.

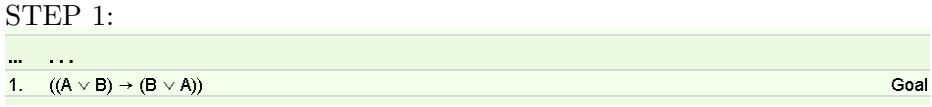
The logical consequences of these axioms — be they basic or definitional — are obtained by the inference rules of natural deduction, that is, by Gentzen's introduction (I-) and elimination (E-) rules for all the standard logical connectives. The rules are used in a strategic and bi-directional way "to close the gap between assumptions and claim"; the resulting *partial derivations* are represented not as trees, but as a *particular* kind of Fitch diagrams.[9] The diagrams indicate which logical steps have been taken and which subgoals are still to be proved. An efficient strategy suggests itself: use E-rules in the forward direction from the top down and I-rules in the backward direction from the bottom up. The resulting representation of "proofs with gaps" is motivated, on the one hand, by the bi-directionality of informal argumentation and, on the other hand, by features of *normal* natural deduction proofs. That explains yet another significant aspect of our rule set. Standard Fitch diagrams are constructed from the top down; thus, one needs an "assumption rule" that can open

[9]Minimal, intuitionist, and classical logic are distinguished through the treatment of negation. The underlying *intercalation calculi* were shown to be complete in [Sieg and Byrnes, 1998] for classical logic and in [Sieg and Cittadini, 2005] for intuitionist logic and some modal logics. Intercalation proofs are in each case uniquely associated with normal natural deduction proofs.

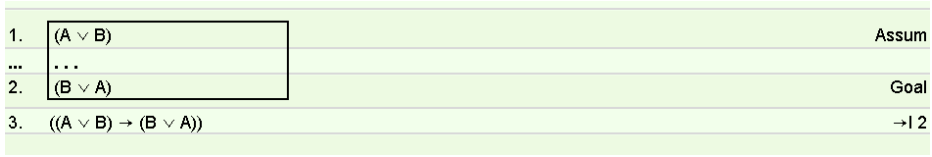
a subproof at any point with any assumption. In contrast, our calculus makes an assumption only through a rule in the context of its discharge, for example, through the backward application of an I-rule or the forward application of an E-rule.

Instead of presenting the logical calculus in full detail, we look at a simple example and prove $(A \vee B) \to (B \vee A)$ as a theorem of sentential logic.

STEP 1:




STEP 2:

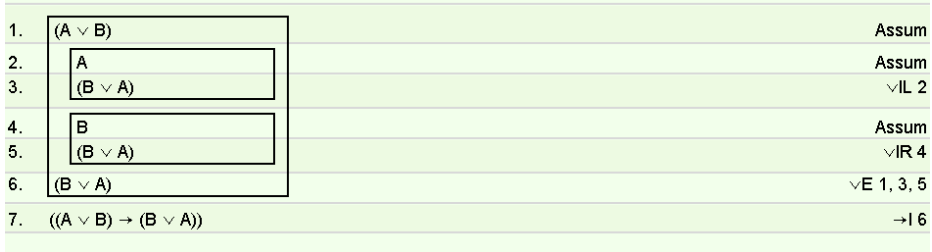



In the next step, we use the rule of $\vee$-Elimination to replace the single gap by a pair of gaps with additional assumptions; here we see that the calculus serves as a multiple-conclusion calculus.

STEP 3:


The next step consists of two steps, namely, completing the first and the second subderivation. The order of completion does not matter.

STEP 4:




*In sum*, constructing a proof is the stepwise local modification of partial Fitch diagrams by logical rules with the aim of filling gaps.[10]

[10] The sequence of such applications is strategically guided by the syntactic configuration and proof theoretic facts. The strategic choices have been implemented in the

There is one final remark on the logical formalism for set theory; it deals with the treatment of bounded universal and existential quantifiers, $(\forall x \in a)\phi(x)$ and $(\exists x \in a)\phi(x)$. The special rules for them avoid repetitive logical moves. Here are the formulations for the bounded universal quantifier. (For the bounded existential one they are analogous.)


The usual restrictive conditions have to be observed. For the elimination rule, the standard substitution restrictions have to be taken into account as well.


The above remarks and diagrams were to indicate briefly the special features of the logical calculus we are using.

There are two significant further expansions of the logical frame. They concern the treatment of definitions and the use of *lemmas-as-rules*. Recall, that the meaning of the logical connectives is captured by their E- and I-rules. The same can be done for definitions through E- and I-rules that are obtained from their defining biconditionals. Here are two examples. The first concerns the powerset operation with the defining axiom $(\forall x)(x \in \wp(a) \leftrightarrow x \subseteq a)$; its associated rules are:

$$\frac{x \subseteq a}{x \in \wp(a)}\ \wp\text{-I} \qquad \frac{x \in \wp(a)}{x \subseteq a}\ \wp\text{-E}$$

The second example presents the rules for the subset relation with the defining axiom $a \subseteq b \leftrightarrow (\forall x \in a)x \in b$:

$$\frac{(\forall x \in a)x \in b}{a \subseteq b}\ \subseteq\text{-I} \qquad \frac{a \subseteq b}{(\forall x \in a)(x \in b)}\ \subseteq\text{-E}$$

complete proof search mechanism AProS that finds natural proofs quite efficiently — in the three logics mentioned in the previous note.

The step from defining axioms to I- and E-rules is a special case of the distinctive and natural inference mechanism we call *lemmas-as-rules.*[11] Assume we have a lemma of the form $H_1, \ldots, H_n \vdash C$ or the conditional $H_1 \ \& \ \ldots \ \& \ H_n \rightarrow C$. In order to appeal to this lemma and to infer $C$ directly, it suffices to have proved the hypotheses $H_i$ on lines that are accessible from the position in the proof at which $C$ is to be concluded. (The Def-I and E-rules are obtained by this mechanism when applied to the defining biconditional in the appropriate direction.) All the important set theoretic operations and notions are introduced in definitional extensions of ZF. Their use and that of additional devices (to shorten proofs) is justified by metamathematical considerations: proofs (of statements in the original language) using them are easily transformed into proofs without them. *The necessary groundwork has been carried out formally with AProS as the checker and can be inspected under CBT at the AProS project website* `http://www.phil.cmu.edu/projects/AProS/`.

**§4. Conceptual hierarchy.** The system of set theory introduced by Zermelo in [Zermelo, 1908] was intended to show, "how the entire theory created by Cantor and Dedekind can be reduced to a few definitions and seven principles, or axioms, which appear to be mutually independent." In the last section we described an expanded frame for our formalization project: a definitional extension of ZF together with a flexible rule-based inferential mechanism. The latter includes not only I- and E-rules for the logical connectives, but also for defined notions. This mechanism is absolutely critical, if one wants to reflect mathematical practice and exploit the conceptual, hierarchical organization of parts of mathematics that are represented in set theory.[12] We consider the basic frame for our project we just described as level 0 of the hierarchy. This conservative extension of ZF can be further expanded to level 1, where relations and functions are introduced as set theoretic objects. That is in full harmony

[11] In Paulson's early work that uses set theory as a frame for verification, "natural deduction rules" are introduced for all the operations in a rather ad-hoc fashion, not through a uniform mechanism connecting the rules to defining axioms of definitional extensions. — In his manuscript [Hamami, 2016], Hamami attempts to make a principled distinction between mathematical and logical inferences. We think, however, that the "mathematical inferences" he discusses, for example, in section 3.3 of his manuscript, can all be recast as applications of our uniform lemmas-as-rules mechanism. He also discusses the proof of the irrationality of the square root of 2 in section 3.1; cf. note 23 below.

[12] [Wiedijk, 2011] pursues a search for intelligible proofs that are close to formal proofs, but also to the informal proofs of mathematical practice. Focused on the Mizar system, it is proposed to consider formal proof sketches that mimic informal proofs, but may be incorrect; correctness is only guaranteed by the full formalization. Formal proof sketches correspond, in our approach, to top-level proofs appealing to lemmas that have not yet been proved (and may be incorrect).

with Zermelo's view of set theory as "that branch of mathematics whose task is to investigate mathematically the fundamental notions 'number', 'order', and 'function', taking them in their pristine, simple form, and to develop thereby the logical foundations of all of arithmetic and analysis; thus it constitutes an indispensable component of the science of mathematics."[Zermelo, 1908, p. 261]

A little more than ten years later, Hilbert discussed in 1920 Zermelo's axiom system and claims that it is the "most comprehensive mathematical system". He supports that claim by a penetrating observation:

> The theory which results from the development of the consequences of this axiom system [Zermelo's] encompasses all mathematical theories (like number theory, analysis, geometry), in the sense that the relations which obtain between the objects of these mathematical disciplines are represented in a perfectly corresponding way by relations which obtain within a subdomain of Zermelo's set theory.[Hilbert, 2013, p. 292]

The first step for such a representation is taken next.

**Level 1: Relations and Functions.**[13] Here we formulate first the defining axiom for unordered pairs: $(\forall x)(x \in \{a, b\} \leftrightarrow x = a \vee x = b)$. Unordered pairs allow us to define ordered pairs $\langle a, b\rangle$ as $\{\{a\}, \{a, b\}\}$ and prove the *principal lemma* for them: $\langle a, b\rangle = \langle c, d\rangle \leftrightarrow a = c \;\&\; b = d$. As the last set theoretic operation we introduce the *Cartesian product* of two sets $a$ and $b$ with the defining axiom:

$$(\forall z)(z \in a \times b \leftrightarrow (\exists x \in a)(\exists y \in b)z = \langle x, y\rangle).$$

*Relations* between (elements of) $a$ and $b$ are subsets of the Cartesian product $a \times b$. Operations on relations like composition and inversion are most significant for functions. A function $f$ from $a$ to $b$, indicated by $f : a \to b$, is defined as a relation between $a$ and $b$ that satisfies single-valued-ness (*Sinval*) and totality (*Total*). Below we will introduce additional operations on functions.

