Logical Methods in Computer Science
Vol. 4 (3:5) 2008, pp. 1–17
www.lmcs-online.org

Submitted Jan. 31, 2007
Published Sep. 9, 2008

EXTRACTING PROGRAMS FROM CONSTRUCTIVE HOL PROOFS VIA IZF SET-THEORETIC SEMANTICS

ROBERT L. CONSTABLE [a] AND WOJCIECH MOCZYDŁOWSKI [b]

[a,b] Department of Computer Science, Cornell University, Ithaca, NY,14853, USA
e-mail address: {rc,wojtek}@cs.cornell.edu

Abstract. Church's Higher Order Logic is a basis for influential proof assistants — HOL and PVS. Church's logic has a simple set-theoretic semantics, making it trustworthy and extensible. We factor HOL into a constructive core plus axioms of excluded middle and choice. We similarly factor standard set theory, ZFC, into a constructive core, IZF, and axioms of excluded middle and choice. Then we provide the standard set-theoretic semantics in such a way that the constructive core of HOL is mapped into IZF. We use the disjunction, numerical existence and term existence properties of IZF to provide a program extraction capability from proofs in the constructive core.

We can implement the disjunction and numerical existence properties in two different ways: one using Rathjen's realizability for IZF and the other using a new direct weak normalization result for IZF by Moczydłowski. The latter can also be used for the term existence property.

1. Introduction

Church's Higher-Order logic [Chu40, Lei94] has been remarkably successful at capturing the intuitive reasoning of mathematicians. It was distilled from *Principia Mathematica*, and is sometimes called the Simple Theory of Types based on that legacy. It incorporates the λ calculus as its notation for functions, including propositional functions, thus interfacing well with computer science, where the λ calculus is fundamental.

One of the reasons Higher-Order logic is successful is that its axiomatic basis is very small, and it has a clean set-theoretic semantics at a low level of the cummulative hierarchy of sets (up to $\omega + \omega$) and can thus be formalized in a small fragment of ZFC set theory. This means it interfaces well with standard mathematics and provides a strong basis for trust. Moreover, the set theory semantics is the basis for many extensions of the core logic; for example, it is straightforward to add arrays, recursive data types, and records to the logic.

Church's theory is the logical basis of two of the most successful interactive provers used in hardware and software verification, HOL [GM93] and PVS [ORS92]. This is due in

1998 ACM Subject Classification: F.4.1.

Key words and phrases: Church's Higher-Order Logic, HOL, PVS, proof assistants, type theory, constructive set theory, program-extraction, proofs-as-programs, λ calculus.

[a,b] The authors have been partly supported by NSF grants DUE-0333526 and 0430161.

part to the two characteristics mentioned above in addition to its elegant automation based on Milner's tactic mechanism and its elegant formulation in the ML metalanguage.

Until recently, one of the few drawbacks of HOL was that its logical base did not allow a way to express a constructive subset of the logic. This issue was considered by Harrison for HOL-light [Har96], and recently Berghofer implemented a constructive version of HOL in the Isabelle implementation [Ber04, BN02] in large part to enable the extraction of programs from constructive proofs. This raises the question of finding a semantics for HOL that justifies this intuitively sound extraction.

The standard justification for program extraction is based on logics that embedded extraction deeply into their semantics; this is the case for the Calculus of Inductive Constructions (CIC) [CPM90, BC04], Minlog [BBS$^+$98], Computational Type Theory (CTT) [ABC$^+$06, C$^+$86] or the closely related Intuitionistic Type Theory (ITT) [ML82, NPS90]. The mechanism of extraction is built deeply into logic and the provers based on it, e.g. Agda [ACN90] on ITT, Coq [The04] on CIC, MetaPRL [HNC$^+$03] and Nuprl [ACE$^+$00] on CTT.

In this paper we show that there is a way to provide a clean set-theoretic semantics for HOL and at the same time use it to semantically justify program extraction. The idea is to first factor HOL into its constructive core, say Constructive HOL, plus the axioms of excluded middle and choice. The semantics for this language can be given in ZFC set theory, and if that logic is factored into its constructive core, called IZF, plus excluded middle and choice (choice is sufficient to give excluded middle), then in the standard semantics, IZF provides the semantics for Constructive HOL. Moreover, we can base program extraction on the IZF semantics.

The constructive content of IZF is not as transparent as in the constructive set theory CZF of Aczel [Acz78], as he is able to interpret CZF in Type Theory, while no such interpretation is known for IZF. However, it is not possible to express the impredicative nature of Higher-Order Logic in CZF. Also, IZF is not as expressive as Howe's ZFC [How96, How98] with inaccessible cardinals and computational primitives, but this makes IZF a more standard theory.

Our semantics is appealing not only because it factors so elegantly, but also because the computational issues and program extraction can be reduced to the standard constructive properties of IZF — the disjunction, numerical existence and term existence properties.

We can implement the disjunction and numerical existence properties in two different ways: one using Rathjen's realizability for CZF [Rat05], recently extended to IZF [Rat06], and the other using a new direct weak normalization result for IZF by Moczydłowski [Moc06a, Moc06b]. The latter can also be used for the term existence property.

In this paper, we provide a set-theoretic semantics for HOL which has the following properties:

- It is as simple as the standard semantics, presented in Gordon and Melham's [GM93].
- It works in constructive set-theory.
- It provides a semantical basis for program extraction.
- It can be applied to the constructive version of HOL recently implemented in Isabelle-HOL as a means of using constructive HOL proofs as programs.

This paper is organized as follows. In section 2 we present a version of HOL. In section 3 we define set-theoretic semantics. Section 4 defines constructive set theory IZF and states its main properties. We show how these properties can be used for program extraction in section 5.

2. Higher-order logic

In this section, we present in detail higher-order logic. There are two syntactic categories: *terms* and *types*. The types are generated by the following abstract grammar:

$$\tau ::= nat \mid bool \mid prop \mid \tau \to \tau \mid \tau \times \tau$$

The distinction between *bool* and *prop* corresponds to the distinction between the two-element type and the type of propositions in type theory, or between the two-element object and the subobject classifier in category theory or, as we shall see, between 2 and the set of all subsets of 1 in constructive set theory.

The terms of HOL are generated by the following abstract grammar:

$$t ::= x_\tau \mid c_\tau \mid (t_{\tau\to\sigma}\ u_\tau)_\sigma \mid (\lambda x_\tau.\ t_\sigma)_{\tau\to\sigma} \mid (t_\tau, s_\sigma)_{\tau\times\sigma}$$

Thus each term t_α in HOL is annotated with a type α, which we call *the type of* t. We will often skip annotating of terms with types, this practice should not lead to confusion, as the implicit type system is very simple. Terms of type *prop* are called *formulas*.

The free variables of a term t are denoted by $FV(t)$ and defined as usual. We consider α-equivalent terms equal. The notation $t[x := u]$ stands for a capture-avoiding substitution and denotes the result of substituting u for x in the term t.

