Boolos's Curious Inference in Isabelle/HOL

Jeffrey Ketland
Faculty of Philosophy, University of Warsaw
jeffreyketland@gmail.com

March 17, 2025

Abstract

In 1987, George Boolos gave an interesting and vivid concrete example of the considerable speed-up afforded by higher-order logic over first-order logic. (A phenomenon first noted by Kurt Gödel in 1936.) Boolos's example concerned an inference I with five premises, and a conclusion, such that the shortest derivation of the conclusion from the premises in a standard system for first-order logic is astronomically huge; while there exists a second-order derivation whose length is of the order of a page or two. Boolos gave a short sketch of that second-order derivation, which relies on the comprehension principle of second-order logic. Here, Boolos's inference is formalized into fourteen lemmas, each quickly verified by the automated-theorem-proving assistant Isabelle/HOL.

Contents

1 Introduction

In 1987, George Boolos ([3]) presented the following "curious inference", I:

	Inference I
(1)	$\forall n\ fn1 \ = \ s1$
(2)	$\forall x\ f1sx \ = \ ssf1x$
(3)	$\forall n \forall x\ fsnsx \ = \ fnfsn, x$
(4)	$D1$
(5)	$\forall x\ (Dx \rightarrow Dsx)$
$\therefore$	
(6)	$Dfssss1ssss1$

Why is I "curious"? There are three points about I which Boolos notes:

(i) I is *valid* in first-order logic.

(ii) In a standard deductive system for first-order logic (the system Boolos focuses on is from [5] and the details are given in the appendix of his paper [3]), the shortest derivation of I's conclusion (6), from its premises (1)–(5), has symbol size at least

$$2^{2^{\cdot^{\cdot^{\cdot^{2}}}}} \Big\} \text{height} = 65{,}536\ 2\text{'s}$$

So, the shortest *first-order* derivation for I is gigantic.

(iii) However, there is a reasonably short derivation of I's conclusion from its premises in a deductive system for *second-order* logic.

This is then a rather concrete example of speed-up, particularly the speed-up of higher-order logical systems over their first-order level—an idea first noticed by Kurt Gödel in 1936 ([4]). Boolos comments:

> But it is well beyond the bounds of physical possibility that any actual or conceivable creature or device should ever write down all the symbols of a complete derivation in a standard system of first-order logic of (6) from (1)–(5): there are far too many symbols in any such derivation for this to be possible. ([3]: 1)

Though the inference I is formalized, one may think of "s" as standing for the successor operation, and "f" as standing for an Ackermann-like function which grows very rapidly.[1] As Boolos puts it,

> f denotes an Ackermann-style function $n, x \mapsto f(n, x)$ defined on the positive integers: $f(1, x) = 2x$; $f(n, 1) = 2$; and $f(n + 1, x + 1) = f(n, f(n + 1, x))$.

[1] The original ideas in [1] and [6]. The so-called "Péter-Ackerman function" is defined

Then the premises (4) and (5) say that the set denoted by the unary predicate symbol "D" contains 1 and is closed under s: in a sense, this set is, thus, *inductive*. We wish to prove that the number $fssss1ssss1$ is in the set D. Roughly speaking, a first-order derivation would need to prove this by proving a "reduction formula", of the form,

$$(R) \quad fssss1ssss1 = \overbrace{ss\ldots s}^{k \text{ iterations}}\ 1 \tag{1}$$

Let t be this term $\overbrace{ss\ldots s}^{k \text{ iterations}}\ 1$, which is clearly a "canonical numeral". Here, again roughly, k is the value of the term "$fssss1ssss1$". Since we have $D1$ and $\forall x(Dx \to Dsx)$, the result of applying Modus Ponens k times will yield Dt. Then, using the reduction formula (R), we obtain $Dfssss1fsss1$, the required conclusion.

How big is k? Well, k is gigantic, and thus the size of the required derivation in then gigantic too. Boolos gives a careful proof-theoretic argument,

> For definiteness, we shall concentrate our attention on the system M of Mates' book *Elementary Logic*. ... What we shall show is that the number of symbols in any derivation of (6) from (l)-(5) in M is at least the value of an exponential stack $2^{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}$ containing 64 K, or 65 536, 2s in all. Do not confuse this number, which we shall call $f(4,4)$, with the number 2^{64K}." ([3]: 3).

which provides the estimate for the lower bound:

$$k \geq f(4,4) = 2^{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}} \tag{2}$$

as noted above.

Despite the extra-ordinary length of any first-order derivation, Boolos pointed out that there is a *reasonably short second-order logic derivation*, which would fit in a few pages if fully formalized. Boolos himself provides such a derivation in the Appendix of his paper:

---

by:

$$\begin{aligned} \mathrm{A}(0,n) &= n+1 \\ \mathrm{A}(m,0) &= \mathrm{A}(m-1,1) \\ \mathrm{A}(m,n) &= \mathrm{A}(m-1,\mathrm{A}(m,n-1)) \end{aligned}$$

Such functions are indeed recursive, though they don't fit the mould of primitive recursion. They grow extremely rapidly—outpacing any primitive recursive function.