We are of course considering special functions like injections, surjections, and bijections with their familiar definitions. We have a substantial number of lemmas concerning composition, inversion, restriction and co-restriction that operate on (special) functions; we consider also images $f[c]$, given $f : a \to b$ and $c \subseteq a$. In the first proof of the Fundamental Lemma we used the *partition lemma* for bijections with disjoint domains and co-domains to obtain a new bijection. In the discussion of further proofs we used *definition by cases* to obtain the new function. In each case we also have to "type" the newly defined function, i.e., indicate its

[13] In the implementation we actually pay a great deal of attention to sorting of sets and typing of functions. That will be discussed in a paper of Sieg, J. Ramsey, and T. Gibson on AProS.

domain and co-domain. After all, $f : a \to b$ is in general different from $f : a \to f[a]$. So, we introduce in addition to composition and inversion operations that modify only domain or co-domain, operations that are usually applied only implicitly. The most important is the *co-expansion*. Given a function $f : a \to b$ and $b \subseteq a$, the co-expansion $'f$ is the function $f$ but with co-domain $a$. Here are some trivial facts about co-domain expansions:

$$\langle x, y\rangle \in f \leftrightarrow \langle x, y\rangle \in {}'f;$$

$$d \subseteq a \to f[d] = {}'f[d];$$

$$f \in \mathrm{bij}(a, b)\ \&\ b \subseteq a \to {}'f \in \mathrm{inj}(a, a).$$

Two other operations will be used in the formal proofs, *restrictions* and *co-restrictions*. The restriction of $f : a \to b$ to a subset $d \subseteq a$ is denoted by $f \restriction d$, whereas its co-restriction $\restriction f$ is $f$ with domain $a$ and co-domain $f[a]$. Here is the important fact:

$$f \in \mathrm{inj}(a, b)\ \&\ d \subseteq a \to \restriction(f \restriction d) \in \mathrm{bij}(d, f[d]).$$

That means, if $f$ is injective, then one can always obtain a bijection between a subset of its domain and that subset's $f$-image. This completes the foundational work that supports the natural formalization of the informal mathematical development in section 2.

**Level 2: Finitary inductive definitions.** Such definitions can be given in a more general way than just as minimal chains of a system $b$ determined by the data $a$, $b \subseteq a$, $h : a \to a$: one can have a number of "generating" functions $h_1 : a \to a$, $h_2 : a \times a \to a$, etc. The inductive generation of the formulae of a formal language is a well-known example. However, even in this general case the inductive definitions can be made explicit on the model of Dedekind's chain of a system, namely, as the intersection of all sets that are closed under the finitely many operations.[14] For the simple case with one generating function we formally verified, for

[14]For this to be feasible within ZF, one has to show the existence of at least one set $a$ that contains the basic elements from $b$ and is closed under the generating operations, say $f_1$ and $f_2$. This can be achieved by generalizing Dedekind's construction that underlies his proof of theorem 131 in *WZ* (and was discussed above for the "approximation from below" in section 2). First, one has to decide which set theoretic operations allow the construction of the intended objects; for example, binary trees generated by a pairing operation from $\{\emptyset\}$. Second, one obtains with the help of replacement a sequence $\langle d_i \mid i \in \mathbb{N}\rangle$, the elements of which are defined recursively by $d_0 = \{\emptyset\}$ and $d_{i+1} = d_i \cup d_i \times d_i$. Then the union axioms ensures the existence of $\bigcup\{d_i \mid i \in \mathbb{N}\}$, the set $a$ of all binary trees with the empty set at their "roots". Third, one defines the generating functions on this $a$ for those binary trees that are to serve as (set theoretic representatives of) formulae. Finally, the set of formulae is obtained as the chain of the "atomic formulae" with these generating functions. One would proceed similarly, if formulae were viewed as finite sequences.

example, the theorem that the chain $c$ of the system $b$ given $h$ is indeed $h$-closed and minimal. We now present a formal proof of the crucial *Structural Lact*: the chain $c$ of the system $b$, given the above data, consists of the elements of $b$ and the $h$-values of $c$'s elements, i.e., $c = b \cup h[c]$. In the formalization we denote $c(a, b, h)$ by $\mathfrak{c}(a, b, h)$.

The proof of the Structural Lemma takes as a first step the representation of the claim through a partial derivation, adds in a second step a theorem in line 3, and then applies $\exists$-elimination to obtain:

| | | |
|---|---|---|
| 1. | h∈func(a,a) | Prem |
| 2. | b⊆a | Prem |
| 3. | (∃x)x = ¢(a,b,h) | Theorem (Chain28) 2, 1 |
| 4. | $z_1$ = ¢(a,b,h) | Assum |
| ... | . . . | |
| 5. | $z_1$ = b ∪ h[$z_1$] | Goal |
| 6. | ¢(a,b,h) = b ∪ h[¢(a,b,h)] | =E 4, 5 |
| 7. | ¢(a,b,h) = b ∪ h[¢(a,b,h)] | ∃E 3, 6 |

FIGURE 2. Pre-processing of the Structural Lemma.

The "pre-processing" with the Theorem and $\exists$-elimination is taken so that we can use the parameter $z_1$ as a name for the term $\mathfrak{c}(a, b, h)$ and have a more readable proof. Here is the description of the strategic proof construction and the completed proof. (The theorems appealed to are listed in Appendix A.)

To prove the set theoretic identity in line 20, it is sufficient to establish the subset relations in lines 9 and 19. (From now on, we will refer to lines simply by their number.) The first claim follows directly from the observations that the chain $z_1$ of $b$ with respect to $h$ is closed under $h$ (7) and contains $b$ (8). For 19 one appeals to the fact that $z_1$ is the minimal chain. So the central part of the proof has to establish (i) Chain($b \cup h[z_1], h$), (ii) $b \subseteq b \cup h[z_1]$, and (iii) $b \cup h[z_1] \subseteq a$. These three facts are established by 10–17, 5, and 6, respectively.

| | | |
|---|---|---|
| 1. | $h \in func(a,a)$ | Prem |
| 2. | $b \subseteq a$ | Prem |
| 3. | $(\exists x)x = \phi(a,b,h)$ | Theorem (Chain28) 2, 1 |
| 4. | $z_1 = \phi(a,b,h)$ | Assum |
| 5. | $b \subseteq b \cup h[z_1]$ | Theorem (Bool1) |
| 6. | $z_1 \subseteq a$ | Theorem (Chain8.2) 4, 2, 1 |
| 7. | $h[z_1] \subseteq z_1$ | Theorem (Chain11.2) 4, 2, 1 |
| 8. | $b \subseteq z_1$ | Theorem (Chain10.2) 4, 2, 1 |
| 9. | $b \cup h[z_1] \subseteq z_1$ | Theorem (Bool6) 8, 7 |
| 10. | $h[h[z_1]] \subseteq h[z_1]$ | Theorem (Func12) 1, 7, 6 |
| 11. | $h[b] \subseteq h[z_1]$ | Theorem (Chain14.2) 4, 2, 1 |
| 12. | $h[b] \cup h[h[z_1]] \subseteq h[z_1]$ | Theorem (Bool6) 11, 10 |
| 13. | $h[z_1] \subseteq a$ | Theorem (Func10) 1, 6 |
| 14. | $h[b] \cup h[h[z_1]] = h[b \cup h[z_1]]$ | Theorem (Func13) 1, 2, 13 |
| 15. | $h[b \cup h[z_1]] \subseteq h[z_1]$ | =E 14, 12 |
| 16. | $h[b \cup h[z_1]] \subseteq b \cup h[z_1]$ | Theorem (Bool4) 15 |
| 17. | $Chain(b \cup h[z_1],h)$ | Defl (Chain) 16 |
| 18. | $b \cup h[z_1] \subseteq a$ | Theorem (Mem6) 9, 6 |
| 19. | $z_1 \subseteq b \cup h[z_1]$ | Theorem (Chain12.2) 4, 17, 5, 18, 1 |
| 20. | $z_1 = b \cup h[z_1]$ | Theorem (Mem4) 19, 9 |
| 21. | $\phi(a,b,h) = b \cup h[\phi(a,b,h)]$ | =E 4, 20 |
| 22. | $\phi(a,b,h) = b \cup h[\phi(a,b,h)]$ | ∃E 3, 21 |

FIGURE 3. Proof of the Structural Lemma

**Remark.** This is a *general* structural fact: we obtain later specific ones for the Dedekind and the Zermelo constructions by instantiating $b$ to $a \setminus d$ and $d \setminus e$, respectively.

The proof of this *Structural Lemma* is complex, if the proofs of all the supporting lemmas are taken into account. The minimal $h$-closed set $c$ can be approximated from below,[15] and it can also be obtained as the least fixed-point of a monotone operator. In the latter case, as we saw

[15]This case is, however, special on account of the fact that the natural numbers are taken for granted.

in section 2, the Structural Lemma is an "immediate" consequence of the general Fixed-Point Theorem.[16] However, we considered a particular monotone function $m$ from $\wp(a)$ to $\wp(a)$ that is canonically defined from the given $h$ and $b \subseteq a$ by $m(z) = b \cup h[z]$. When showing that $c$ is equal to the smallest fixed-point of this $m$, the proof of the fixed-point property $m(c) = c$ just *is* the proof of the structural fact. After all, $c$ is $c(a, b, h)$ and is defined as the intersection of all subsets $z$ of $a$ satisfying the condition $b \subseteq z \ \& \ (\forall x \in z)h(x) \in z$. How is the minimal fixed-point $\mathrm{fix}(a, b, h)$ given for the above $m$? It is defined as the intersection of all subsets $z$ of $a$ satisfying the condition $b \cup h[z] \subseteq z$; this is obviously equivalent to the defining condition for $c(a, b, h)$.

No matter in which way $c$ is obtained, its minimality guarantees the principle of *proof by induction*:

$$(\forall x \in b)\psi(x) \ \& \ (\forall x \in c)(\psi(x) \to \psi(h(x))) \to (\forall x \in c)\psi(x).$$

When the closure operations are injective they yield "deterministic" elementary inductive definitions. For deterministic inductive definitions Dedekind proved, in the special case of natural numbers, the principle of *definition by (structural) recursion*. (See case (i) under "Examples and challenges" at level 3.)