Our version of HOL has a set of built-in constants. To increase readability, we write $c : \tau$ instead of c_τ to provide the information about the type of c. If the type of a constant involves α, it is a constant *schema*, there is one constant for each type τ substituted for α. There are thus constants $=_{bool}$, $=_{nat}$ and so on.

$$\bot : prop \qquad \top : prop \qquad =_\alpha : \alpha \times \alpha \to prop$$

$$\to : prop \times prop \to prop \qquad \land : prop \times prop \to prop \qquad \lor : prop \times prop \to prop$$

$$\forall_\alpha : (\alpha \to prop) \to prop \qquad \exists_\alpha : (\alpha \to prop) \to prop \qquad \varepsilon_\alpha : (\alpha \to prop) \to \alpha$$

$$0 : nat \qquad S : nat \to nat \qquad false : bool \qquad true : bool$$

We present the proof rules for HOL in a sequent-based natural deduction style. A *sequent* is a pair (Γ, t), where Γ is a list of formulas and t is a formula. The free variables of a context are the free variables of all its formulas. A sequent (Γ, t) is written as $\Gamma \vdash t$. We write binary constants (equality, implication, etc.) using infix notation. We use standard abbreviations for quantifiers: $\forall a : \tau.\ \phi$ abbreviates $\forall_\tau(\lambda a_\tau.\ \phi)$, similarly with $\exists a : \tau.\ \phi$. The proof rules for HOL are as follows:

$$\frac{}{\Gamma \vdash t}\ t \in \Gamma \qquad \frac{}{\Gamma \vdash t = t} \qquad \frac{\Gamma \vdash t = s}{\Gamma \vdash \lambda x_\tau.\ t = \lambda x_\tau.\ s}\ x_\tau \notin FV(\Gamma)$$

$$\frac{\Gamma \vdash t \quad \Gamma \vdash s}{\Gamma \vdash t \land s} \qquad \frac{\Gamma \vdash t \land s}{\Gamma \vdash t} \qquad \frac{\Gamma \vdash t \land s}{\Gamma \vdash s} \qquad \frac{}{\Gamma \vdash \top}$$

$$\frac{\Gamma \vdash t}{\Gamma \vdash t \lor s} \qquad \frac{\Gamma \vdash s}{\Gamma \vdash t \lor s} \qquad \frac{\Gamma \vdash t \lor s \quad \Gamma, t \vdash u \quad \Gamma, s \vdash u}{\Gamma \vdash u}$$

$$\frac{\Gamma, t \vdash s}{\Gamma \vdash t \to s} \qquad \frac{\Gamma \vdash s \to t \quad \Gamma \vdash s}{\Gamma \vdash t} \qquad \frac{\Gamma \vdash s = u \quad \Gamma \vdash t[x := u]}{\Gamma \vdash t[x := s]}$$

$$\frac{\Gamma \vdash f_{\alpha\to prop}\ t_\alpha}{\Gamma \vdash \exists_\alpha(f_{\alpha\to prop})} \qquad \frac{\Gamma \vdash \exists_\alpha(f_{\alpha\to prop}) \quad \Gamma, f_{\alpha\to prop}\ x_\alpha \vdash u}{\Gamma \vdash u}\ x_\alpha \text{ new}$$

Finally, we list HOL axioms.

(1) (FALSE) $\bot = \forall b : prop.\ b$.
(2) (FALSENOTTRUE) $false = true \to \bot$.
(3) (BETA) $(\lambda x_\tau.\ t_\sigma)s_\tau = t_\sigma[x_\tau := s_\tau]$.
(4) (ETA) $(\lambda x_\tau.\ f_{\tau\to\sigma}\ x_\tau) = f_{\tau\to\sigma}$, where $x \notin FV(f)$.
(5) (FORALL) $\forall_\alpha = \lambda P_{\alpha\to prop}.\ (P = \lambda x_\alpha.\ \top)$.
(6) (P3) $\forall n : nat.\ (0 = S(n)) \to \bot$.
(7) (P4) $\forall n, m : nat.\ S(n) = S(m) \to n = m$.
(8) (P5) $\forall P : nat \to prop.\ P(0) \land (\forall n : nat.\ P(n) \to P(S(n))) \to \forall n : nat.\ P(n)$.
(9) (BOOL) $\forall x : bool.\ (x = false) \lor (x = true)$.
(10) (EM) $\forall x : prop.\ (x = \bot) \lor (x = \top)$.
(11) (CHOICE) $\forall P : \alpha \to prop.\ \forall x : \alpha.\ P\ x \to P(\varepsilon_{(\alpha\to prop)\to\alpha}(P))$.

Our choice of rules and axioms is redundant. Propositional connectives, for example, could be defined in terms of quantifiers and *bool*. However, we believe that this makes the account of the semantics clearer and shows how easy it is to define a sound semantics for such system. Our presentation is based on the core part of the theory of [GM93]. It does not include type definitions and parametric polymorphism. We believe extending it to incorporate these features should not be very difficult.

The theory CHOL (Constructive HOL) arises by taking away from HOL the axioms (CHOICE) and (EM).

We write $\vdash_H \phi$ and $\vdash_C \phi$ to denote that HOL and CHOL, respectively, proves ϕ. We will generally use letters $\mathcal{P}, \mathcal{Q}$ to denote proof trees. A notation $\mathcal{P} \vdash_C \phi$ means that $\mathcal{P}$ is a proof tree in CHOL of ϕ.

3. Semantics

3.1. **Set theory.** The set-theoretic semantics needs a small part of the cumulative hierarchy — $R_{\omega+\omega}$ is sufficient to carry out all the constructions. The Axiom of Choice is necessary in order to define the meaning of the ε constant. For this purpose, C will denote a[1] necessarily non-constructive function such that for any $X, Y \in R_{\omega+\omega}$:

- If X is non-empty, then $C(X, Y) \in X$.
- If X is empty and Y is non-empty, then $C(X, Y) \in Y$.
- Otherwise, $C(X, Y)$ is $\emptyset$.

Recall that in the world of set theory, $0 = \emptyset$, $1 = \{0\}$ and $2 = \{0, 1\}$. Classically $P(1)$, the set of all subsets of 1, is equal to 2. This is not the case constructively; there is no uniform way of transforming an arbitrary subset of 1 into an element of 2. In fact, it is easy to see that $P(1) = 2$ entails the law of excluded middle:

Lemma 3.1. If $P(1) = 2$, then for any ϕ, ϕ or $\neg\phi$.