Boolos's Second-Order Derivation (sketch)

By the comprehension principle of second-order logic, $\exists N \forall z(Nz \iff \forall X[X1 \,\&\, \forall y(Xy \to Xsy) \to Xz])$, and then for some N, $\exists E \forall z(Ez \iff Nz \,\&\, Dz)$.
LEMMA 1: $N1$, $\forall y(Ny \to Nsy)$; $Nssss1$; $E1$, $\forall y(Ey \to Esy)$; $Es1$.
LEMMA 2: $\forall n(Nn \to \forall x(Nx \to Efnx))$.
Proof. By comprehension, $\exists M \forall n(Mn \iff \forall x(Nx \to Efnx))$. We want $\forall n(Nn \to Mn)$. Enough to show $M1$ and $\forall n(Mn \to Msn)$, for then if Nn, Mn.
$M1$: Want $\forall x(Nx \to Ef1x)$. By comprehension $\exists Q \forall x(Qx \iff Ef1x)$. Want $\forall x(Nx \to Qx)$. Enough to show $Q1$ and $\forall x(Qx \to Qsx)$.
$Q1$: Want $Ef11$. But $f11 = s1$ by (1) and $Es1$ by Lemma 1.
$\forall x(Qx \to Qsx)$: Suppose Qx, i.e. $Ef1x$. By (2) $f1sx = ssf1x$; by Lemma 1 twice, $Ef1sx$. Thus Qsx and $M1$.
$\forall n(Mn \to Msn)$: Suppose Mn, i.e. $\forall x(Nx \to Efnx)$. Want Msn, i.e. $\forall x(Nx \to Efsnx)$. By comprehension, $\exists P \forall x(Px \iff Efsnx)$. Want $\forall x(Nx \to Px)$. Enough to show $P1$ and $\forall x(Px \to Psx)$.
$P1$: Want $Efsn1$. But $fsn1 = s1$ by (1) and $Es1$ by Lemma 1.
$\forall x(Px \to Psx)$: Suppose Px, i.e. $Efsnx$; thus $Nfsnx$. Want $Efsnsx$. Since $Nfsnx$ and Mn, $Efnfsnx$. But by (3) $fnfsnx = fsnsx$; thus $Efsnsx$. By Lemma 1, $Nssss1$. By Lemma 2, $Efssss1ssss1$. Thus $Dfssss1ssss1$, as desired.

Obviously, this is highly condensed![2] This is not quite fully formalized, but clearly the missing logical inference steps, in each small sublemma, will not add a large overhead.

An idea worth examining is then to see if this second-order inference can be formalized and verified in an automated reasoning system. There are quite a few of these to work with, and an important one is Isabelle/HOL, originally designed by Lawrence Paulson at Cambridge.[3]

Below, in §2, we construct a formalization in Isabelle following Boolos's proof fairly closely.[4] With some definitions (slightly different from Boolos's) and some coaxing, Isabelle finds the required derivations. We use a "locale" to define the primitive symbols and five premises, and along with a definition of "*inductive*" and four definitions for Boolos's predicates "N", "E", "M" and "Q". Boolos's two main Lemmas then turn into some eighteen formalization lemmas. Isabelle quickly verifies each of these, using its own proof search algorithms.[5]

[2] I believe that Boolos's phrase "for some N" in the second line is unintentional.

[3] The theorem prover Isabelle was designed by Lawrence Paulson in the late 1980s in Cambridge. See [7] for the current Isabelle user's manual.

[4] The Boolos curious inference has also been put into MIZAR and OMEGA in 2007 in [2].

[5] The Isabelle formalization §2 does not use Boolos's predicate "P", which is defined using a *parameter* (i.e. "n"). In my initial attempt at formalization, I found this generated a difficulty in properly expressing the formalization. A similar difficulty is encountered in

Because the main ideas behind the second-order proof are, I believe, independently interesting, in §3, I give a rigorous, but semi-formal, and more "mathematical-looking" proof of the conclusion (6) from the premises (1)–(5). This is structured into fourteen lemmas. We then construct a separate Isabelle/HOL formalization of that in §4. This now has fourteen formalized lemmas, but the definitions adopted match those using in the semi-formal proof (and are again slightly different from Boolos's). These fourteen lemmas are organized into five groups for clarity.

In each case, I do not provide the machine proofs in Isabelle's Isar language of these lemmas, since they aren't very instructive. The informal proofs in in §3 are more instructive, and could, with coaxing, be parlayed into machine proofs.

Boolos uses a notation for function terms and atomic predicates which avoids brackets. We shall prefer to write the inference I slightly differently from Boolos's presentation. The premises (axioms) are:

$$\begin{array}{rc} A1: & F(x,e) = s(e) \\ A2: & F(e,s(y)) = s(s(F(e,y))) \\ A3: & F(s(x),s(y)) = F(x,F(s(x),y)) \\ A4: & D(e) \\ A5: & D(x) \to D(s(x)) \end{array}$$

The result we wish to prove is:

$$D(F(s(s(s(s(e)))),s(s(s(s(e))))))$$

2 Isabelle Formalization I (Based on Boolos's Proof Given in §1)

---

the semi-formal proof at Lemma 11. The subproof for Lemma 11 defines a set A, which implicitly depends on a parameter.

```
theory Boo1 imports Main
begin

text "Boolos's inference"

locale boolax_1 =
  fixes F ::  " 'a × 'a ⇒ 'a "
  fixes s ::  " 'a ⇒  'a "
  fixes D ::  " 'a ⇒ bool "
  fixes e ::  " 'a "
  assumes A1:  "F(x, e) = s(e)"
  and A2:  "F(e, s(y)) = s(s(F(e, y)))"
  and A3:  "F(s(x), s(y)) = F(x, F(s(x), y))"
  and A4:  "D(e)"
  and A5:  "D(x) ⟶ D(s(x))"

context boolax_1
begin

text "Definitions"

definition (in boolax_1) induct ::  "'a set ⇒ bool"
  where
 "induct X ≡ e ∈ X ∧ (∀ x.  (x ∈ X ⟶ s(x) ∈ X))"
definition (in boolax_1) N ::  "'a ⇒ bool"
  where
  "N x ≡ (∀ X. (induct X → x ∈ X))"
definition (in boolax_1) E ::  "'a ⇒ bool"
  where
  "E x ≡ (N x ∧ D x)"
definition (in boolax_1) M ::  "'a ⇒ bool"
  where
  "M x ≡ (∀ y.  (N y ⟶ E(F(x, y))))"
definition (in boolax_1) Q ::  "'a ⇒ bool"
  where
  "Q x ≡ E(F(e, x))"
```

```
text "Lemmas"
lemma lem1: "N e"
  by (simp add: N_def induct_def)
lemma lem2: "N x ⟶ N(s(x))"
  by (simp add: N_def induct_def)
lemma lem3: "N(s(s(s(s(e)))))"
   by (simp add: lem1 lem2)
lemma lem4: "E e"
   using A4 E_def lem1 by auto
lemma lem5: "E x ⟶ E(s(x))"
   by (simp add: A5 E_def lem2)
lemma lem6: "E(s(e))"
   by (simp add: lem4 lem5)
lemma lem7: "Q e"
   by (simp add: A1 Q_def lem6)
lemma lem8: "Q x ⟶ Q(s(x))"
   by (simp add: A2 Q_def lem5)
lemma lem9: "N x ⟶ Q x"
   by (metis N_def induct_def lem7 lem8 mem_Collect_eq)
lemma lem10: "M e"
   by (meson Q_def bool_ax.M_def bool_ax_axioms lem9)
lemma lem11: "E (F(s(n), e))"
   by (simp add: A1 lem6)
lemma lem12: "M x ∧ E (F(s(x), y)) ⟶ E (F(s(x), s(y)))"
   by (simp add: A3 E_def M_def)
lemma lem13: "M x ⟶ induct {y. E (F(s(x), y))}"
   using A1 induct_def lem12 lem6 by auto
lemma lem14: "M x ⟶ M(s(x))"
   by (metis CollectD M_def N_def lem13)
lemma lem15: "N x ⟶ M x"
   by (metis N_def induct_def lem10 lem14 mem_Collect_eq)
lemma lem16: "N x ∧ N y ⟶ E(F(x,y))"
   using M_def lem15 by blast
lemma lem17: "E(F(s(s(s(s(e)))), s(s(s(s(e))))))"
  by (simp add: lem16 lem3)
lemma lem18: "D(F(s(s(s(s(e)))), s(s(s(s(e))))))"
  using E_def lem17 by auto
end
end
```

3 Standard Mathematical Proof

3.1 Main Idea

The main idea behind the short, *second-order* proof is to define the notion of an "inductive set" and define a specific "closure" or "container set" $\mathbb{N}$ to be "the smallest inductive set". These definitions, which are second-order, are:

$$\begin{aligned} \mathsf{Df(ind)}: \quad & X \text{ is inductive} := (e \in X \wedge \forall x(x \in X \to s(x) \in X)) && (3) \\ \mathsf{Df}(\mathbb{N}): \quad & \mathbb{N} := \{x \mid \forall Y(Y \text{ is inductive} \to x \in Y))\} && (4) \end{aligned}$$

So, a set is inductive just if it contains e and is closed under applying s. And the set $\mathbb{N}$ is defined to be the smallest inductive set. Thus,

$$\mathbb{N} = \{e, s(e), s(s(e)), s(s(s(e))), \dots\} \tag{5}$$

Notice that we don't require the usual "Peano properties", of non-surjectivity and injectivity, for e and s.[6]

It is straightforward to prove (these are Lemma 1 and Lemma 2 below):

$$\mathbb{N} \text{ is inductive} \tag{6}$$

$$X \text{ is inductive} \to \mathbb{N} \subseteq X \tag{7}$$

One can easily prove (this is Lemma 4 below),

$$s(s(s(s(e)))) \in \mathbb{N} \tag{8}$$

Now A4 and A5 say that (this is Lemma 3 below),

$$\{x \mid D(x)\} \text{ is inductive,} \tag{9}$$

So, we easily obtain:

$$\mathbb{N} \subseteq \{x \mid D(x)\}. \tag{10}$$

Given these definitions, and the premises A1–A5, the key target is to prove the following claim (this is Lemma 13 below):

[6] I.e., we don't require axioms stating non-surjectivity, $\forall x(s(x) \neq e)$, or injectivity, $\forall x \forall y(s(x) = s(y) \to x = y)$.

$$\text{(Closure)} \quad (\forall x \in \mathbb{N})(\forall y \in \mathbb{N})\ F(x, y) \in \mathbb{N} \tag{11}$$

This claim, (Closure), states that the "container" $\mathbb{N}$ is *closed* under the binary operation F.

It will then follow from (Closure) that:

$$F(s(s(s(s(e)))), s(s(s(s(e))))) \in \mathbb{N}. \tag{12}$$

So, we obtain:

$$D(F(s(s(s(s(e)))), s(s(s(s(e)))))), \tag{13}$$

This is the required conclusion (this is Lemma 14 below).

However, how are we to prove (Closure)? Intuitively, we shall prove this by a *double induction*: an "outer induction" on x, and an "inner induction" on y (where x is a parameter). Note first that (Closure) is logically equivalent to,

$$\forall x(x \in \mathbb{N} \to (\forall y(y \in \mathbb{N} \to F(x, y) \in \mathbb{N})) \tag{14}$$

But (14) is clearly logically equivalent to,

$$\forall x(x \in \mathbb{N} \to \mathbb{N} \subseteq \{y \mid F(x, y) \in \mathbb{N}\}) \tag{15}$$

So, if we define

$$P_1(x, y) \quad := \quad F(x, y) \in \mathbb{N} \tag{16}$$
$$P_2(x) \quad := \quad \mathbb{N} \subseteq \{y \mid P_1(x, y)\} \tag{17}$$