**Level 3: Structural definitions.** A distinctive feature of modern mathematics is the definition of "abstract concepts" that impose a structure on systems of mathematical objects, i.e., specify relationships between the elements of such systems independently of the nature of these objects. One pertinent question asks, what is the connection between different systems falling under the same abstract concept. Here we give two examples. In the first example we formulate a representation theorem that, in turn, implies the categoricity of the concept. The second example serves to showcase a non-categorical concept. In both cases we state theorems that can be proved directly in the informal mathematical way — having available the set theoretic apparatus we have implemented. Thus, the formalization of these theorems can be tackled.

**Examples and challenges.**

i. *Simply infinite systems.* This was Dedekind's name for the abstract concept characterizing natural numbers up to isomorphism. The defining or characteristic conditions of this notion are best known as the Dedekind Peano Axioms. — The axiom of infinity guarantees in ZF the existence of an "inductive set", i.e., a set $v$ that contains zero (the empty set) and

[16]The fixed-point approach to inductive definitions was taken in [Paulson, 1994] to formulate "packages to associate I- and E-rules for (co-) inductive definitions". Packages and examples are described, but neither proofs of the supporting theorems nor proofs involving the inductively defined notions are presented. The advantages Paulson lists for his approach on p. 148 also accrue to our treatment.

that is closed under von Neumann's successor operation. Restricting the successor operation to $v$, we have a successor function *suc* and can define "the" natural numbers $\mathbb{N}$ as the chain of the system $\{\emptyset\}$ given the function *suc* from $v$ to $v$. We have the induction principle for $\mathbb{N}$ and can take on the challenge of formalizing the proofs of three theorems: (1) the Recursion Theorem (proving the existence of set theoretic functions that satisfy the schema of primitive recursion), (2) the Representation Theorem (implying that all simply infinite systems are isomorphic), and (3) various forms of the Pigeonhole Principle.

ii. *Groups.* This structural definition goes back to the 19th century as a paradigm of an "algebraic structure" in Bourbaki's terminology. There are two important, though elementary, facts that can be taken on as challenges for the formalization of their proofs: (1) Cayley's Representation Theorem and (2) the Isomorphism Theorem. Zipperer formally verified the latter theorem with the proof assistant Lean in his MS Thesis [Zipperer, 2016]. Much of his work was devoted to reconstructing set theory within the higher type theory that underlies Lean. We conjecture that it is straightforward to formalize a proof of the theorem within AProS.

But let us return from what is now imagined to be (easily) formalizable to the real task at hand. We saw in section 2, how chains of a system and fixed-points of a monotone function are crucial for characterizing the set that is obtained from an initial one by finitely iterating the application of a function. That, in turn, is crucial for the proofs of Dedekind's *Fundamental Lemma* from which the Cantor-Bernstein Theorem follows quite directly.

**§5. The formalized proofs.** We are going to present, finally, the AProS-certified proofs of the Cantor-Bernstein Theorem and of Dedekind's *Fundamental Lemma*. The goal of our formalization work has been the same over many years, namely, to give mathematically perspicuous and conceptually well-organized derivations that reflect the structure of deeply analyzed proofs. Our methodological guide has been the other *Dirichlet Principle* that asks us to overcome problems "with a minimum of blind calculation and a maximum of perceptive thought". That meant for us bringing out the leading idea of each proof and to articulate as auxiliary lemmas the "technical" steps that are immediate and mathematically meaningful at the lower level. Note again that the auxiliary lemmas that are used in the formal proofs are listed in Appendix A.

As we saw in section 2, the *Fundamental Lemma* yields the Cantor-Bernstein Theorem. Here is the formal proof of the latter from the former. The informal proof is seen as guiding the construction of the derivation.

| | | |
|---|---|---|
| 1. | f∈inj(a,b) | Prem |
| 2. | g∈inj(b,a) | Prem |
| 3. | 1(g ∘ f)∈bij(a,g ∘ f[a]) | Theorem (Core12) 1, 2 |
| 4. | g[b]⊆a | Theorem (Func17) 2 |
| 5. | g ∘ f[a]⊆g[b] | Theorem (Comp11) 1, 2 |
| 6. | a≈g[b] | Theorem (Fundamental Lemma) 3, 4, 5 |
| 7. | b≈g[b] | Theorem (Equi4) 2 |
| 8. | a≈b | Theorem (Equi8) 6, 7 |

FIGURE 4. Proof of the Cantor-Bernstein Theorem

We focus now on the proof of the Fundamental Lemma; in it, two central facts are used: the *Partition Lemma* (for obtaining a bijection from bijections on the parts of the partitions) and the *Structural Lemma* that is crucial to obtain the two different pairs of partitions of $a$ and $d$ discussed in section 2. With these facts firmly in the background, the informal proof is the blueprint for the formal proof: it gives the "directions" for the construction steps. Here is the sequence of line numbers in the order in which they are inserted: 15; backward with the partition theorem yielding the subgoals in 8, 9, 12, 14; 7 as justification for 8; 9 justified from 6, 3, 2, 1; 13 justifies 14; 10 and 11 justify 12.

Below is the formal proof of the *Fundamental Lemma*; please do not follow the proof in its pure linear order, but rather in the order indicated by the above sequence. The Zermelo construction can be formally be obtained as well — from the appropriate second pair of partitions we discussed earlier. The proof of the Fundamental Lemma will be followed by the proof of the underlying Dedekind partitioning. For our purposes, we use the predicate $\mathbf{Binpart}(a, b, c)$ which is defined as $\mathbf{Binpart}(a, b, c) \leftrightarrow a \subseteq c \;\&\; c \setminus a = b$.

| | | |
|---|---|---|
| 1. | h∈bij(a,e) | Prem |
| 2. | d⊆a | Prem |
| 3. | e⊆d | Prem |
| 4. | e⊆a | Theorem (Mem6) 3, 2 |
| 5. | (∃x)x = ¢(a,a \ d,'h) | Theorem (Chain26) 1, 2, 3 |
| 6. | $z_1$ = ¢(a,a \ d,'h) | Assum |
| 7. | $z_1$⊆a | Theorem (CBT_D1) 6, 3, 2, 1 |
| 8. | Binpart(a \ $z_1$,$z_1$,a) | Theorem (Bool21) 7 |
| 9. | Binpart(a \ $z_1$,'h[$z_1$],d) | Theorem (CBT_D4) 6, 3, 2, 1 |
| 10. | 'h∈inj(a,a) | Theorem (CE4) 1, 4 |
| 11. | ↾('h ↾ $z_1$)∈bij($z_1$,'h[$z_1$]) | Theorem (Equi2) 10, 7 |
| 12. | (∃f)f∈bij($z_1$,'h[$z_1$]) | ∃I 11 |
| 13. | $_1$(a \ $z_1$)∈bij(a \ $z_1$,a \ $z_1$) | Theorem (Id8) |
| 14. | (∃f)f∈bij(a \ $z_1$,a \ $z_1$) | ∃I 13 |
| 15. | a≈d | Theorem (FUn10) 9, 8, 12, 14 |
| 16. | a≈d | ∃E 5, 15 |

FIGURE 5. Proof of the Fundamental Lemma.

Here is the proof of the Dedekind partitioning with the crucial use of the Structural Lemma (Chain 24) in 12:

| | | |
|---|---|---|
| 1. | h∈bij(a,e) | Prem |
| 2. | d⊆a | Prem |
| 3. | e⊆d | Prem |
| 4. | z = ¢(a,a \ d,'h) | Prem |
| 5. | d \ 'h[z] = ((a \ d) ∪ d) \ ((a \ d) ∪ 'h[z]) | Theorem (Bool15) |
| 6. | a \ d⊆a | Theorem (Bool9) |
| 7. | a \ z⊆d | Theorem (CBT_D3) 4, 3, 2, 1 |
| 8. | (a \ d) ∪ d = a | Theorem (Bool11) 2 |
| 9. | d \ 'h[z] = a \ ((a \ d) ∪ 'h[z]) | =E 8, 5 |
| 10. | e⊆a | Theorem (Mem6) 3, 2 |
| 11. | 'h∈func(a,a) | Theorem (CE5) 1, 10 |
| 12. | (a \ d) ∪ 'h[z] = z | Theorem (Chain24.2) 4, 6, 11 |
| 13. | 'h[z]⊆d | Theorem (CBT_D2) 4, 3, 2, 1 |
| 14. | d \ 'h[z] = a \ z | =E 12, 9 |
| 15. | d \ (a \ z) = 'h[z] | Theorem (Bool22) 14, 13 |
| 16. | (a \ z⊆d & d \ (a \ z) = 'h[z]) | &I 7, 15 |
| 17. | Binpart(a \ z,'h[z],d) | DefI (Binary Partition) 16 |

FIGURE 6. Proof of the Dedekind partition of $d$.

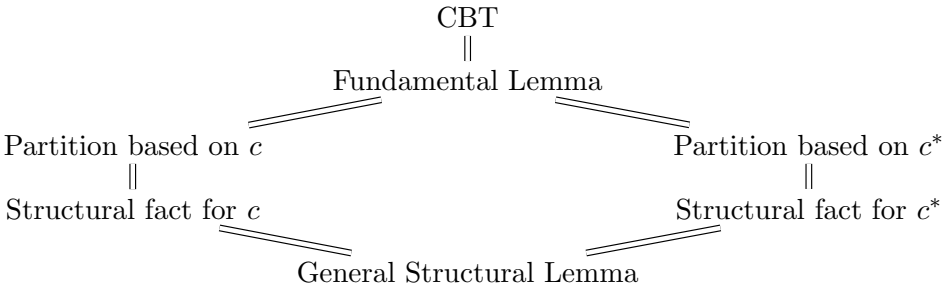
We followed Dedekind's argument and used in particular his way of explicitly defining "from above" sets that are inductively generated. In section 2 we compared Zermelo's proof with Dedekind's. The difference between their proofs lies in the choice of the initial set on which $h$ is being iterated: Dedekind uses $a \setminus d$, whereas Zermelo starts with $d \setminus e$. This leads to exactly two partitions and two canonical bijections $h^*$ and $h^{**}$. The proofs of the Structural Lemmas proceed in exactly the same way, whether we are considering chains of the sets $a \setminus d$ or $d \setminus e$ with respect to $h$: they are instances of the proof of the (general) Structural Lemma we gave above. In section 2 we also indicated how the consideration for the Structural Lemmas can be carried out via fixed-point constructions with the canonically defined monotone operations $m$ and $m^*$. The proofs are almost exactly the same as for Dedekind's or Zermelo's version with one appeal to an additional Lemma that asserts the identity of the minimal chains $c$ and $c^*$ with the minimal fixed-points of $m$ and $m^*$.

| | | |
|---|---|---|
| 1. | h∈func(a,a) | Prem |
| 2. | b⊆a | Prem |
| 3. | fix(a,b,h) = ∩({z∈℘(a)\|b ∪ h[z]⊆z}) | Prem |
| 4. | fix(a,b,h) = ¢(a,b,h) | Theorem (CBT_FP2) 3, 2, 1 |
| 5. | (∃x)x = ¢(a,b,h) | ∃I 4 |
| 6. | $z_1$ = ¢(a,b,h) | Assum |
| ... | . . . | |
| 7. | $z_1$ = b ∪ h[$z_1$] | Goal |
| 8. | ¢(a,b,h) = fix(a,b,h) | Sym= 4 |
| 9. | ¢(a,b,h) = b ∪ h[¢(a,b,h)] | =E 6, 7 |
| 10. | fix(a,b,h) = b ∪ h[fix(a,b,h)] | =E 9, 8 |
| 11. | fix(a,b,h) = b ∪ h[fix(a,b,h)] | ∃E 5, 10 |

FIGURE 7. Set-up for the fixed-point version of the structural fact.