Proof. Suppose $P(1) = 2$ and take a formula ϕ. Consider $A = \{x \in 1 \mid \phi\}$ and $B = \{x \in 1 \mid \neg\phi\}$. Since $A \cup B \in P(1)$, $A \cup B \in 2$, so either $A \cup B = 0$ or $A \cup B = 1$. In the former case, $0 \notin A$ and $0 \notin B$. Then we have $\neg\phi$ because from ϕ we obtain $0 \in A$, which is a contradiction. But we also have $\neg\neg\phi$ because from $\neg\phi$ we obtain $0 \in B$, which is also a

[1]Note that if we want to pinpoint C, we need to assume more than AC, as the existence of a definable choice function for $R_{\omega+\omega}$ is not provable in ZFC.

contradiction. Thus we have refuted the assumption $A \cup B = 0$, so $A \cup B = 1$. Therefore $0 \in A \cup B$, so either $0 \in A$ in which case ϕ, or $0 \in B$ in which case $\neg\phi$. So either ϕ or $\neg\phi$. □

The following helpful lemma, however, does hold in a constructive world:

Lemma 3.2. If $A \in P(1)$, then $A = 1$ iff $0 \in A$.

Let us also define precisely the function application operation in set theory. We borrow the definition from [Acz99].

$$App(f, x) = \{z \mid \exists y.\ z \in y \wedge (x, y) \in f\}$$

The advantage of using this definition over an intuitive one ("the unique y such that $(x, y) \in f$") is that it is defined for all sets f and x. Partiality of *App* would entail serious problems in the constructive setting. This definition is equivalent to the standard one when f is a function:

Lemma 3.3. If f is a function from A to B and $x \in A$, then $App(f, x)$ is the unique y such that $(x, y) \in f$.

Proof. Let y be the unique element of B such that $(x, y) \in f$. If $z \in App(f, x)$ then there is y' such that $z \in y'$ and $(x, y') \in f$. Since $y' = y$, $z \in y$. For the other direction, if $z \in y$, then obviously $z \in App(f, x)$. □

From now on, the notation $f(x)$ means $App(f, x)$. We will also use a lambda notation in set theory to define functions: $\lambda x \in A.\ B(x)$ means $\{(x, B(x)) \mid x \in A\}$.

3.2. **The definition of the semantics.** We first define a meaning $[\![\tau]\!]$ of a type τ by structural induction on τ.

- $[\![nat]\!] = \mathbb{N}$.
- $[\![bool]\!] = 2$.
- $[\![prop]\!] = P(1)$.
- $[\![\tau \times \sigma]\!] = [\![\tau]\!] \times [\![\sigma]\!]$, where $A \times B$ denotes the cartesian product of sets A and B.
- $[\![\tau_1 \to \tau_2]\!] = [\![\tau_1]\!] \to [\![\tau_2]\!]$, where $A \to B$ denotes the set of all functions from A to B.

The meaning of a constant c_α is denoted by $[\![c_\alpha]\!]$ and is defined as follows.

- $[\![=_\alpha]\!] = \lambda(x_1, x_2) \in [\![\alpha]\!] \times [\![\alpha]\!].\ \{x \in 1 \mid x_1 = x_2\}$.
- $[\![\to]\!] = \lambda(b_1, b_2) \in [\![prop]\!] \times [\![prop]\!].\ \{x \in 1 \mid x \in b_1 \to x \in b_2\}$.
- $[\![\vee]\!] = \lambda(b_1, b_2) \in [\![prop]\!] \times [\![prop]\!].\ b_1 \cup b_2$.
- $[\![\wedge]\!] = \lambda(b_1, b_2) \in [\![prop]\!] \times [\![prop]\!].\ b_1 \cap b_2$.
- $[\![false]\!] = [\![\bot]\!] = 0$.
- $[\![true]\!] = [\![\top]\!] = 1$.
- $[\![\forall_\alpha]\!] = \lambda f \in [\![\alpha]\!] \to [\![prop]\!].\ \bigcap_{a \in [\![\alpha]\!]} f(a)$.
- $[\![\exists_\alpha]\!] = \lambda f \in [\![\alpha]\!] \to [\![prop]\!].\ \bigcup_{a \in [\![\alpha]\!]} f(a)$.
- $[\![\varepsilon_\alpha]\!] = \lambda P \in [\![\alpha]\!] \to [\![prop]\!].\ C(P^{-1}(\{1\}), [\![\alpha]\!])$.
- $[\![0]\!] = 0$.
- $[\![S]\!] = \lambda n \in \mathbb{N}.\ n + 1$

Standard semantics, presented for example by Gordon and Melham in [GM93], uses a truth table approach — implication $\phi \to \psi$ is false iff ϕ is true and ψ is false etc. It is easy to see that with excluded middle, our semantics is equivalent to the standard one.

Lemma 3.4 (ZF)**.** For any $A, B \in P(1)$, $[\![\to]\!](A, B) = 0$ iff $A = 1$ and $B = 0$.

Proof. Suppose $[\![\to]\!](A, B) = 0$. Then $\{x \in 1 \mid x \in A \to x \in B\} = 0$, so $0 \notin \{x \in 1 \mid x \in A \to x \in B\}$, so it is not the case that $0 \in A \to 0 \in B$, so $0 \in A$ and $0 \notin B$. Thus, $A = 1$ and $B = 0$. The other direction is easy. $\square$

The definition of our semantics is not original. The meaning of logical constants is essentially a combination of the fact that any complete lattice with pseudo-complements is a model for higher-order logic and that $P(1)$ is a complete lattice with pseudo-complement defined in the clause for $\to$ [RS63]. Similar semantics for HOL have also been provided in category-theoretical setting [LS86]. The novelty of our approach lies in *utilizing* this kind of semantics for the purpose of program extraction in Section 5.

To present the rest of the semantics, we need to introduce environments. An *environment* is a function from HOL variables to sets such that $\rho(x_\tau) \in [\![\tau]\!]$. We will use the symbol ρ exclusively for environments. The meaning $[\![t]\!]_\rho$ of a term t is parameterized by an environment ρ and defined by structural induction on t:

- $[\![c_\tau]\!]_\rho = [\![c_\tau]\!]$.
- $[\![x_\tau]\!]_\rho = \rho(x_\tau)$.
- $[\![s\ u]\!]_\rho = App([\![s]\!]_\rho, [\![u]\!]_\rho)$.
- $[\![\lambda x_\tau.\ u]\!]_\rho = \{(a, [\![u]\!]_{\rho[x_\tau:=a]}) \mid a \in [\![\tau]\!]\}$.
- $[\![(s, u)]\!]_\rho = ([\![s]\!]_\rho, [\![u]\!]_\rho)$.

3.3. **The properties of the semantics.** There are several standard properties of the semantics we have defined.

Lemma 3.5 (Substitution Lemma)**.** For any terms t, s and environments ρ, $[\![t]\!]_{\rho[x:=[\![s]\!]_\rho]} = [\![t[x := s]]\!]_\rho$.