Then (Closure) is logically equivalent (using definitions) to,

$$\forall x(x \in \mathbb{N} \to P_2(x)) \tag{18}$$

In turn, (18), and therefore (Closure), is equivalent (using the definition of $\subseteq$) to,

$$\mathbb{N} \subseteq \{x \mid P_2(x)\} \tag{19}$$

And (19), given the definitions of "inductive" and of $\mathbb{N}$, and therefore (Closure), will follow from a proof of:

$$\{x \mid P_2(x)\} \text{ is inductive} \tag{20}$$

And (20), in turn, by the definition of "inductive", will follow from proofs of:

$$P_2(e) \tag{21}$$
$$P_2(x) \to P_2(s(x)) \tag{22}$$

In summary, we need to establish $P_2(e)$ and $P_2(x) \to P_2(s(x))$. These are Lemmas 9 and 11 below. To prove these, we shall need Lemmas 5, 6, 7 below, and these rely on the premises A1–A3, along with the four definitions, and Lemma 2.

These imply that $\{x \mid P_2(x)\}$ is inductive (Lemma 12 below). Given the meaning of "inductive", this tells us that $\mathbb{N} \subseteq \{x \mid P_2(x)\}$ (this uses Lemma 1 below). From this, we conclude $\forall x(x \in \mathbb{N} \to P_2(x))$, and, deabbreviating, $\forall x(x \in \mathbb{N} \to \mathbb{N} \subseteq \{y \mid F(x,y) \in \mathbb{N}\})$. This fairly quickly implies, $\forall x \forall y(x \in \mathbb{N} \wedge y \in \mathbb{N} \to F(x,y) \in \mathbb{N})$, which is (Closure), and is Lemma 13 below.

The rest of the proof, which leads to Lemma 14 below, follows from (Closure) and earlier lemmas,

$$\{x \mid D(x)\} \text{ is inductive} \tag{23}$$
$$s(s(s(s(e)))) \in \mathbb{N} \tag{24}$$

which are Lemmas 3 and 4, as explained above.

3.2 Proof

Here, we give rigorous but semi-formal proofs of the fourteen lemmas referred to above. The four definitions we shall use are the following:

$\mathsf{Df(ind)}$	X is inductive	$:= (e \in X \wedge \forall x(x \in X \to s(x) \in X))$
$\mathsf{Df}(\mathbb{N})$	$\mathbb{N}$	$:= \{x \mid \forall X(X \text{ is inductive} \to x \in X)\}$
$\mathsf{Df(P_1)}$	$P_1(x,y)$	$:= F(x,y) \in \mathbb{N}$
$\mathsf{Df(P_2)}$	$P_2(x)$	$:= \mathbb{N} \subseteq \{y \mid P_1(x,y)\}$

Lemma 1. If X is inductive, then $\mathbb{N} \subseteq X$.

Proof. Using $\mathsf{Df}(\mathbb{N})$.
Suppose (a) X is inductive and (b) $x \in \mathbb{N}$. From $\mathsf{Df}(\mathbb{N})$, we conclude that $\forall Y(Y \text{ is inductive} \to x \in Y)$. And therefore, X is inductive $\to x \in X$. But, by (a), X is inductive. So, $x \in X$. So, discharging (b), $x \in \mathbb{N} \to x \in X$. Since x is arbitrary, therefore $\mathbb{N} \subseteq X$, as claimed. □

Lemma 2. $\mathbb{N}$ is inductive.

Proof. Using $\mathsf{Df(ind)}$ and $\mathsf{Df}(\mathbb{N})$.
For a contradiction, suppose $\mathbb{N}$ is not inductive. From $\mathsf{Df(ind)}$, we have: X is inductive if and only if $e \in X$ and, for all x, if $x \in X$, then $s(x) \in X$. So, either (a) $e \notin \mathbb{N}$ or $\exists x(x \in \mathbb{N} \wedge s(x) \notin \mathbb{N})$.

Now, assume (a) holds. from $\mathsf{Df}(\mathbb{N})$, $e \in \mathbb{N}$ iff $\forall Y(Y$ is inductive $\rightarrow e \in Y)$. Since $e \notin \mathbb{N}$, there is an inductive set Y such that $e \notin Y$. But since Y is inductive, $e \in Y$. This is a contradiction. Therefore, (b) holds. The statement (b) is existential. Let a witness for (b) be a: so $a \in \mathbb{N}$ and $s(a) \notin \mathbb{N}$. Using $\mathsf{Df}(\mathbb{N})$ and some simplification, it follows that $\forall Y(Y$ is inductive $\rightarrow a \in Y)$, and $\exists Y(Y$ is inductive $\wedge\, s(a) \notin Y)$. The second claim is an existential one, and let a witness for this inductive set be A. So, we have: A is inductive, $a \in A$ and $s(a) \notin A$. From the fact that A is inductive and $\mathsf{Df(ind)}$, it follows that $\forall x(x \in A \rightarrow s(x) \in A)$, and thus $a \in A \rightarrow s(a) \in A$. Hence, $s(a) \in A$, contradicting the above.

Thus, $\mathbb{N}$ is inductive. □

Lemma 3. $\{x \mid D(x)\}$ is inductive.

Proof. Using A4, A5 and $\mathsf{Df(ind)}$.
From $\mathsf{Df(ind)}$, $\{x \mid D(x)\}$ is inductive if and only if $e \in \{x \mid D(x)\}$, and $\forall y(y \in \{x \mid D(x)\} \rightarrow s(y) \in \{x \mid D(x)\})$. So, to establish that $\{x \mid D(x)\}$ is inductive, we need to establish that $D(e)$ and $\forall y(D(y) \rightarrow D(s(y))$. Clearly these follow immediately from premises A4 and A5. □

Lemma 4. $s(s(s(s(e)))) \in \mathbb{N}$.