In this partial derivation we gave only the "pre-processing" for the fixed-point version of the structural fact. The gap between 1–6 and the goal $z_1 = b \cup h[z_1]$ is literally filled by lines 5–19 of the earlier proof (adjusting the line references for the justifications).

The arguments, first from the Structural Lemmas to the two ways of partitioning the sets $a$ and $d$, then to the Fundamental Lemma and, finally, from the latter to CBT are fully fixed. As the proofs of the Structural Lemmas (for the two different ways of making an inductively generated set explicit without appealing to $\mathbb{N}$) are in each case instances of the same proof, one might be justified in claiming that there is *essentially one proof* of the Cantor-Bernstein Theorem given the choice of the subsets on which to finitely iterate $h$:



The seeming multitude of proofs of the Cantor-Bernstein Theorem has been reduced by analyzing crucial concepts and related techniques, so that we understand the identical structure of some proofs and evaluate differences between others. Here we have an example of "proof theoretic" investigations that are quite different from the standard ones in pursuit of (modified) Hilbert Programs. However, the issue itself can be traced back to Hilbert. Thiele, in [Thiele, 2003, 2005], found what was supposed to be the 24th problem of Hilbert's Paris talk in 1900. It was asking for "Criteria of simplicity, or proof of the greatest simplicity of certain proofs." Hilbert's manuscript continues:

> Develop a theory of the method of proof in mathematics in general. Under a given set of conditions there can be but one simplest proof. Quite generally, if there are two proofs for a theorem, you must keep going until you have derived each from the other, or until it becomes quite evident what variant conditions (and aids) have been used in the two proofs.

This is a more specific aspect of what Hilbert called for in his [Hilbert, 1918], namely, a theory of "the specifically mathematical proof". We think we have taken, with our mathematical analysis and the strictly formal development of proofs of the Cantor-Bernstein Theorem, a step in the direction Hilbert pointed to.

However, we have to realize that we have been using formalization in its strict form that was not yet in Hilbert's chest of tools, neither in 1900 nor in 1917. So let us reflect on the emergence of the very idea and contrast it with that of *natural formalization*.

**§6. Formalization's historical roots.** From the programmatic perspective of formalizing mathematics with the support of computers[17] one understands *groundedness* and *correctness* of these considerations in purely syntactic terms: the global starting principles for proofs are [the] axioms [of ZF] formulated in a formal language; the basic inferential principles are those of a (correct) logical calculus. This broad understanding conforms to a long mathematical and logical tradition. Indeed, proponents of the formalization of mathematics at the turn from the 19th to

[17]The very close and deep connection between, on the one hand, ordinary human and, on the other hand, computer-supported formalization is clear from Turing's 1936 [Turing, 1936] analysis of "mechanical procedures" that are carried out by humans. The use of his digital computing machines for mathematical work was clearly initiated by Turing himself; see, for example, [Turing, 1948a] and [Turing, 1948b]. — In [Wiedijk, 2008, p. 1414] it is asserted that the "revolution" to formal mathematics took place "in the late twentieth and early twenty first centuries"; the timeframe for these developments is more correctly given by a shift of almost exactly one century locating the revolution in "the late nineteenth and early twentieth centuries".

the 20th century viewed it in just this way.[18] Consider, for example, Frege's remarks on "the ideal of a strictly scientific method in mathematics", when he compares his way of developing a theory with Euclid's:

> It cannot be required that we should prove everything, because that is impossible; but we can that all propositions used without proof should be expressly mentioned as such ... Furthermore I demand — and in this I go beyond Euclid — that all methods of inference used must be mentioned in advance. [Frege, 1980, p. 117]

The core difference, as Frege saw it, is the appeal to explicitly formulated logical principles that allow us to take inferential steps and achieve "gapless" proofs. However, to avoid an infinite regress, these steps must be of an elementary kind and allow for mechanical or algorithmic checkability; the latter is to draw on only minimal cognitive resources. Frege claimed that in his logical system "inference is conducted like a calculation" and observed:

> I do not mean this in a narrow sense, as if it were subject to an algorithm the same as ... ordinary addition or multiplication, but only in the sense that there is an algorithm at all, i.e., a totality of rules which governs the transition from one sentence or from two sentences to a new one in such a way that nothing happens except in conformity with these rules. [Frege, 1984, p. 237]

These observations apply also to Whitehead and Russell's *Principia Mathematica*.[19] For the Hilbert School during the 1920s, when the finitist consistency program was being pursued, the in-principle formalizability of mathematics either was taken for granted, was thought to be carried out easily, or was justified by reference to *Principia Mathematica*.

In a lecture given to the Mathematical Association of America in December of 1933, Gödel viewed it as an "important fact that all of mathematics can be reduced to a few formal axioms and rules of inference". He expressed the Fregean requirement on inference rules as follows:

> [T]he outstanding feature of the rules of inference being that they are purely formal, i.e., refer only to the outward structure of the formulas, not to their meaning, so that they could be applied by someone who knew nothing about mathematics, or by a machine. [Gödel, 1933, p. 45]

[18] Our focus is restricted to the logical setting of first-order logic. That restriction evolved, however, only significantly later beginning in the 1930s.

[19] It was motivating for Emil Post when he sought a decision procedure for all of *Principia Mathematica* in the early 1920s; see [Sieg, Szabó, and McLaughlin, 2017]

The "formal" character of the rules had become by then a topic of intense discussion, as a proper mathematical characterization was needed — for the general formulation and proof of the Incompleteness Theorems Gödel had discovered in late 1930. Church and Turing gave such a characterization a few years later in their pursuit of a negative solution to the *Entscheidungsproblem* for first-order logic. In particular, Turing based his famous machine characterization on an analysis of *mechanical procedures* carried out by humans.

Referring back to Turing's analysis Gödel wrote in 1964 [Gödel, 1964], "A formal system can simply be defined to be any mechanical procedure for producing formulas, called provable formulas." He emphasized that it is of the essence of the concept of a formal system "that reasoning is completely replaced by mechanical operations on formulas". Some have argued that the strict formalization of sophisticated theories is not possible for us humans: Bourbaki claimed this for mathematical theories, whereas Suppes and, in the strongest possible way, Stegmüller asserted this for physical theories.[20] That seems to be right, if all theorems have to be formulated in the basic language and all proofs have to appeal only to the fundamental rules and axioms of the theory. However, such requirements should not be imposed, as they are in deep conflict with axiomatic developments ever since Euclid. At this point the central tools of natural formalization come in.

We constructed an *inferential mechanism* and a hierarchical conceptual structure that allowed us to formally parallel arguments of mathematical practice. The conceptual organization for our case study is grounded in ZF as the *methodological framework*[21], but we reemphasize that other frameworks could be used as well. *Practical and humanly intelligible* formalization is made feasible by (1) analyzing mathematical practice in detail, (2) mimicking the analyzed practice with salient logical tools, and (3) embedding the local proofs in a conceptually well-structured deductive framework. (2) and (3) provide tools for the investigation of "the concept of the specifically mathematical proof" underlying (1). The eventual

[20] See [Bourbaki, 2004, p. 9], [Suppes, 2002, p. 27] and [Stegmüller, 1979, pp. 4–7]. In Bourbaki one finds this remark: "[N]o great experience is necessary to perceive that such a project [of full formalization] is absolutely unrealizable: the tiniest proof at the beginning of the *Theory of Sets* would already require hundreds of signs for its complete formalization. ... Formalized mathematics cannot be written down in full, ... We shall therefore very quickly abandon formalized mathematics." Stegmüller actually formulates in his [Stegmüller, 1979, p. 5] a thesis that formal axiomatics "is not humanly possible", whereas structural axiomatics can be carried out; he views this difference as "a purely practical or psychological one" [Stegmüller, 1979, p. 6].