Proof. By structural induction on t. Case t of:

- c — the claim is obvious.
- x. Then $[\![x]\!]_{\rho[x:=[\![s]\!]_\rho]} = [\![s]\!]_\rho = [\![x[x := s]]\!]_\rho$.
- $u\ v$. Then $[\![u\ v]\!]_{\rho[x:=[\![s]\!]]} = App([\![u]\!]_{\rho[x:=[\![s]\!]_\rho]}, [\![v]\!]_{\rho[x:=[\![s]\!]_\rho]})$. By the inductive hypothesis, this is equal to $App([\![u[x := s]]\!]_\rho, [\![v[x := s]]\!]_\rho) = [\![u[x := s]\ v[x := s]]\!]_\rho = [\![t[x := s]]\!]_\rho$.
- (u, v). Similar to the previous case.
- $\lambda y_\tau.\ u$. Without loss of generality we may assume that $y \notin \{x\} \cup FV(s)$. Then $[\![t]\!]_{\rho[x:=s]} = \{(a, [\![u]\!]_{\rho[x:=[\![s]\!]_\rho][y:=a]}) \mid a \in [\![\tau]\!]\}$. By the inductive hypothesis, this is equal to $\{(a, [\![u[x := s]]\!]_{\rho[y:=a]}) \mid a \in [\![\tau]\!]\} = [\![(\lambda y_\tau.\ u[x := s])]\!]_\rho = [\![t[x := s]]\!]_\rho$. $\square$

Lemma 3.6. For any type α, $\exists x.\ x \in [\![\alpha]\!]$.

Proof. Easy. $\square$

Lemma 3.7. If $x_\sigma \notin FV(t)$, then for any $b \in [\![\sigma]\!]$, $[\![t]\!]_\rho = [\![t]\!]_{\rho[x_\sigma:=b]}$.

Proof. Straightforward induction on t. We only show the case when $t = \lambda y_\tau.\ u$. Without loss of generality we can assume that $y \neq x$. We have $[\![t]\!]_\rho = \{(a, [\![u]\!]_{\rho[y:=a]}) \mid a \in [\![\tau]\!]\}$. Since $x \notin FV(u)$, by the inductive hypothesis this is equal to $\{(a, [\![u]\!]_{\rho[y:=a][x:=b]}) \mid a \in [\![\tau]\!]\}$. Since $x \neq y$, this is also equal to $\{(a, [\![u]\!]_{\rho[x:=b][y:=a]}) \mid a \in [\![\tau]\!]\} = [\![\lambda y_\tau.\ u]\!]_{\rho[x:=b]}$. $\square$

Lemma 3.8. For any ρ, $[\![t_\alpha]\!]_\rho \in [\![\alpha]\!]$.

By induction on t. Case t of:

- x_τ. The claim follows by the definition of environments.
- c_τ. We proceed by case analysis of c. We show the interesting cases.
  - $\forall_\alpha$. The type of c is $(\alpha \to \mathit{prop}) \to \mathit{prop}$. We need to show that if f is a function from $[\![\alpha]\!]$ to $P(1)$, then $\bigcap_{a\in[\![\alpha]\!]} f(a)$ is in $P(1)$. Since for any $a \in [\![\alpha]\!]$, $f(a) \in P(1)$ and $P(1)$ is closed under intersections, the claim follows.
  - $\exists_\alpha$. The proof is similar and follows by the fact that $P(1)$ is closed under unions.
  - ε_α. The type of ε_α is $(\alpha \to \mathit{prop}) \to \alpha$. Take any function F from $[\![\alpha]\!]$ to $P(1)$. Then $F^{-1}(\{1\}) \subseteq [\![\alpha]\!]$. By the definition of C, if $F^{-1}(\{1\}) \neq \emptyset$, then $[\![\varepsilon_\alpha]\!](F) \in [\![\alpha]\!]$. So suppose $F^{-1}(\{1\}) = \emptyset$. By Lemma 3.6, $[\![\alpha]\!]$ is not empty, so by the definition of C, $[\![\varepsilon_\alpha]\!](F) \in [\![\alpha]\!]$ as well.

In particular, this implies that for any formula t, $[\![t]\!]_\rho \subseteq 1$. So if we want to prove that $[\![t]\!]_\rho = 1$, then by Lemma 3.2 it suffices to show that $0 \in [\![t]\!]_\rho$.

3.4. **Soundness.**

The soundness theorem establishes validity of the proof rules and axioms with respect to the semantics.

Definition 3.9. We write $[\![\Gamma]\!]_\rho = 1$ if $[\![t_1]\!]_\rho = 1, \ldots, [\![t_n]\!]_\rho = 1$, where $\Gamma = t_1, t_2, \ldots, t_n$.

Theorem 3.10 (Soundness). *If $\Gamma \vdash t$ then for any ρ, if $[\![\Gamma]\!]_\rho = 1$, then $[\![t]\!]_\rho = 1$.*

Proof. Straightforward induction on $\Gamma \vdash t$. We show several interesting cases.

-
$$\frac{}{\Gamma \vdash t}\; t \in \Gamma$$
  The claim is trivial.
-
$$\frac{\Gamma \vdash t = s}{\Gamma \vdash \lambda x_\tau.\ t = \lambda x_\tau.\ s}$$
  We need to show that $\{(a, [\![t]\!]_{\rho[x_\tau:=a]}) \mid a \in [\![\tau]\!]\} = \{(a, [\![s]\!]_{\rho[x_\tau:=a]}) \mid a \in [\![\tau]\!]\}$. That is, that for any $a \in [\![\tau]\!]$, $[\![t]\!]_{\rho[x_\tau:=a]} = [\![s]\!]_{\rho[x_\tau:=a]}$. Let $\rho' = \rho[x_\tau := a]$. We get the claim by the inductive hypothesis.
-
$$\frac{\Gamma, t \vdash s}{\Gamma \vdash t \to s}$$
  Suppose $[\![\Gamma]\!]_\rho = 1$. We need to show that $0 \in \{x \in 1 \mid x \in [\![t]\!]_\rho \to x \in [\![s]\!]_\rho\}$. Since $0 \in 1$, assume $0 \in [\![t]\!]_\rho$. Then $[\![\Gamma, t]\!]_\rho = 1$. By the inductive hypothesis $[\![s]\!]_\rho = 1$ thus also $0 \in [\![s]\!]_\rho$.
-
$$\frac{\Gamma \vdash t \to s \quad \Gamma \vdash t}{\Gamma \vdash s}$$
  Suppose $[\![\Gamma]\!]_\rho = 1$. By the inductive hypothesis, $0 \in \{x \in 1 \mid x \in [\![t]\!]_\rho \to x \in [\![s]\!]_\rho\}$ and $0 \in [\![t]\!]_\rho$, so easily $0 \in [\![s]\!]_\rho$.
-
$$\frac{\Gamma \vdash s = u \quad \Gamma \vdash t[x := u]}{\Gamma \vdash t[x := s]}$$
  Assume $[\![\Gamma]\!]_\rho = 1$. By the inductive hypothesis, $[\![s]\!]_\rho = [\![u]\!]_\rho$ and $[\![t[x := u]]\!]_\rho = 1$. Using the Substitution Lemma we get $[\![t[x := u]]\!]_\rho = [\![t]\!]_{\rho[x:=[\![u]\!]_\rho]} = [\![t]\!]_{\rho[x:=[\![s]\!]_\rho]} = [\![t[x := s]]\!]_\rho$.