Proof. Using $\mathsf{Df(ind)}$ and Lemma 2.
By Lemma 2, $\mathbb{N}$ is inductive. Using $\mathsf{Df(ind)}$, it follows that $e \in \mathbb{N}$ and $\forall x(x \in \mathbb{N} \rightarrow s(x) \in \mathbb{N})$. Thus, $e \in \mathbb{N}$. And likewise, $s(e) \in \mathbb{N}$; and $s(s(e)) \in \mathbb{N}$; and $s(s(s(e))) \in \mathbb{N}$; and $s(s(s(s(e)))) \in \mathbb{N}$. □

Lemma 5. $P_1(e, e)$.

Proof. Using A1, $\mathsf{Df(P_1)}$, $\mathsf{Df(ind)}$ and Lemma 2.
We wish to prove that $P_1(e, e)$. Using $\mathsf{Df(P_1)}$, we need to prove $F(e, e) \in \mathbb{N}$.

From A1, we have: $F(x, e) = s(e)$. Hence, $F(e, e) = s(e)$. But we already have shown that $s(e) \in \mathbb{N}$, in the proof of Lemma 4, which relied on $\mathsf{Df(ind)}$ and Lemma 2. So, $F(e, e) \in \mathbb{N}$. □

Lemma 6. $P_1(e, x) \rightarrow P_1(e, s(x))$.

Proof. Using A2, $\mathsf{Df(P_1)}$, $\mathsf{Df(ind)}$ and Lemma 2.
Let us suppose $P_1(e, x)$ holds. By the definition $\mathsf{Df(P_1)}$, we have: $P_1(z, x) \iff F(z, x) \in \mathbb{N}$, and therefore, $F(e, x) \in \mathbb{N}$. Since $\mathbb{N}$ is inductive (Lemma 2), using $\mathsf{Df(ind)}$, it follows that $s(s(F(e, x))) \in \mathbb{N}$. From A2, we have $F(e, s(y)) =$

$s(s(F(e,y)))$. So, relabelling variables, $F(e,s(x)) = s(s(F(e,x)))$. But $s(s(F(e,x))) \in \mathbb{N}$. And, therefore, $F(e,s(x)) \in \mathbb{N}$, as claimed. $\square$

Lemma 7. $\{x \mid P_1(e,x)\}$ is inductive.

Proof. Using $\mathsf{Df}(\mathsf{ind})$, Lemmas 5, and Lemma 6.
By $\mathsf{Df}(\mathsf{ind})$, we need $P_1(e,e)$ and $P_1(e,x) \to P_1(e,s(x))$. But these are Lemmas 5, 6. $\square$

Lemma 8. $P_1(s(x),e)$.

Proof. Using A1, $\mathsf{Df}(\mathsf{P_1})$, $\mathsf{Df}(\mathsf{ind})$ and Lemma 2.
Using $\mathsf{Df}(\mathsf{P_1})$, we claim $F(s(x),e) \in \mathbb{N}$.

From the proof of Lemma 4 (which depends on $\mathsf{Df}(\mathsf{ind})$ and Lemma 2) we have $s(e) \in \mathbb{N}$. From A1, we have $F(x,e) = s(e)$ and thus $F(s(x),e) = s(e)$. So, $F(s(x),e) \in \mathbb{N}$, as claimed. $\square$

Lemma 9. $P_2(e)$.

Proof. Using $\mathsf{Df}(\mathsf{P_2})$, Lemma 1 and Lemma 7.
By $\mathsf{Df}(\mathsf{P_2})$, $P_2(x)$ holds iff $\mathbb{N} \subseteq \{y \mid P_1(x,y)\}$. So, $P_2(e)$ holds iff $\mathbb{N} \subseteq \{y \mid P_1(e,y)\}$. By Lemma 7, $\{x \mid P_1(e,x)\}$ is inductive. And by Lemma 1, it follows that $\mathbb{N} \subseteq \{y \mid P_1(e,y)\}$, and thus $P_2(e)$, as claimed. $\square$

Lemma 10. $P_2(x) \to \forall y(P_1(s(x),y) \to P_1(s(x),s(y)))$.

Proof. Using A3, $\mathsf{Df}(\mathsf{P_1})$ and $\mathsf{Df}(\mathsf{P_2})$.
Let us suppose $P_2(x)$. From $\mathsf{Df}(\mathsf{P_2})$, this implies: $\mathbb{N} \subseteq \{y : P_1(x,y)\}$. We claim: $\forall y(P_1(s(x),y) \to P_1(s(x),s(y)))$.

Suppose $P_1(s(x),y)$. Thus, using $\mathsf{Df}(\mathsf{P_1})$, $F(s(x),y) \in \mathbb{N}$. We claim: $P_1(s(x),s(y)))$.

Since $\mathbb{N} \subseteq \{y : P_1(x,y)\}$, we have $\forall y(y \in \mathbb{N} \to P_1(x,y))$. And thus, $\forall z(z \in \mathbb{N} \to F(x,z) \in \mathbb{N})$. It follows that $F(s(x),y) \in \mathbb{N} \to F(x,F(s(x),y)) \in \mathbb{N})$. Since $F(s(x),y) \in \mathbb{N}$, we have: $F(x,F(s(x),y)) \in \mathbb{N}$. By A3, we have: $F(s(x),s(y)) = F(x,F(s(x),y))$. And therefore, $F(s(x),s(y)) \in \mathbb{N}$. Hence, $P_1(s(x),s(y)))$, as claimed. $\square$

Lemma 11. $P_2(x) \to P_2(s(x))$.