[21] "Methodological framework" is pointing to Bernays' concept of "Methodischer Rahmen" that is to provide, by suitable construction procedures, axiomatically characterized domains in which mathematical practice can be represented; cf. [Sieg, 1997].

touchstone for the success of such an investigation will be its use as the basis for a *fully automated proof search* in the style of a mathematician — with logical sensibilities. The steps we have taken up to now, when going beyond logic, is to turn AProS into a supportive *Proof Assistant* that checks and helps to create (new) mathematical work, while being guided by its user-mathematician

**§7. Human-centered techniques and more.** Barendregt and Wiedijk viewed, in [Barendregt and Wiedijk, 2005], developing a proof assistant as a real challenge. In their joint paper, they asserted then that no system was useful for this purpose. They suggested that the situation would change in the "next few decades". For such a change to happen, some conditions have to be satisfied in their view: (i) proofs have to be presented in a mathematical style, (ii) an extensive library of definitions and theorems must be available, (iii) efficient decision procedures for particular problems can be appealed to and, crucially, (iv) the system has to provide "support for reasoning with gaps".[22]

Point (iv) is elaborated with references to *formal sketches* [Wiedijk, 2011], *proof planning* [Bundy, 1991], and *proof refinement* as proposed in [Lamport, 1995]. As to the latter they write [Barendregt and Wiedijk, 2005, p. 2370]:

> In Lamport (1995) a proof style is described in which proofs are incrementally developed by *refining* steps in the proof into more detailed steps. Although that paper does not talk about proofs in the computer, and although we are not sure that the specific proof display format that is advocated in that paper is optimal, it is clear that this *style of working* should be supported by

[22] [Barendregt and Wiedijk, 2005, pp. 2368–70]. The significance of points (i) – (iii) is also asserted in [Harrison, 2008, p. 1405]. Harrison's book [Harrison, 2009] has an informative chapter entitled "Interactive theorem proving"; this chapter is also interesting for its quick dismissal of natural deduction calculi in spite of the fact that they are viewed (p. 472) as "relatively good for formalizing typical human proofs". Barendregt and Wiedijk give on p. 2370 the following reason for the importance of condition (iv) on a computer system: "If one just wants to use a Proof Assistant to order one's thought, or to communicate something to another mathematician, then fully working out all proofs is just not practical. In that case one would like to just give a sketch of the proof inside the formal system, as described in Wiedijk (2004)." Wiedijk, on p. 379 of the 2004 paper quoted, attributes to Peter Aczel the claim "that in order to get mathematicians involved with the formalization of mathematics, technology is needed for reasoning with gaps, where one can leave out the details of a formalization that one considers to be obvious or well-known, and one only needs to formalize the interesting parts." Wiedijk asserts, "This vision is the focus of this paper." The vision is realized, at least in part, through our approach; see note 12 and the articulation of proof sketches as partial proofs.

> systems for computer mathematics, in order to be accepted by the mathematical community.

Gowers also emphasizes in his interview [Diaz-Lopez, 2016] that mathematicians often use a *top-down approach*. "In such an approach, you have some vague plan for how it works and you try to put in the details and some things work well and other things don't work and you have to modify them and so on." He continued, "Getting all that to work on a computer is an extremely ambitious goal, obviously." This perspective is presented in his paper with Ganesalingam [Ganesalingam and Gowers, 2013] in most informative detail. We discuss their general perspective and solve a problem of theirs with our approach in Appendix B; that particular example shows a genuine parallelism between their work and ours.

Let us go back from Gowers' *top-down approach* to the related, but more specific *reasoning with gaps* that Barendregt and Wiedijk stress. They claim, as we saw, "It is clear that *this style of working* should be supported for computer mathematics, in order to be accepted by the mathematical community." *This style of working* refers back to Lamport's proof presentation that is viewed as a "refinement of natural deduction" [Lamport, 1995, p. 600]. As a matter of fact, Lamport's proofs are organized like standard *Fitch diagrams*: their hierarchical structure is not indicated graphically through the containment relation of boxes, but by numerical indices that reflect exactly that relation; the rules, including an unrestricted assumption rule, are applied only in the forward direction. The refinement steps are taken on *completed* derivations in which lemmas have been appealed to as rules, and this appeal is replaced by a derivation of the lemma. Such replacements lead ultimately to a single derivation that uses only basic rules and will in general be unsurveyable.

The intercalation calculus implemented in AProS is a version of natural deduction. However, it does not generate derivations of a formula from premises and assumptions, but rather *partial derivations*; cf. the examples of proofs in sections 4 and 5. Partial derivations are *formal derivations with gaps* indicating their subgoals together with the premises and assumptions that are directly accessible from a particular subgoal. Gaps can be closed by forward (and always goal-directed) or backward applications of logical *and* definitional rules as well as the mechanism of lemmas-as-rules; the goal directedness of forward moves is crucial for *extraction*, that combines several elimination moves. The intercalation calculus supports "reasoning with gaps" in the most direct way, since the *bi-directional* way of constructing proofs is its essence. This is also the opening for efficient automated proof search that systematically seeks to connect goals,

premises and assumptions, exploiting uniform logical strategies as well as specific heuristics for particular parts of mathematics.[23]

In a certain way we are trying to bring to life Hilbert's perspective when he claims in [Hilbert, 1927], "The fundamental idea of my proof theory is none other than to describe the activity of our understanding, to make a protocol of the rules according to which our thinking actually proceeds." The rules Hilbert alludes to cannot just be those of logic, but must include conceptual and heuristic ones. The "fundamental idea" was articulated in a complementary way by one of the last logic students in Göttingen, Saunders Mac Lane: to uncover the "highly rational structure" of mathematical proofs. That was done in the hope that this rational structure would allow shaping sequences of rule applications and finding humanly intelligible proofs.[24] Eighty years later and after a breathtaking development of computational power, we can go beyond the mathematical analysis and formal representation of proofs. We can implement suitable procedures and run computer experiments, observing how strategic, humanly motivated modifications affect search.

This direction of our efforts is related to Gowers' "extreme human-oriented automatic theorem-proving" [Diaz-Lopez, 2016]. He thinks, as

[23]Lamport discusses [Lamport, 1995, pp. 601–604] in great detail the proof of the irrationality of the square root of 2. AProS finds, fully automatically, a proof of that claim from premises that correspond to the lemmas Lamport appeals to in his top-level sketch; see [Sieg and Cittadini, 2005, pp. 334–5]. To obtain a "fully verified" proof it would have to be grounded in an appropriate methodological framework. In this case, we would have to prove the premises in elementary number theory. This is an absolutely classical way of proceeding: Book I of Euclid's "Elements" is organized to prove the Pythagorean Theorem: the proof itself uses a few lemmas (as local axioms); that is preceded by the systematic reduction of the complexity of lemmas to the five Euclidean postulates (viewed as global axioms).

Here is a second example: with the help of representability and derivability conditions as local axioms and a general heuristic moving between object- and meta-language, AProS fully automatically finds proofs of Gödel's Incompleteness Theorems and Löb's Theorem; see [Sieg and Cittadini, 2005]. To obtain a fully verified proof for ZF*, i.e., ZF without the axiom of infinity, the local axioms have to be proved in an appropriate theory of syntax. Paulson in his [Paulson, 2014] gives "a machine-assisted proof of the incompleteness theorems" for ZF*. However, no intelligible proof is actually presented; Paulson only "describes" the machine proofs.

[24]Mac Lane completed his thesis *Abgekürzte Beweise im Logikkalkul* [Mac Lane, 1934] in 1934 and published a year later a summary that emphasized its programmatic features. He pointed out, in particular, that proofs are not "mere collections of atomic processes, but are rather complex combinations with a highly rational structure". He reflected in 1979 [Mac Lane, 1979] on this early logical work and remarked: "There remains the real question of the actual structure of mathematical proofs and their strategy. It is a topic long given up by mathematical logicians, but one which still — properly handled — might give us some real insight."

we do, that programming computers "to do" mathematics forces us to formulate principles that underlie mathematical research and, in particular, proof finding:

> [O]ne of the best ways of thinking in depth about how humans find proofs is thinking about how would you automate the process of finding proofs. If you can explain to a computer how to find a proof, you can probably explain it to a human. [Diaz-Lopez, 2016, p. 1027]

Such explanations might convey crucial skills in the training of mathematicians, but also more generally in mathematics education.[25] Gowers thinks, however, that discovering automated proof procedures for computers is still far away, because "computers need a lot more help" than humans. In his [Gowers, 2007], Gowers reflects on the distance between "the linguistic superstructure of mathematics" and "the low-level formalities of rigorous proof".

How can one shrink that distance? Gowers proposes to make precise, as far as possible, notions that are used in the "superstructure" to evaluate proofs; he focuses on the notion of *being memorable*. There may be a strong connection between (easily) "memorizing" and (deeply) "understanding". However, he points to *two objective features* of proofs that are rooted in mathematical and computational experience and make them memorable.[26] There should be, first of all, a clearly statable *main idea* to structure a proof, and, secondly, proofs should be short or, as Gowers puts it, of *low-width*. Remembering all the details of a proof is not as efficient as remembering "a few important ideas and develop the technical skill to convert them quickly into a formal proof". It is better still "if the ideas themselves are not so much memorized as *understood*, so that one feels that they arise naturally". Main ideas can trigger, after all, the

[25]There is some empirical support for such a claim. In [Alcock, Hodds, Roy, and Inglis, 2015] it is shown that even modest training in "self-explanation" when reading proofs improves subsequent proof comprehension. Students are encouraged, in their training, to apply "a series of techniques"; those techniques are described in the training material as follows: "*After reading each line*: Try to identify and elaborate the main ideas of the proof. Attempt to explain each line in terms of previous ideas. These may be ideas from the information in the proof, examples from previous theorems/proofs, or ideas from your own prior knowledge of the topic area. *Before proceeding to the next line of the proof you should ask yourself the following*: Do I understand the ideas used in that lien? Do I understand why those ideas have been used? How do those ideas link to other ideas in the proof, other theorems, or prior knowledge that I may have? Does the self-explanation I have generated help to answer the questions I am asking?" This material is found at `www.setmath.lboro.ac.uk.`

[26]The investigation of this very issue is viewed as an interdisciplinary one that should involve of course mathematicians, but also philosophers, psychologists and, we would add, computer scientists.

generation of proof steps and thus allow a low-width presentation.[27] That is also our principal concern, namely, to have a conceptual organization that makes for strategically constructed, easily surveyable, and deeply understood proofs.

There is an extremely informal principle that is the nexus of all the human-centered approaches and techniques we have discussed. To Gowers seems almost "too obvious to be worth mentioning" [Gowers, 2007, p. 46]. It is helpful that he does mention it; after all, how often are truly obvious principles hidden from view? His "very general principle" is articulated as follows, "one should always be consciously aware of what it is one is trying to prove, and what one already knows." This is, in a very open-ended way, at the heart of Bundy's proof plans and Wiedijk's sketches. It is also the imperative for our approach of bi-directional reasoning with gaps: the gaps in arguments diagrammatically indicate which goals are being pursued, which local assumptions and premises can be appealed to, and what other facts in the deductive framework might be useful to reach the current goal. In this very precise, formal setting we can bring to bear logical strategies and mathematical heuristics.