- $$\frac{\Gamma \vdash f\ t_\alpha}{\Gamma \vdash \exists_\alpha(f_{\alpha\to prop})}$$
  Assume $[\![\Gamma]\!]_\rho = 1$. We have to show that $0 \in \bigcup_{a\in[\![\alpha]\!]}([\![f]\!]_\rho(a))$, so that there is $a \in [\![\alpha]\!]$ such that $0 \in [\![f]\!]_\rho(a)$. By Lemma 3.8, $[\![t_\alpha]\!]_\rho \in [\![\alpha]\!]$, so taking $a = [\![t_\alpha]\!]_\rho$ we get the claim by the inductive hypothesis.
- $$\frac{\Gamma \vdash \exists_\alpha(f_{\alpha\to prop}) \quad \Gamma, f\ x_\alpha \vdash u}{\Gamma \vdash u}\ x_\alpha \text{ new}$$
  Suppose $[\![\Gamma]\!]_\rho = 1$. By the inductive hypothesis, there is $a \in [\![\alpha]\!]$ such that $0 \in [\![f]\!]_\rho(a)$. Let $\rho' = \rho[x_\alpha := a]$. By the inductive hypothesis we get $0 \in [\![u]\!]_{\rho'}$. As $x_\alpha \notin FV(u)$, by Lemma 3.7 $[\![u]\!]_\rho = 1$. □

Having verified the soundness of the HOL proof rules, we proceed to verify the soundness of the axioms.

Theorem 3.11. *For any axiom t of HOL and any ρ defined on $FV(t)$, $0 \in [\![t]\!]_\rho$.*

Proof. We proceed axiom by axiom and sketch the respective proofs.

- (FALSE) $[\![\bot]\!]_\rho = \emptyset = \bigcap_{a\in P(1)} a = [\![\forall b : prop.\ b]\!]_\rho$. The second equality follows by $0 \in P(1)$.
- (BETA) We have $[\![(\lambda x_\tau.\ t_\sigma)\ s_\tau]\!]_\rho = App([\![\lambda x_\tau.\ t_\sigma]\!]_\rho, [\![s_\tau]\!]_\rho) = App(\{(a, [\![t]\!]_{\rho[x:=a]}) \mid a \in [\![\tau]\!]\}, [\![s_\tau]\!]_\rho) = [\![t]\!]_{\rho[x_\tau := [\![s_\tau]\!]_\rho]} =$ (by the Substitution Lemma) $= [\![t_\sigma[x_\tau := s_\tau]]\!]_\rho$.
- (ETA) $[\![\lambda x_\tau.\ f_{\tau\to\sigma} x_\tau]\!]_\rho = \{(a, [\![f\ x_\tau]\!]_{\rho[x_\tau:=a]}) \mid a \in [\![\tau]\!]\} = \{(a, App([\![f]\!]_{\rho[x_\tau:=a]}, a)) \mid a \in [\![\tau]\!]\} =$ (since $x_\tau \notin FV(f)$) $= \{(a, [\![f]\!]_\rho(a)) \mid a \in [\![\tau]\!]\} = [\![f]\!]_\rho$, as by Lemma 3.8, $[\![f]\!]_\rho \in [\![\tau]\!] \to [\![\sigma]\!]$ and functions in set theory are represented by their graphs.
- (FORALL) We have:
  $$[\![\forall_\alpha]\!]_\rho = \{(F, \bigcap_{a\in[\![\alpha]\!]} F(a)) \mid F \in [\![\alpha]\!] \to P(1)\}$$
  Furthermore:
  $$[\![\lambda F_{\alpha\to prop}.\ F = \lambda x_\alpha.\ \top]\!]_\rho = \{(F, \{z \in 1 \mid F = \lambda x \in [\![\alpha]\!].\ 1\}) \mid F \in [\![\alpha]\!] \to P(1)\}$$
  So take any $F \in [\![\alpha]\!] \to P(1)$. It suffices to show that $\bigcap_{a\in[\![\alpha]\!]} F(a) = \{z \in 1 \mid F = \lambda x \in [\![\alpha]\!].\ 1\}$. We have $x \in \bigcap_{a\in[\![\alpha]\!]} F(a)$ iff for all $a \in [\![\alpha]\!]$, $x \in F(a)$ and $x = 0$. This happens if and only if $x = 0$ and for all $a \in [\![\alpha]\!]$, $F(a) = 1$ which is equivalent to $x \in \{z \in 1 \mid P = \lambda x \in [\![\alpha]\!].\ 1\}$. The claim follows.
- The axioms $P3, P4, P5$ follow by the fact that natural numbers satisfy the respective Peano axioms.
- (BOOL) We need to show that $[\![\forall_{bool}.\ (\lambda x_{bool}.\ x = false \vee x = true)]\!]_\rho = 1$. Unwinding the definition, this is equivalent to $\bigcap_{x\in 2}(\{z \in 1 \mid x = 0\} \cup \{z \in 1 \mid x = 1\}) = 1$. and furthermore to: for all $x \in 2$ and y, $y \in \{z \in 1 \mid x = 0\} \cup \{z \in 1 \mid x = 1\}$ iff $y = 0$. Take any $x \in 2$ and y. The left-to-right direction is obvious, for the right-to-left direction, either $x = 0$ or $x = 1$. In the former case, $0 \in \{z \in 1 \mid x = 0\}$, in the latter $0 \in \{z \in 1 \mid x = 1\}$.
- (EM) We need to show that $[\![\forall_{prop}.\ (\lambda x_{prop}.\ x = \bot \vee x = \top)]\!]_\rho = 1$. Reasoning as in the case of (BOOL), we find that this is equivalent to: for all $x \in P(1)$ and y, $y \in \{z \in 1 \mid x = 0\} \cup \{z \in 1 \mid x = 1\}$ iff $y = 0$. Suppose $x \in P(1)$. At this point, it is impossible

- **Extensionality** Two sets are equal if they have the same elements.
- **Empty Set** There is an empty set.
- **Pairing** For any sets a, b, there is a set consisting of a and b.
- **Infinity** There is a set closed under the successor operation and containing the empty set.
- **Union** For any set a, there is a set $\bigcup a$ which is a union of all elements of a.
- **Power Set** For any set a, there is a set of all subsets of a.
- **Separation** For any formula ϕ, for any set a, there is a set of all elements of a satisfying ϕ.
- **Replacement** For any formula $\phi(x, y, \overline{z})$, for any set a, if for all $x \in a$ there is exactly one y such that $\phi(x, y, \overline{z})$ holds, then there is a set b such that for all $x \in a$ there is $y \in b$ such that $\phi(x, y, \overline{z})$ holds.
- **$\in$-Induction** For any formula $\phi(a, \overline{z})$, if for all sets b $(\forall x \in b.\phi(x, \overline{z}))$ implies $\phi(b, \overline{z})$, then for all a, $\phi(a, \overline{z})$ holds.