Proof. Using $\mathsf{Df}(\mathsf{P_2})$, $\mathsf{Df}(\mathsf{ind})$, Lemma 1, Lemma 8, and Lemma 10.
Suppose $P_2(x)$. By $\mathsf{Df}(\mathsf{P_2})$, we have: $\mathbb{N} \subseteq \{y : P_1(x,y)\}$. We claim $P_2(s(x))$, i.e. $\mathbb{N} \subseteq \{y : P_1(s(x),y)\}$.

By Lemma 8, we have: $P_1(s(x),e)$. By Lemma 10, we have: $P_2(x) \to \forall y(P_1(s(x),y) \to P_1(s(x),s(y)))$. Thus, we have: $\forall y(P_1(s(x),y) \to P_1(s(x),s(y)))$.

Let $A = \{y \mid P_1(s(x),y)\}$. Thus, by $\mathsf{Df}(\mathsf{ind})$, we conclude that A is inductive. By Lemma 1, we conclude that $\mathbb{N} \subseteq A$, and therefore, $\mathbb{N} \subseteq \{y : P_1(s(x),y)\}$, as claimed. $\square$

Lemma 12. $\{x \mid P_2(x)\}$ is inductive.

Proof. Using $\mathsf{Df(ind)}$, Lemma 9 and Lemma 11.
Using $\mathsf{Df(ind)}$, we claim $P_2(e)$ and $\forall x(P_2(x) \rightarrow P_2(s(x)))$. These are Lemma 9 and Lemma 11, respectively. $\square$

Lemma 13. $x \in \mathbb{N} \wedge y \in \mathbb{N} \rightarrow F(x, y) \in \mathbb{N}$.

Proof. Using $\mathsf{Df(P_1)}$, $\mathsf{Df(P_2)}$, Lemma 1, and Lemma 12.
From Lemma 12, we have: $\{x \mid P_2(x)\}$ is inductive. And thus, by Lemma 1, $\mathbb{N} \subseteq \{x \mid P_2(x)\}$. Let us suppose $x \in \mathbb{N}$ and $y \in \mathbb{N}$. We claim: $F(x, y) \in \mathbb{N}$.

Since $x \in \mathbb{N}$, we conclude, $P_2(x)$. And therefore, using $\mathsf{Df(P_2)}$, we conclude $\mathbb{N} \subseteq \{z \mid P_1(x, z)\}$. But also $y \in \mathbb{N}$. So, $P_1(x, y)$. And therefore, $F(x, y) \in \mathbb{N}$, as claimed. $\square$

Lemma 14. $D(F(s(s(s(s(e)))), s(s(s(s(e)))))).$

Proof. Using Lemma 1, Lemma 3, Lemma 4, and Lemma 13.
By Lemma 4, $s(s(s(s(e)))) \in \mathbb{N}$. So, by Lemma 13, $F(s(s(s(s(e)))), s(s(s(s(e))))) \in \mathbb{N}$. By Lemma 3, $\{x \mid D(x)\}$. Hence, by Lemma 1, $\mathbb{N} \subseteq \{x \mid D(x)\}$.
Thus, $F(s(s(s(s(e)))), s(s(s(s(e))))) \in \{x \mid D(x)\}$. So, $D(F(s(s(s(s(e)))), s(s(s(s(e))))))$, as claimed. $\square$

We now convert this semi-formal proof into an Isabelle formalization in the next section. We merely ask Isabelle to *verify* these lemmas using its own automated proof algorithms, and we don't give the detailed subproofs of each lemma (in Isabelle's Isar language).

4 Isabelle Formalization II (Based on Proof Given in §3)

4.1 Formalization

```
theory Boo2 imports Main
begin

text "Boolos's inference"

locale boolax_2 =
  fixes F ::  " 'a × 'a ⇒ 'a "
  fixes s ::  " 'a ⇒  'a "
  fixes D ::  " 'a ⇒ bool "
  fixes e ::  " 'a "
  assumes A1:  "F(x, e) = s(e)"
  and A2:  "F(e, s(y)) = s(s(F(e, y)))"
  and A3:  "F(s(x), s(y)) = F(x, F(s(x), y))"
  and A4:  "D(e)"
  and A5:  "D(x) ⟶ D(s(x))"
```

```
context boolax_2
begin

text "Definitions"

definition (in boolax_2) induct ::  "'a set ⇒ bool"
   where
  "induct X ≡ e ∈ X ∧ (∀ x.  (x ∈ X ⟶ s(x) ∈ X))"

definition (in boolax_2) N ::  "'a set"
   where
  "N = {x.  (∀ Y. (induct Y ⟶ x ∈ Y))}"

definition (in boolax_2) P1 ::  "'a ⇒ 'a ⇒ bool"
  where
  "P1 x y ≡ F(x,y) ∈ N"

definition (in boolax_2) P2 ::  "'a ⇒ bool"
  where
  P2 x ≡ N ⊆ {y.  P1 x y}"

text "Lemmas"

text "I. Basic Lemmas"

lemma Induction_wrt_N: "induct X ⟶ N ⊆ X"
  using N_def by auto
lemma N_is_inductive:  "induct N"
  by (simp add:  N_def induct_def)
lemma D_is_inductive:  "induct {x.  D(x)}"
  using A4 A5 induct_def by auto
lemma Four_in_N: "s(s(s(s(e)))) ∈ N"
  using induct_def N_is_inductive by auto

text "II. Proof that {x.  P1 e x} is inductive"

lemma P1ex_basis:  "P1 e e"
   using A1 P1_def induct_def N_is_inductive by auto
lemma P1ex_closed:  "P1 e x ⟶ P1 e (s(x))"
   using A2 P1_def induct_def N_is_inductive by auto
lemma P1ex_inductive:  "induct {x.  P1 e x}"
   using induct_def P1ex_basis P1ex_closed by auto
```

```
text "III. Proof that {x.  P2 x} is inductive"

lemma P1sx_basis:  P1 (s(x)) e"
  using A1 P1_def induct_def N_is_inductive by auto
lemma P2_basis:  "P2 e"
  by (simp add:  P2_def Induction_wrt_N P1ex_inductive)
lemma P2_closeda:  "P2 x ⟶ (∀ y.  (P1 (s(x)) y ⟶ P1 (s(x)) (s(y))))"
  using A3 P1_def P2_def by auto
lemma P2_closedb:  "P2 x ⟶ P2 (s(x))"
  using P2_def induct_def Induction_wrt_N P1sx_basis P2_closeda by auto
lemma P2_inductive:  "induct {x.  P2 x}"
  using induct_def P2_basis P2_closedb by auto

text "IV. Proof that N is closed under F"

lemma N_closed_F: "x ∈ N ∧ y ∈ N ⟶ F(x,y) ∈ N"
  using Induction_wrt_N P1_def P2_def P2_inductive by auto

text "V. Conclusion"

lemma F_Four_in_D: "D(F(s(s(s(s(e)))), s(s(s(s(e))))))"
 using D_is_inductive Four_in_N N_closed_F Induction_wrt_N by auto

end
end
```