We have brought into the investigation of "the concept of the specifically mathematical proof" not only a special way of formally representing proofs, but also systematic aspects of their discovery; that certainly enriches Hilbert's theory of proofs. The proof theory of natural deduction, as shaped by Gentzen and Prawitz, has contributed to the discovery aspects by making it possible to narrow strategic choices in the process of natural formalization. If we explicitly incorporate main ideas as heuristics into that process, we may be lead to more subtle measures of the simplicity of (the search for) proofs and to significant connections with memorability. However, this way of proceeding could certainly stimulate another kind of interdisciplinary work, namely, building sophisticated computational models of proof construction that are deeply informed by the practice of mathematics and, thus, would help us to gain insights into the mathematical capacities of the human mind.

[27]These observations are reflected very concretely in the central features of natural formalization. Two examples where parallel leading ideas are employed are discussed in [Sieg, 2010]: the *Pythagorean Theorem* with partitioning of figures and establishing equality of areas; the core object of this paper, *CBT*, with partitioning of sets and establishing the equinumerosity of their respective parts. The question of the identity of proofs, raised in [Gowers, 2007, p. 57], has been pursued in our analysis of proofs of the CBT emphasizing that there is essentially *one* proof. Identity of proofs is for us also relative: relative to a methodological frame, relative to an answer to the question, which part of the foundational work is being considered as a "real" part of a proof?, and relative to a particular conceptual level.

## Appendix A. List of theorems.

| | |
|---|---|
| Mem4 | $a \subseteq b, b \subseteq a \vdash a = b$ |
| Mem6 | $a \subseteq b, b \subseteq c \vdash a \subseteq c$ |
| Bool1 | $a \subseteq a \cup b$ |
| Bool4 | $a \subseteq b \vdash a \subseteq c \cup b$ |
| Bool6 | $a \subseteq c, b \subseteq c \vdash a \cup b \subseteq c$ |
| Bool9 | $a \setminus c \subseteq a$ |
| Bool11 | $b \subseteq a \vdash a = (a \setminus b) \cup b$ |
| Bool15 | $a \setminus b = ((c \setminus a) \cup a) \setminus ((c \setminus a) \cup b)$ |
| Bool21 | $b \subseteq a \vdash \mathrm{Binpart}(a \setminus b, b, a)$ |
| Bool22 | $a \setminus b = d \setminus e, b \subseteq a \vdash a \setminus (d \setminus e) = b$ |
| Equi2 | $f \in \mathrm{inj}(a, b), d \subseteq a \vdash \restriction(f \restriction d) \in \mathrm{bij}(d, f[d])$ |
| Equi4 | $f \in \mathrm{inj}(a, b) \vdash a \approx f[a]$ |
| Equi8 | $a \approx b, a \approx c \vdash b \approx c$ |
| Func10 | $f \in \mathrm{func}(a, b), x \subseteq a \vdash f[x] \subseteq b$ |
| Func12 | $f \in \mathrm{func}(a, b), d \subseteq c, c \subseteq a \vdash f[d] \subseteq f[c]$ |
| Func13 | $f \in \mathrm{func}(c, c), a \subseteq c, b \subseteq c \vdash f[a \cup b] = f[a] \cup f[b]$ |
| Func17 | $f \in \mathrm{inj}(a, b) \vdash f[a] \subseteq b$ |
| Comp11 | $f \in \mathrm{inj}(a, b), g \in \mathrm{inj}(b, c) \vdash g \circ f[a] \subseteq g[b]$ |
| Core12 | $f \in \mathrm{inj}(a, b), g \in \mathrm{inj}(b, c) \vdash \restriction(g \circ f) \in \mathrm{bij}(a, g \circ f[a])$ |
| Id8 | $\vdash {}_1(a) \in \mathrm{bij}(a, a)$ |
| CE4 | $f \in \mathrm{bij}(a, b), b \subseteq a \vdash {}'f \in \mathrm{inj}(a, a)$ |
| CE5 | $f \in \mathrm{bij}(a, b), b \subseteq a \vdash {}'f \in \mathrm{func}(a, a)$ |
| Chain8 | $f \in \mathrm{func}(c, c), a \subseteq c \vdash ¢(c, a, f) \subseteq c$ |
| Chain10 | $f \in \mathrm{func}(c, c), a \subseteq c \vdash a \subseteq ¢(c, a, f) \subseteq c$ |
| Chain11 | $f \in \mathrm{func}(c, c), a \subseteq c \vdash f[¢(c, a, f) \subseteq c] \subseteq ¢(c, a, f) \subseteq c$ |
| Chain12 | $f \in \mathrm{func}(c, c), a \subseteq c, b \subseteq a, \mathrm{Chain}(a, f) \vdash ¢(c, b, f) \subseteq a$ |
| Chain14 | $f \in \mathrm{func}(c, c), a \subseteq c \vdash f[a] \subseteq ¢(c, a, f)$ |
| Chain24 | $f \in \mathrm{func}(c, c), a \subseteq c \vdash ¢(c, a, f) = a \cup f[¢(c, a, f)]$ |
| Chain28 | $f \in \mathrm{func}(c, c), b \subseteq a \vdash (\exists x)x = ¢(c, a, f)$ |
| FUn10 | $\mathrm{Binpart}(x, y, a), \mathrm{Binpart}(w, z, b), (\exists f)f \in \mathrm{bij}(x, w), (\exists f)f \in \mathrm{bij}(y, z)$ $\vdash a \approx b$ |
| CBT_D1 | $h \in \mathrm{bij}(a, e), d \subseteq a, e \subseteq d, z = ¢(a, a \setminus d, {}'h) \vdash z \subseteq a$ |
| CBT_D2 | $h \in \mathrm{bij}(a, e), d \subseteq a, e \subseteq d, z = ¢(a, a \setminus d, {}'h) \vdash {}'h[z] \subseteq d$ |
| CBT_D3 | $h \in \mathrm{bij}(a, e), d \subseteq a, e \subseteq d, z = ¢(a, a \setminus d, {}'h) \vdash a \setminus z \subseteq d$ |
| CBT_D4 | $h \in \mathrm{bij}(a, e), d \subseteq a, e \subseteq d, z = ¢(a, a \setminus d, {}'h) \vdash \mathrm{Binpart}(a \setminus z, {}'h[z], d)$ |
| CBT_D5 | $h \in \mathrm{bij}(a, e), d \subseteq a, e \subseteq d \vdash a \approx d$ |

**Appendix B.** The construction of proofs in the main part of our paper is strategically guided. These strategies are sufficiently general to automatically generate a proof of a problem that is a focal point in [Ganesalingam and Gowers, 2013]. There it is used to illustrate the authors' approach to human-centered automated proof search. Ganesalingam and Gowers characterize the problem as "routine", but consider its solution as paradigmatic for their approach. Here is the problem:

THEOREM B.1. *Let $x$ be a complete metric space and $a \subseteq x$; if $a$ is closed then $a$ is complete.*

The proof to which our strategies lead is slightly more formal than that given by Ganesalingam and Gowers. Its construction proceeds in almost exactly the same steps as theirs. That is remarkable in that their motivations are presented as mathematical, whereas ours are almost purely logical appealing to mathematical insights through (syntactically linked) lemmas.

Neither Ganesalingam and Gowers nor we refer in our respective proofs to the definition of a metric space $[M(x)]$, of a Cauchy sequence with elements from $a$ $[y \in cau(a)]$ or of convergence of such a sequence $y$ to a point $w$ $[CON(y, w)]$. The proofs don't use the definition of a closed set $[CLO(a)]$ but rather an equivalent characterization through Lemma 2 below.[28] In the two formal definitions below we consider $a$ to be a subset of the metric space $x$.

**Definitions.**

1. $C(a)[a \text{ is complete}] \leftrightarrow (\forall y \in cau(a))(\exists w \in a)CON(y, w)$.
2. $w \in lim(a)[w \text{ is a limit point of } a] \leftrightarrow (\exists y \in cau(a))CON(y, w)$.

In our proof we appeal to two lemmas, as do Ganesalingam and Gowers.

**Lemmas.**

1. $(M(x)\ \&\ a \subseteq x) \rightarrow (\forall y \in cau(a))y \in cau(x)$.
2. $(M(x)\ \&\ a \subseteq x) \rightarrow (CLO(a) \leftrightarrow (\forall w \in lim(a))w \in a)$.

The formal statement of the problem is now given as follows:

THEOREM B.2.

$$(M(x)\ \&\ C(x)\ \&\ a \subseteq x) \rightarrow (CLO(a) \rightarrow C(a)).$$

Here is the initial partial proof, i.e., the problem to be solved with the above definitions and lemmas in the "accessible" background.

[28] This is really the substantive mathematical observation; for its standard proof see, for example, (Hewitt and Stromberg), Theorem 6.7 on p. 56.

| | | |
|---|---|---|
| 1. | M(x) | Prem |
| 2. | C(x) | Prem |
| 3. | a⊆x | Prem |
| ... | . . . | |
| 4. | (CLO(a) → C(a)) | Goal |

STEP 1 leads immediately via →-I to the next partial proof:

| | | |
|---|---|---|
| 1. | M(x) | Prem |
| 2. | C(x) | Prem |
| 3. | a⊆x | Prem |
| 4. | CLO(a) | Assum |
| ... | . . . | |
| 5. | C(a) | Goal |
| 6. | (CLO(a) → C(a)) | →I 5 |

This step is strategically immediate, as no other logical move can be made at this point. — The full proof is given next and, following it, we will describe the strategic steps for its construction.

| | | |
|---|---|---|
| 1. | M(x) | Prem |
| 2. | C(x) | Prem |
| 3. | a⊆x | Prem |
| 4. | CLO(a) | Assum |
| 5. | y∈cau(a) | Assum |
| 6. | y∈cau(x) | Theorem (Lemma 1) 5, 3, 1 |
| 7. | (∃w∈x)CON(y,w) | Theorem (Definition of Complete 2) 6, 2 |
| 8. | ($z_1$∈x & CON(y,$z_1$)) | Assum |
| 9. | CON(y,$z_1$) | &ER 8 |
| 10. | (∃y∈cau(a))CON(y,$z_1$) | ∃∈I 9, 5 |
| 11. | $z_1$∈lim(a) | Theorem (Definition of lim) 10 |
| 12. | $z_1$∈a | Theorem (Lemma 2) 11, 4, 3, 1 |
| 13. | (∃w∈a)CON(y,w) | ∃∈I 9, 12 |
| 14. | (∃w∈a)CON(y,w) | ∃∈E 7, 13 |
| 15. | (∀y∈cau(a))(∃w∈a)CON(y,w) | ∀∈I 14 |
| 16. | C(a) | Theorem (Definition of complete) 15 |
| 17. | (CLO(a) → C(a)) | →I 16 |