Figure 1: The axioms of IZF with Replacement

to proceed further constructively, all we know is that x is a subset of 1, which does not provide enough information to decide whether $x = 0$ or $x = 1$. However, classically, using the rule of excluded middle, $P(1) = 2$ and we proceed as in the previous case.

- (CHOICE) We argue classically, so in particular $P(1) = 2$. We need to show that:

$$\begin{array}{ll} [\![\forall_{\alpha\to prop}(\lambda P_{\alpha\to prop}.\ \forall_\alpha(\lambda x_\alpha.\ Px \to P(\varepsilon_{(\alpha\to prop)\to\alpha}(P))]\!] = 1, & \text{which is equivalent to} \\ \bigcap_{P\in[\![\alpha]\!]\to 2}[\![\forall_\alpha(\lambda x_\alpha.\ Px \to P(\varepsilon_{(\alpha\to prop)\to\alpha}(P))]\!] = 1, & \text{which is equivalent to} \\ \bigcap_{P\in[\![\alpha]\!]\to 2}\bigcap_{x\in[\![\alpha]\!]}[\![Px \to P(\varepsilon_{(\alpha\to prop)\to\alpha}(P))]\!] = 1, & \text{which is equivalent to} \end{array}$$

$$\bigcap_{P\in[\![\alpha]\!]\to 2}\ \bigcap_{x\in[\![\alpha]\!]} \{a \in 1 \mid a \in P(x) \to a \in P(C(P^{-1}(\{1\}), [\![\alpha]\!]))\} = 1.$$

To show this, it suffices to show that for all $P \in [\![\alpha]\!] \to 2$, for all $x \in [\![\alpha]\!]$, if $0 \in P(x)$ then $0 \in P(C(P^{-1}(\{1\}), [\![\alpha]\!]))$. Take any P and x. Suppose $0 \in P(x)$. Then $P(x) = 1$, so $x \in P^{-1}(\{1\})$. Therefore $C(P^{-1}(\{1\}), [\![\alpha]\!])) \in P^{-1}(\{1\})$, so $P(C(P^{-1}(\{1\}), [\![\alpha]\!]) = 1$, which shows the claim. □

Corollary 3.12. HOL is consistent: it is not the case that $\vdash_H \bot$.

Proof. Otherwise we would have $[\![\bot]\!] = [\![\top]\!]$, that is $0 = 1$. □

4. IZF

The essential advantage of the semantics in the previous section over a standard one is that for the constructive part of HOL this semantics can be defined in constructive set theory IZF.

An obvious approach to creating a constructive version of ZFC set theory is to replace the underlying first-order logic with intuitionistic first-order logic. As many authors have explained [Myh73, Bee85, McC86, Š85], the ZF axioms need to be reformulated so that they do not imply the law of excluded middle.

In a nutshell, to get IZF from ZFC, the Axiom of Choice and Excluded Middle are taken away and Foundation is reformulated as $\in$-induction. The axioms of IZF are thus

Extensionality, Union, Infinity, Power Set, Separation, Replacement or Collection[2] and $\in$-Induction. The list of axioms for the version with Replacement can be found in Figure 1. A detailed account of the theory can be found for example in Friedman [Fri73]. Besoon's book [Bee85] and Ščedrov's paper [Š85] contain a lot of information on metamathematical properties of IZF and related set theories. For convenience, we assume that the first-order logic has built-in bounded quantifiers ($\forall x \in a.\ \phi$ and $\exists x \in a.\ \phi$), defined as abbreviations in the standard way. We also include in the signature all the set terms corresponding to the axioms of IZF — $\mathbb{N}, \bigcup t, P(a)$ etc. For the full list, see [Moc07].

Myhill [Myh73] have proved several important properties of IZF:

- Disjunction Property (DP) : If IZF $\vdash \phi \vee \psi$, then IZF $\vdash \phi$ or IZF $\vdash \psi$.
- Numerical Existence Property (NEP) : If IZF $\vdash \exists x \in \mathbb{N}.\ \phi(x)$, then there is a natural number n such that IZF $\vdash \phi(\overline{n})$, where $\overline{n} = S(S(\ldots(0)))$ and $S(x) = x \cup \{x\}$.
- Term Existence Property (TEP) : If IZF $\vdash \exists x.\ \phi(x)$, then for some term t, IZF $\vdash \phi(t)$.

Moreover, the semantics and the soundness theorem for CHOL work in IZF, as neither Choice nor Excluded Middle are necessary to carry out these developments. Note that the existence of $P(1)$ is crucial for the semantics.

All the properties are constructive — there is a recursive procedure extracting a natural number, a disjunct or a term from a proof. A trivial one is to look through all the proofs for the correct one. For example, if IZF $\vdash \phi \vee \psi$, a procedure could enumerate all theorems of IZF looking for either ϕ or ψ; its termination would be ensured by DP. We discuss more efficient alternatives in section 5.3.

5. Extraction

We will show that the semantics we have defined can serve as a basis for program extraction from proofs. All that is necessary for program extraction from constructive HOL proofs is provided by the semantics and the soundness proof. Therefore, if one wants to provide an extraction mechanism for the constructive part of the logic, it may be sufficient to carefully define set-theoretic semantics, prove the soundness theorem and the extraction mechanism for IZF would take care of the rest. We speculate on practical uses of this approach in section 6.

5.1. IZF Extraction.

We first describe extraction from IZF proofs. To facilitate the description, we will use a very simple fragment of type theory, which we call TT^0.

The *types* of TT^0 are generated by the following abstract grammar. They should not be confused with HOL types; the context will make it clear which types we refer to.

$$\tau ::= * \mid P_\phi \mid nat \mid bool \mid \tau \times \tau \mid \tau + \tau \mid \tau \to \tau$$

We associate with each type τ of TT^0 a set of its elements, which are finitistic objects. The set of elements of τ is denoted by $El(\tau)$ and defined by structural induction on τ:

- $El(*) = \{*\}$.
- $El(P_\phi)$ is the set of all IZF proofs of the formula ϕ.
- $El(nat) = \mathbb{N}$, the set of natural numbers.
- $El(bool) = \{true, false\}$.

[2]There is a difference, in particular the version with Collection does not satisfy Term Existence Property (TEP), defined on the next page. A concerned reader can replace IZF with IZF$_R$ whenever TEP is used.