4.2 Correspondence

The correspondence between the Lemmas of the semi-formal mathematical proof in §3 and the Lemmas of the Isabelle formalization in §4.1 is given in the table below.

Semi-formal lemma	Isabelle lemma
Lemma 1	`lemma Induction_wrt_N: "induct X ⟶ N ⊆ X"`
Lemma 2	`lemma N_is_inductive:  "induct N"`
Lemma 3	`lemma D_is_inductive:  "induct {x.  D(x)}"`
Lemma 4	`lemma Four_in_N: "s(s(s(s(e)))) ∈ N"`
Lemma 5	`lemma P1ex_basis:  "P1 e e"`
Lemma 6	`lemma P1ex_closed:  "P1 e x ⟶ P1 e (s(x))"`
Lemma 7	`lemma P1ex_inductive:  "induct {x.  P1 e x}"`
Lemma 8	`lemma P1sx_basis:  P1 (s(x)) e"`
Lemma 9	`lemma P2_basis:  "P2 e"`
Lemma 10	`lemma P2_closeda:  "P2 x ⟶ (∀ y.  (P1 (s(x)) y ⟶ P1 (s(x)) (s(y))))"`
Lemma 11	`lemma P2_closedb:  "P2 x ⟶ P2 (s(x))"`
Lemma 12	`lemma P2_inductive:  "induct {x.  P2 x}"`
Lemma 13	`lemma N_closed_F: "x ∈ N ∧ y ∈ N ⟶ F(x,y) ∈ N"`
Lemma 14	`lemma F_Four_in_D: "D(F(s(s(s(s(e)))), s(s(s(s(e))))))"`

5 Isabelle Formalization I

theory *Boo1* **imports** *Main*
begin

Boolos's inference

locale *boolax-1* =
 fixes $F :: \ 'a \times 'a \Rightarrow 'a$
 fixes $s :: \ 'a \Rightarrow 'a$
 fixes $D :: \ 'a \Rightarrow \ bool$
 fixes $e :: \ \ 'a$
 assumes *A1*: $F(x, e) = s(e)$
 and *A2*: $F(e, s(y)) = s(s(F(e, y)))$
 and *A3*: $F(s(x), s(y)) = F(x, F(s(x), y))$
 and *A4*: $D(e)$
 and *A5*: $D(x) \longrightarrow D(s(x))$

context *boolax-1*
begin

Definitions

definition (**in** *boolax-1*) $induct :: 'a\ set => bool$
 where $induct\ X \equiv e \in X \wedge (\forall x.\ (x \in X \longrightarrow s(x) \in \ X))$

definition (**in** *boolax-1*) $N :: 'a \Rightarrow bool$
 where $N\ x \equiv (\forall X.\ (induct\ X \longrightarrow x \in X))$

definition (**in** *boolax-1*) $E :: 'a \Rightarrow bool$
 where $E\ x \equiv (N\ x \wedge D\ x)$

definition (**in** *boolax-1*) $M :: 'a \Rightarrow bool$
 where $M\ x \equiv (\forall y.\ (N\ y \longrightarrow \ E(F(x, y))))$

definition (**in** *boolax-1*) $Q :: 'a \Rightarrow bool$
 where $Q\ x \equiv E(F(e, x))$

Lemmas

lemma *lem1*: $N\ e$ $\langle proof \rangle$

lemma *lem2*: $N\ x \longrightarrow N(s(x))$ $\langle proof \rangle$

lemma *lem3*: $N(s(s(s(s(e)))))$ $\langle proof \rangle$

lemma *lem4*: $E\ e$ $\langle proof \rangle$

lemma *lem5*: $E\ x \longrightarrow E(s(x))$ $\langle proof \rangle$

lemma *lem6*: $E(s(e))$ $\langle proof \rangle$

lemma *lem7*: $Q\ e$ ⟨*proof*⟩

lemma *lem8*: $Q\ x \longrightarrow Q(s(x))$ ⟨*proof*⟩

lemma *lem9*: $N\ x \longrightarrow Q\ x$ ⟨*proof*⟩

lemma *lem10*: $M\ e$ ⟨*proof*⟩

lemma *lem11*: $E\ (F(s(n),\ e))$ ⟨*proof*⟩

lemma *lem12*: $M\ x \wedge E\ (F(s(x),\ y)) \longrightarrow E\ (F(s(x),\ s(y)))$ ⟨*proof*⟩