Here is the description of the steps taken in the strategic, indeed fully automatic proof construction. STEP 2 yields 15 by DEF-I backward. The next move of $\forall\in$-I backward in STEP 3 leads to the additional assumption in 5 and the new goal in 14. To prove the existential statement in 14 we search for an extractable existential statement in one of the premises or assumptions; the definition of $C(x)$ contains $(\exists w \in x)CON(y, w)$ as a strictly positive subformula. To obtain that statement by elimination rules the goal $y \in cau(x)$ has to be obtained from 1, 3, 5 by Lemma 1; that is STEP 4. So we have obtained in 7 $(\exists w \in x)CON(y, w)$; the point of this strategic move was to use next, in STEP 5, $\exists\in$-E and introduce the new assumption $z_1 \in x \ \& \ CON(y, z_1)$ in 8 for pursuing our goal in 13. With $\exists\in$-I applied backward (with $z_1$) in STEP 6 we have two new goals, namely, 9 (that is immediately obtained from 8) and 12. The latter goal leads to its extraction in STEP 7 from Lemma 2; that requires, STEP 8, the proof of 11, $z_1 \in lim(a)$. The defining condition of a limit point in $a$, $(\exists y \in cau(a))CON(y, z_1)$, is the new goal and obtainable in STEP 9 from 5, 9 by $\exists\in$-I.

Our strategic approach leads directly to this derivation. We could not have chosen a more perfect example for illustrating the power of the strategies used in our natural formalization. (The strategies, as emphasized in section 3, are an expansion of AProS's complete search procedure for normal natural deduction proofs in first-order logic.) We should mention that Ganselingam and Gowers put great emphasis on a human-friendly output

from programs that are to mimic human mathematical thought processes: "We would like to create human-style write-ups *while being faithful to its thought processes.*" We don't intend to output more than an informal English version of the final proof *without* presenting any of the dead ends and false starts. The internal "proof object" of AProS records the order in which the lines were introduced in the final proof. Given that information together with the bi-directional construction of proofs it will be relatively straightforward to associate with a completed derivation a very informative description of its construction. Indeed, Sieg's undergraduate student Stephen Wu has done exactly that for sentential logic in his 2017 honors thesis — as a proof of concept.

**Acknowledgements.** Before listing the formal references and connections to our work, we mention the informal debts we have incurred over a long period of time; the project of formalizing proofs of the Cantor-Bernstein theorem was started in 2003. For having worked on aspects of the project, we are most grateful to: Adam Conkey, Ian Kash, Davin Lafon, Einam Livnat, Conor Mayo-Wilson, Alex Smith and, most importantly, Tyler Gibson and Dawn McLaughlin. The work was done in two complementary directions: the more theoretical, academic one that found expression in [Kash, 2004], [Gibson, 2006] and [Livnat, 2011] and the more practical, computational one of expanding the ProofLab for the construction of proofs in set theory; the latter project was started in 2010 and would have been impossible without Dawn's and Tyler's sustained, excellent support. Very important was the work in the summer of 2015, when Patrick Walsh joined the project and formalized, together with Aeyaz Kayani, Dedekind's considerations in his [Dedekind, 1888].

## REFERENCES

[2015] Lara Alcock, Mark Hodds, Somali Roy, and Matthew Inglis, *Investigating and improving undergraduate proof comprehension*, ***Notice of the American Mathematical Society***, vol. 62 (2015), no. 7, pp. 742–753.

[1924] Stefan Banach, *Un théorème sur les transformations biunivoques*, ***Fundamenta Mathematicae***, vol. 1 (1924), no. 6, pp. 236–239.

[2005] Henk Barendregt and Freek Wiedijk, *The challenge of computer mathematics*, ***Philosophical Transactions of the Royal Society A***, vol. 363 (2005), no. 1835, pp. 2351–2375.

[1981] Lee Harrison Blaine, *Programs for structured proofs*, ***[Suppes, 1981, pp. 81–120]***, 1981.

[1921] Émile Borel, ***Leçons sur la théorie des fonctions***, Gauthier-Villars, Paris, 1921, first edition was published in 1898.

[2004] Nicolas Bourbaki, ***Theory of sets***, Addison-Wesley, 2004.

[1991] Alan Bundy, *A science of reasoning*, ***Computational logic: essays in honor of Alan Robinson*** (J.L. Lassez and G. Plotkin, editors), MIT Press, Cambridge, 1991, pp. 178–198.

[1932] Georg Cantor, *Gesammelte Abhandlungen mathematischen und philosophischen Inhalts*, (Ernst Zermelo, editor), Springer, Berlin, 1932.

[1937] ———, *Briefwechsel Cantor-Dedekind*, (Jean Cavaillès and Emmy Noether, editors), Hermann, Paris, 1937.

[2015] Martin Davis and Wilfried Sieg, *Conceptual confluence in 1936: Post and Turing*, ***Turing's revolution: The impact of his ideas about computability*** (G. Sommaruga and T. Strahm, editors), Birkhäuser, 2015, pp. 3–27.

[1872] Richard Dedekind, ***Stetigkeit und irrationale Zahlen***, Vieweg, Braunschweig, 1872, translated in [Ewald, 1996, pp. 765–779].

[1888] ———, ***Was sind und was sollen die Zahlen?***, Vieweg, Braunschweig, 1888, translated in [Ewald, 1996, pp. 787–833].

[1932] ———, *Gesammelte mathematische Werke*, (Robert Fricke, Emmy Noether, and Øystein Ore, editors), vol. 3, Vieweg, Braunschweig, 1932.

[2010] Oliver Deiser, *Introductory note to 1901*, ***[Zermelo, 2010, pp. 52–70]***, vol. 1, 2010.

[2016] Alexander Diaz-Lopez, *Interview with Sir Timothy Gowers*, ***Notices of the American Mathematical Society***, vol. 63 (2016), no. 9, pp. 1026–1028.

[1996] William Bragg Ewald (editor), ***From Kant to Hilbert: Readings in the foundations of mathematics***, Oxford University Press, Oxford, 1996, two volumes.

[1993] José Ferreirós, *On the relations between Georg Cantor and Richard Dedekind*, ***Historia mathematica***, vol. 20 (1993), no. 4, pp. 343–363.

[1999] ———, ***Labyrinth of thought: a history of set theory and its role in modern mathematics***, Birkhäuser Verlag, Basel, 1999, second edition 2008.

[1893] Gottlob Frege, *Grundgesetze der Arithmetik*, Pohle Verlag, Jena, 1893.

[1980] ———, *Translations from the philosophical writings of Gottlob Frege*, (Peter Geach and Max Black, editors), Blackwell, Oxford, 1980.

[1984] ———, *Collected papers on mathematics, logic, and philosophy*, (Brian McGuinness, editor), Oxford University Press, Oxford, 1984.

[2013] Mohan Ganesalingam and William Timothy Gowers, *A fully automatic problem solver with human-style output*, arXiv:1309.4501, 2013.

[1936] Gerhard Gentzen, *Die Widerspruchsfreiheit der reinen Zahlentheorie*, ***Mathematische Annalen***, vol. 112 (1936), no. 1, pp. 493–565.

[2006] Tyler Gibson, *Proof search in first-order logic with equality*, ***Master's thesis***, Carnegie Mellon University, 2006.

[1931] Kurt Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, ***Monatshefte für Mathematik und Physik***, vol. 38 (1931), no. 1, pp. 173–198, reprinted and translated in [Gödel, 1986, pp. 126–195].

[1933] ———, *The present situation in the foundations of mathematics*, ***[Gödel, 1995]***, 1933, pp. 36–53.

[1934] ———, *On undecidable propositions of formal mathematical systems*, ***in [Gödel, 1986]***, 1934, pp. 346–369.

[1964] ———, *Postscriptum to [Gödel, 1934]*, in [Gödel, 1986], 1964.

[1986] ———, *Collected Works, vol. 1: Publications 1929–1936*, (Solomon Feferman, editor), Oxford University Press, Oxford, 1986.

[1995] ———, *Collected Works, vol. 3: Unpublished essays and lectures*, (Solomon Feferman, editor), Oxford University Press, Oxford, 1995.

[2007] William Timothy Gowers, *Mathematics, memory, and mental arithmetic*, ***Mathematical knowledge*** (Mary Leng, A. Paseau, and M. Potter, editors), Oxford University Press, 2007, pp. 33–58.

[2000] Ivor Grattan-Guinness, ***The search for mathematical roots, 1870–1940: logics, set theories and the foundations of mathematics from Cantor through Russell to Gödel***, Princeton University Press, Princeton, 2000.

[2008] Thomas C. Hales, *Formal proof*, ***Notices of the American Mathematical Society***, vol. 55 (2008), no. 11, pp. 1370–1380.

[1988] Michael Hallett, ***Cantorian set theory and limitation of size***, Oxford University Press, Oxford, 1988.

[2016] Yamin Hamami, *Mathematical inference and logical inference*, manuscript, 2016.

[2008] John Harrison, *Formal proof—theory and practice*, ***Notices of the American Mathematical Society***, vol. 55 (2008), no. 11, pp. 1395–1406.

[2009] ———, ***Handbook of practical logic and automated reasoning***, Cambridge University Press, Cambridge, 2009.

[1969] Edwin Hewitt and Karl Stromberg, ***Real and abstract analysis: a modern treatment of the theory of functions of a real variable***, Springer-Verlag,

Berlin, 1969.

[1918] David Hilbert, *Axiomatisches Denken*, ***Mathematische Annalen***, vol. 78 (1918), pp. 405–415.

[1927] ———, *Die Grundlagen der Mathematik*, ***Abhandlungen aus dem mathematischen Seminar der Hamburgischen Universität***, (1927), no. 6, pp. 65–85.