- $El(\tau_1 \times \tau_2) = El(\tau_1) \times El(\tau_2)$.
- $M \in El(\tau_1 + \tau_2)$ iff either $M = inl(M_1)$ and $M_1 \in El(\tau_1)$ or $M = inr(M_1)$ and $M_1 \in El(\tau_2)$.
- $M \in El(\tau_1 \to \tau_2)$ iff M is a method which given any element of $El(\tau_1)$ returns an element of $El(\tau_2)$.

In the last clause, we use an abstract notion of "method". It will not be necessary to formalize this notion, but for the interested reader, all "methods" we use are functions provably recursive in $ZF + Con(ZF)$, where $Con(ZF)$ denotes consistency of ZF.

The notation $M : \tau$ stands for $M \in El(\tau)$.

We call a TT^0 type *pure* if it does not contain $*$ and P_ϕ. There is a natural mapping of pure types TT^0 to sets. It is so similar to the meaning of the HOL types that we will use the same notation.

- $[\![nat]\!] = \mathbb{N}$.
- $[\![bool]\!] = 2$.
- $[\![\tau \times \sigma]\!] = [\![\tau]\!] \times [\![\sigma]\!]$.
- $[\![\tau + \sigma]\!] = [\![\tau]\!] + [\![\sigma]\!]$, the disjoint union of $[\![\tau]\!]$ and $[\![\sigma]\!]$.
- $[\![\tau \to \sigma]\!] = [\![\tau]\!] \to [\![\sigma]\!]$.

If a set (and a corresponding IZF term) is in a codomain of the map above, we call it *type-like*. If a set A is type-like, then there is a unique pure type τ such that $[\![\tau]\!] = A$. We denote this type $Type(A)$. Thus, type-like sets are these "generated" by pure TT^0 types via natural semantics. Formally, we define a recursive set TL of IZF terms such that for any $t \in TL$, t is type-like and we can find effectively $Type(A)$. The definition of TL follows the definition above: TL is the smallest set such that $\mathbb{N}, 2 \in TL$ and if $t, u \in TL$, then $t \times u$, $t + u$ and $t \to u$ are also elements of TL. Thus, the sentence "A is type-like" stands for "$A \in TL$". Note that for any term $t \in TL$ we can find a term t' such that IZF $\vdash t = t'$ and $t' \notin TL$ — it suffices to take $t' \equiv t \cup \emptyset$.

Before we proceed further, let us extend TT^0 with a new type Q_τ, where τ is any pure type of TT^0. Intuitively, Q_τ is the provable counterpart of $[\![\tau]\!]$. Formally, the members of $El(Q_\tau)$ are pairs $(t, \mathcal{P})$ such that $\mathcal{P} \vdash_{IZF} t \in [\![\tau]\!]$ ($\mathcal{P}$ is an IZF proof of $t \in [\![\tau]\!]$). Note that there is a natural mapping from closed HOL terms M of type τ into Q_τ — it is easy to construct using Lemma 3.8 a proof $\mathcal{P}$ of the fact that $[\![M]\!]_\rho \in [\![\tau]\!]$, so the pair $([\![M]\!]_\rho, P) : Q_\tau$. In particular, any natural number n can be injected into Q_{nat}. The set of pure types stays unchanged.

We are going to tailor extraction from IZF proofs to the HOL logic. For this purpose, we will specify which elements of IZF proofs/formulas carry interesting computational content for us. We will use the type $*$ to mark the parts of proofs we are not interested in.

We first define a helper function T, which takes a pure type τ and returns another type. Intuitively, $T(\tau)$ is the type of the extract from a statement $\exists x.\ x \in [\![\tau]\!]$. The function T is defined by induction on τ:

- $T(bool) = bool$.
- $T(nat) = nat$.
- $T(\tau \times \sigma) = T(\tau) \times T(\sigma)$.
- $T(\tau + \sigma) = T(\tau) + T(\sigma)$.
- $T(\tau \to \sigma) = Q_\tau \to T(\sigma)$. The rationale for this definition is that in order to utilize an IZF function from $[\![\tau]\!]$ to $[\![\sigma]\!]$ we need to supply an element of a set $[\![\tau]\!]$, which is an element of Q_τ.

Furthermore, we assign to each formula ϕ of IZF a TT^0 type $\overline{\phi}$, which intuitively describes the *computational content* of an IZF proof of ϕ. We do it by induction on ϕ:

- $\overline{a \in b} = *$.
- $\overline{a = b} = *$ (atomic formulas carry no useful computational content).
- $\overline{\phi_1 \vee \phi_2} = \overline{\phi_1} + \overline{\phi_2}$.
- $\overline{\phi_1 \wedge \phi_2} = \overline{\phi_1} \times \overline{\phi_2}$.
- $\overline{\phi_1 \to \phi_2} = P_{\phi_1} \to \overline{\phi_2}$.
- $\overline{\exists a \in A.\ \phi_1} = T(Type(A)) \times \overline{\phi_1}$, if A is type-like.
- $\overline{\exists a \in A.\ \phi_1} = *$, if A is not type-like.
- $\overline{\exists a.\ \phi_1} = *$.
- $\overline{\forall a \in A.\ \phi_1} = Q_{Type(A)} \to \overline{\phi_1}$, if A is type-like.
- $\overline{\forall a \in A.\ \phi_1} = *$, if A is not type-like.
- $\overline{\forall a.\ \phi_1} = *$.

The definition is tailored for HOL logic and could be extended to allow meaningful extraction from a larger class of formulas. For example, we could extract a term from $\exists a.\ \phi_1$ using Term Existence Property.

We present several natural examples of our translation in action:

(1) $\overline{\exists x \in \mathbb{N}.\ x = x} = nat \times *$.
(2) $\overline{\forall x \in \mathbb{N}.\ \exists y \in \mathbb{N}.\ \phi} = Q_{nat} \to nat \times \overline{\phi}$.
(3) $\overline{\forall f \in \mathbb{N} \to \mathbb{N}.\ \exists x \in \mathbb{N}.\ f(x) = 0} = Q_{nat \to nat} \to nat \times *$.

These types are richer than what we intuitively would expect — nat in the first case, $nat \to nat$ in the second and $(nat \to nat) \to nat$ in the third, because any closed HOL term of type nat or $nat \to nat$ can be injected into Q_{nat} or $Q_{nat \to nat}$ via the soundness theorem. The extra $*$ can be easily discarded from types (and extracts).

Lemma 5.1. For any IZF term t, which is not type-like, $\overline{\phi[a := t]} = \overline{\phi}$.

Proof. Straightforward induction on ϕ. □

Lemma 5.2 (IZF)**.** $(\exists a \in 2.\ \phi(a))$ iff $\phi(0) \vee \phi(1)$.