lemma *lem13*: $M\ x \longrightarrow induct\ \{y.\ E\ (F(s(x),\ y))\}$ ⟨*proof*⟩

lemma *lem14*: $M\ x \longrightarrow M(s(x))$ ⟨*proof*⟩

lemma *lem15*: $N\ x \longrightarrow M\ x$ ⟨*proof*⟩

lemma *lem16*: $N\ x \wedge N\ y \longrightarrow E(F(x,y))$ ⟨*proof*⟩

lemma *lem17*: $E(F(s(s(s(s(e)))),\ s(s(s(s(e))))))$ ⟨*proof*⟩

lemma *lem18*: $D(F(s(s(s(s(e)))),\ s(s(s(s(e))))))$ ⟨*proof*⟩

end

end

6 Isabelle Formalization II

theory *Boo2* **imports** *Main*
begin

Boolos's inference

locale *boolax-2* =
fixes $F :: \ 'a \times 'a \Rightarrow \ 'a$
fixes $s :: \ 'a \Rightarrow 'a$
fixes $D :: \ 'a \Rightarrow bool$
fixes $e :: \ 'a$
assumes *A1*: $F(x,\ e) = s(e)$
and *A2*: $F(e,\ s(y)) = s(s(F(e,\ y)))$
and *A3*: $F(s(x),\ s(y)) = F(x,\ F(s(x),\ y))$
and *A4*: $D(e)$
and *A5*: $D(x) \longrightarrow D(s(x))$

context *boolax-2*
begin

Definitions

definition (**in** *boolax-2*) *induct* :: *'a set* ⇒ *bool*
where *induct* $X \equiv (e \in X \wedge (\forall x.\ (x \in X \longrightarrow s(x) \in X)))$

definition (**in** *boolax-2*) N :: *'a set*
where $N = \{x.\ (\forall Y.\ (\mathit{induct}\ Y \longrightarrow x \in Y))\}$

definition (**in** *boolax-2*) *P1* :: *'a* ⇒ *'a* ⇒ *bool*
where $\mathit{P1}\ x\ y \equiv F(x,y) \in N$

definition (**in** *boolax-2*) *P2* :: *'a* ⇒ *bool*
where $\mathit{P2}\ x \equiv N \subseteq \{y.\ \mathit{P1}\ x\ y\}$

Lemmas

I. Basic Lemmas

lemma *Induction-wrt-N*: $\mathit{induct}\ X \longrightarrow N \subseteq X$ ⟨*proof*⟩

lemma *N-is-inductive*: $\mathit{induct}\ N$ ⟨*proof*⟩

lemma *D-is-inductive*: $\mathit{induct}\ \{x.\ D(x)\}$ ⟨*proof*⟩

lemma *Four-in-N*: $s(s(s(s(e)))) \in N$ ⟨*proof*⟩

II. Proof that $x.P1ex$ is inductive

lemma *P1ex-basis*: $\mathit{P1}\ e\ e$ ⟨*proof*⟩

lemma *P1ex-closed*: $\mathit{P1}\ e\ x \longrightarrow \mathit{P1}\ e\ (s(x))$ ⟨*proof*⟩

lemma *P1ex-inductive*: $\mathit{induct}\ \{x.\ \mathit{P1}\ e\ x\}$ ⟨*proof*⟩

III. Proof that $x.P2x$ is inductive

lemma *P1sx-basis*: $\mathit{P1}\ (s(x))\ e$ ⟨*proof*⟩

lemma *P2-basis*: $\mathit{P2}\ e$ ⟨*proof*⟩

lemma *P2-closeda*: $\mathit{P2}\ x \longrightarrow (\forall y.\ (\mathit{P1}\ (s(x))\ y \longrightarrow \mathit{P1}\ (s(x))\ (s(y))))$ ⟨*proof*⟩

lemma *P2-closedb*: $\mathit{P2}\ x \longrightarrow \mathit{P2}(s(x))$ ⟨*proof*⟩

lemma *P2-inductive*: $\mathit{induct}\ \{x.\ \mathit{P2}\ x\}$ ⟨*proof*⟩

IV. Proof that N is closed under F

lemma *N-closed-F*: $x \in N \wedge y \in N \longrightarrow F(x,y) \in N$ ⟨*proof*⟩

V. Conclusion

lemma *F-Four-in-D*: $D(F(s(s(s(s(e)))),\ s(s(s(s(e))))))$ ⟨*proof*⟩

end
end

7 Acknowledgements

The author acknowledges support of the Polish National Science Center (Narodowe Centrum Nauki w Krakowie (NCN), Kraków, Poland), project number 2020/39/B/HS1/02020.

References

[1] W. Ackermann. Zum Hilbertschen Aufbau der reellen Zahlen. *Mathematische Annalen*, 99:118–133, 1928.

[2] C. Benzmüller and C. Brown. The curious inference of Boolos in MIZAR and OMEGA. In R. Matuszewski and A. Zalewska, editors, *From Insight to Proof — Festschrift in Honour of Andrzej Trybulec*, pages 299–388. The University of Białystok, Białystok, Poland, 2007. Online: http://mizar.org/trybulec65/20.pdf.

[3] G. Boolos. A curious inference. *Journal of Philosophical Logic*, 16:1–12, 1987.

[4] K. Gödel. Über die Länge von Beweisen. *Ergebinisse Eines Mathematischen Kolloquiums*, 7:23–24, 1936. Reprinted with English translation, Vol 1 of Gödel's collected works.

[5] B. Mates. *Elementary Logic*. Oxford University Press, New York, 1972.

[6] R. Péter. Konstruktion nichtrekursiver Funktionen. *Mathematische Annalen*, 111:42–60, 1935.

[7] M. Wenzel et al. The Isabelle/Isar reference manual, 2020. The manual is available online.
https://isabelle.in.tum.de/dist/Isabelle/doc/isar-ref.pdf.