[2013] ———, *David Hilbert's lectures on the foundations of arithmetic and logic: 1917-1933*, (William Ewald and Wilfried Sieg, editors), vol. 3, Springer-Verlag, 2013.

[2013] Arie Hinkis, ***Proofs of the Cantor-Bernstein theorem: A mathematical excursion***, Science Networks, Historical Studies, vol. 45, Birkhäuser Verlag, Basel, 2013.

[1997] Akihiro Kanamori, *The mathematical import of Zermelo's well-ordering theorem*, this Bulletin, vol. 3 (1997), no. 3, pp. 281–311.

[2004] ———, *Zermelo and set theory*, this Bulletin, vol. 10 (2004), no. 4, pp. 487–553.

[2004] Ian Kash, *A partially automated proof of the Cantor-Bernstein theorem*, ***Senior Thesis***, Carnegie Mellon University, 2004.

[1928] Bronisław Knaster and Alfred Tarski, *Un théorème sur les fonctions d'ensembles*, ***Annales de la Société Polonaise des Mathématiques***, vol. 6 (1928), pp. 133–134.

[1906] Julius König, *Sur la théorie des ensembles*, ***Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences***, vol. 143 (1906), pp. 110–112.

[1911] Alwin Korselt, *Über einen Beweis des Äquivalenzsatzes*, ***Mathematische Annalen***, vol. 70 (1911), no. 2, pp. 294–296.

[1995] Leslie Lamport, *How to write a proof*, ***The American Mathematical Monthly***, vol. 102 (1995), no. 7, pp. 600–608.

[2011] Einam Livnat, *The Cantor-Bernstein theorem in AProS*, ***Master's thesis***, Carnegie Mellon University, 2011.

[1934] Saunders Mac Lane, *Abgekürzte Beweise im Logikkalkül*, ***Ph.D. thesis***, Göttingen, 1934.

[1935] ———, *A logical analysis of mathematical structure*, ***The Monist***, vol. 45 (1935), no. 1, pp. 118–130.

[1979] ———, *A late return to a thesis in logic*, ***Saunders Maclane — Selected papers*** (I. Kaplansky, editor), Springer, 1979, pp. 63–66.

[1984] James McDonald and Patrick Suppes, *Student use of an interactive theorem prover*, ***Contemporary Mathematics***, vol. 29 (1984), pp. 315–360.

[2013] Eugenio G. Omodeo, *Proof verification within set theory*, ***CILC***, 2013.

[2002] Dominique Pastre, *Strong and weak points of the MUSCADET theorem prover–examples from CASC-JC*, ***AI Communications***, vol. 15 (2002), no. 2, 3, pp. 147–160.

[2011] ———, *MUSCADET version 4.1 user's manual*, 2011.

[1993] Lawrence C. Paulson, *Set theory for verification I: From foundations to functions*, ***Journal of Automated Reasoning***, vol. 11 (1993), no. 3, pp. 353–389.

[1994] ———, *A fixedpoint approach to implementing (co)inductive definitions*, ***Automated deduction CADE-12***, Lecture Notes in Computer Science, no. 814, 1994, pp. 148–181.

[1995] ———, *Set theory for verification II: Induction and recursion*, ***Journal of Automated Reasoning***, vol. 15 (1995), no. 2, pp. 167–215.

[2014] ———, *A machine-assisted proof of Gödel's incompleteness theorems for the theory of hereditarily finite sets*, ***The Review of Symbolic Logic***, vol. 7 (2014), no. 3, pp. 484–498.

[1906a] Giuseppe Peano, *Super theorema de Cantor-Bernstein*, ***Rendiconti del Circolo Matematico di Palermo (1884-1940)***, vol. 21 (1906), pp. 360–366.

[1906b] ———, *Super theorema de Cantor-Bernstein*, ***Revisita di Matematica***, vol. 8 (1906), pp. 136–143.

[1906] Henri Poincaré, *Les mathématiques et la logique*, ***Revue de métaphysique et de morale***, vol. 14 (1906), no. 3, pp. 294–317.

[1898] Ernst Schröder, *Über zwei Definitionen der Endlichkeit und G. Cantorsche Sätze*, ***Nova Acta Academiae Caesareae Leopoldino-Carolinae***, vol. 71 (1898), pp. 303–362.

[1997] Wilfried Sieg, *Aspects of mathematical experience*, ***reprinted in [Sieg, 2013]***, 1997, pp. 329–343.

[2010] ———, *Searching for proofs (and uncovering capacities of the mathematical mind)*, ***reprinted in [Sieg, 2013]***, 2010, pp. 377–401.

[2013] ———, ***Hilbert's programs and beyond***, Oxford University Press, Oxford, 2013.

[1998] Wilfried Sieg and John Byrnes, *Normal natural deduction proofs (in classical logic)*, ***Studia Logica***, vol. 60 (1998), no. 1, pp. 67–106.

[2005] Wilfried Sieg and Saverio Cittadini, *Normal natural deduction proofs (in non-classical logics).*, ***Mechanizing mathematical reasoning*** (D. Hutter and W. Stephan, editors), Lecture Notes in Computer Science, vol. 2605, Springer, 2005, pp. 169–191.

[2005] Wilfried Sieg and Clinton Field, *Automated search for Gödel's proofs*, ***Annals of Pure and Applied Logic***, vol. 133 (2005), pp. 319–338.

[2015] Wilfried Sieg and Rebecca Morris, *Dedekind's structuralism: Creating concepts and deriving theorems*, ***Logic, philosophy of mathematics, and their history: Essays in honor of W. W. Tait*** (Erich Reck, editor), 2015, forthcoming.

[2005] Wilfried Sieg and Dirk Schlimm, *Dedekind's analysis of number: Systems and axioms*, ***Synthese***, vol. 147 (2005), no. 1, pp. 121–170.

[2014] ———, *Dedekind's abstract concepts: Models and mappings*, ***Philosophia Mathematica***, (2014).

[2017] Wilfried Sieg, Máté Szabó, and Dawn McLaughlin, *Why Post did [not] have Turing's thesis*, ***Martin Davis on computability, computational logic, and mathematical foundations*** (Eugenio G. Omodeo and Alberto Policriti, editors), Birkhäuser, 2017, pp. 175–208.

[1979] Wolfgang Stegmüller, ***The structuralist view of theories: A possible analogue of the Bourbaki programme in physical science***, Springer-Verlag, 1979.

[1981] Patrick Suppes (editor), ***University-level computer-assisted instruction at Stanford: 1968–1980***, Institute for Mathematical Studies in the Social Sciences, Stanford, 1981.

[2002] ———, ***Representation and invariance of scientific structures***, CSLI Publications, Stanford, 2002.

[1955] Alfred Tarski, *A lattice-theoretical fixpoint theorem and its applications*, ***Pacific Journal of Mathematics***, vol. 5 (1955), no. 2, pp. 285–309.

[2003] Rüdiger Thiele, *Hilbert's twenty-fourth problem*, ***The American Mathematical Monthly***, vol. 110 (2003), no. 1, pp. 1–24.

[2005] ———, *Hilbert and his twenty-four problems*, ***Mathematics and the historians' craft*** (G. van Brummeln and M. Kinyon, editors), Springer, 2005, pp. 243–295.

[1936] Alan M. Turing, *On computable numbers, with an application to the Entscheidungsproblem*, ***Proceedings of the London Mathematical Society***, vol. 42 (1936), pp. 230–265.

[1948a] ———, *Intelligent machinery*, ***Collected works of A.M. Turing, Mechanical Intelligence*** (D.C. Ince, editor), 1992, North Holland, 1948, originally written as a report for the national physical labratory, pp. 107–127.

[1948b] ———, *Practical forms of type theory*, ***The Journal of Symbolic Logic***, vol. 13 (1948), no. 2, pp. 80–94.

[1967] Jean van Heijenoort (editor), ***From Frege to Gödel: a source book in mathematical logic, 1879–1931***, Harvard University Press, Cambridge, 1967.

[1902] Alfred North Whitehead, *On cardinal numbers*, ***American Journal of Mathematics***, vol. 24 (1902), no. 4, pp. 367–394.

[1910] Alfred North Whitehead and Bertrand Russell, ***Principia Mathematica***, vol. 1, Cambridge University Press, 1910.

[1912] ———, ***Principia Mathematica***, vol. 2, Cambridge University Press, 1912.

[1913] ———, ***Principia Mathematica***, vol. 3, Cambridge University Press, 1913.

[2008] Freek Wiedijk, *Formal proof: getting started*, ***Notices of the American Mathematical Society***, vol. 55 (2008), no. 11.

[2011] ———, *Formal proof sketches*, ***Types for proofs and programs; international workshop, Types 2003, Turin***, Lecture Notes in Computer Science, vol. 3085, Springer, 2011, pp. 378–393.

[2006] Wolfgang Windsteiger, *An automated prover for Zermelo-Fraenkel set theory in Theorema*, ***Journal of Symbolic Computation***, vol. 41 (2006), no. 3–4, pp. 435–470.

[1901] Ernst Zermelo, *Über die Addition transfiniter Kardinalzahlen*, ***Nachrichten von der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse aus dem Jahre 1901***, vol. 1901 (1901), pp. 34–38, reprinted in [Zermelo, 2010].

[1908] ———, *Untersuchungen über die Grundlagen der Mengenlehre I*, ***Mathematische Annalen***, vol. 65 (1908), no. 2, pp. 261–281, translated in [van Heijenoort, 1967, pp. 199–215].

[1930] ———, *Über Grenzzahlen und Mengenbereiche*, ***Fundamenta Mathematicae***, vol. 16 (1930), pp. 29–47, translated in [Ewald, 1996, pp. 1219–1233].

[2010] ———, *Collected Works, Gesammelte Werke, 1: Set theory, Miscellanea, Mengenlehre, Varia*, (Heinz-Dieter Ebbinghaus, C. Fraser, and A. Kanamori, editors), Springer Verlag, Berlin, 2010.

[2016] Andrew Zipperer, *A formalization of elementary group theory in the proof assistant Lean*, ***Master's thesis***, Carnegie Mellon University, 2016.

DEPARTMENT OF PHILOSOPHY
CARNEGIE MELLON UNIVERSITY
PITTSBURGH, PA 15213, USA