We are now ready to describe the extraction function E, which takes an IZF proof $\mathcal{P}$ of a formula ϕ and returns an object of TT^0 type $\overline{\phi}$. We do it by induction on ϕ, checking on the way that the object returned is of type $\overline{\phi}$. Recall that DP, TEP and NEP denote Disjunction, Term and Numerical Existence Property, respectively. Case ϕ of:

- $a \in b$ — return $*$. We have $* : *$.
- $a = b$ — return $*$. We have $* : *$, too.
- $\phi_1 \vee \phi_2$. Apply DP to $\mathcal{P}$ to get a proof $\mathcal{P}_1$ of either ϕ_1 or ϕ_2. In the former case return $inl(E(\mathcal{P}_1))$, in the latter return $inr(E(\mathcal{P}_1))$. By the inductive hypothesis, $E(\mathcal{P}_1) : \overline{\phi_1}$ (or $E(\mathcal{P}_1) : \overline{\phi_2}$), so $E(\mathcal{P}) : \overline{\phi}$ follows.
- $\phi_1 \wedge \phi_2$. Then there are proofs $\mathcal{P}_1$ and $\mathcal{P}_2$ such that $\mathcal{P}_1 \vdash \phi_1$ and $\mathcal{P}_2 \vdash \phi_2$. Return a pair $(E(\mathcal{P}_1), E(\mathcal{P}_2))$. By the inductive hypothesis, $E(\mathcal{P}_1) : \overline{\phi_1}$ and $E(\mathcal{P}_2) : \overline{\phi_2}$, so $(E(\mathcal{P}_1), E(\mathcal{P}_2)) : \overline{\phi_1 \wedge \phi_2}$.
- $\phi_1 \to \phi_2$. Return a function G which takes an IZF proof $\mathcal{Q}$ of ϕ_1, applies $\mathcal{P}$ to $\mathcal{Q}$ (using the modus-ponens rule of the first-order logic) to get a proof $\mathcal{R}$ of ϕ_2 and returns $E(\mathcal{R})$. By the inductive hypothesis, any such $E(\mathcal{R})$ is in $El(\overline{\phi_2})$, so $G : P_{\phi_1} \to \overline{\phi_2}$.
- $\exists a \in A.\ \phi_1(a)$, where A is type-like. Let $T = Type(A)$. We proceed by induction on T. Case T of:

- *bool*. By Lemma 5.2, we have $\phi_1(0) \vee \phi_1(1)$. Apply DP to get a proof $\mathcal{Q}$ of either $\phi_1(0)$ or $\phi_1(1)$. Let b be $false$ or $true$, respectively. Return a pair $(b, E(\mathcal{Q}))$. By the inductive hypothesis, $E(\mathcal{Q}) : \overline{\phi_1(\llbracket b \rrbracket)}$. By Lemma 5.1, since $\llbracket b \rrbracket_\rho$ is not type-like, $E(\mathcal{Q}) : \overline{\phi_1}$, so $(b, E(Q)) : T(bool) \times \overline{\phi} = \overline{\exists a \in 2.\ \phi_1(a)}$.
- *nat*. Apply NEP to $\mathcal{P}$ to get a natural number n and a proof $\mathcal{Q}$ of $\phi_1(\overline{n})$. Return a pair $(n, E(\mathcal{Q}))$. By the inductive hypothesis, $E(\mathcal{Q}) : \overline{\phi_1(\overline{n})}$. By Lemma 5.1, since we can assume without loss of generality that $\overline{n}$ is not type-like, $E(\mathcal{Q}) : \overline{\phi_1}$, so $(n, E(\mathcal{Q})) : T(nat) \times \overline{\phi_1}$.
- (τ, σ). Construct a proof $\mathcal{Q}$ of $\exists a_1 \in \llbracket \tau \rrbracket \exists a_2 \in \llbracket \sigma \rrbracket.\ a = \langle a_1, a_2 \rangle \wedge \phi_1$. Let $M = E(\mathcal{Q})$. By the inductive hypothesis M is a pair $\langle M_1, M_2 \rangle$ such that $M_1 : T(\tau)$ and $M_2 : \overline{\exists a_2 \in \llbracket \sigma \rrbracket.\ a = \langle a_1, a_2 \rangle \wedge \phi_1}$. Therefore M_2 is a pair $\langle M_{21}, M_{22} \rangle$, $M_{21} : T(\sigma)$ and $M_{22} : \overline{a = \langle a_1, a_2 \rangle \wedge \phi_1}$. Therefore M_{22} is a pair $\langle N, O \rangle$, where $O : \overline{\phi_1}$. Therefore $\langle M_1, M_{21} \rangle : T(\tau \times \sigma)$, so $\langle \langle M_1, M_{21} \rangle, O \rangle : T(\tau \times \sigma) \times \overline{\phi_1}$ and we are justified to return $\langle \langle M_1, M_{21} \rangle, O \rangle$.
- $\tau + \sigma$. Construct a proof $\mathcal{Q}$ of $(\exists a \in \llbracket \tau \rrbracket.\ \phi_1) \vee (\exists a \in \llbracket \sigma \rrbracket.\ \phi_1)$. Apply DP to get the proof $\mathcal{Q}_1$ of (without loss of generality) $\exists a \in \llbracket \tau \rrbracket.\ \phi_1$. Let $M = E(\mathcal{Q}_1)$. By the inductive hypothesis, $M = \langle M_1, M_2 \rangle$, where $M_1 : T(\tau)$ and $M_2 : \overline{\phi_1}$. Return $\langle inl(M_1), M_2 \rangle$, which is of type $(T(\tau + \sigma), \overline{\phi_1})$.
- $\tau \to \sigma$. Use TEP to get a term f such that $(f \in \llbracket \tau \rrbracket \to \llbracket \sigma \rrbracket) \wedge \phi_1(f)$. Construct proofs $\mathcal{Q}_1$ of $\forall x \in \llbracket \tau \rrbracket \exists y \in \llbracket \sigma \rrbracket . f(x) = y$ and $\mathcal{Q}_2$ of $\phi_1(f)$. Without loss of generality, we can assume that f is not type-like. By the inductive hypothesis and Lemma 5.1, $E(\mathcal{Q}_2) : \overline{\phi}$. Let G be a function which works as follows: G takes a pair $(t, \mathcal{R})$ such that $\mathcal{R} \vdash t \in \llbracket \tau \rrbracket$, applies $\mathcal{Q}_1$ to $t, \mathcal{R}$ to get a proof $\mathcal{R}_1$ of $\exists y \in \llbracket \sigma \rrbracket.\ f(t) = y$ and calls $E(\mathcal{R}_1)$ to get a term M. By the inductive hypothesis, $M : \overline{\exists y \in \llbracket \sigma \rrbracket.\ f(t) = y}$, so $M = \langle M_1, M_2 \rangle$, where $M_1 : T(\sigma)$. The function G returns M_1. Our extraction procedure $E(\mathcal{P})$ returns $\langle G, E(\mathcal{Q}_2) \rangle$. The type of $\langle G, E(\mathcal{Q}_2) \rangle$ is $(\mathcal{Q}_\tau \to T(\sigma)) \times \overline{\phi_1}$ which is exactly $(T(\tau \to \sigma)) \times \overline{\phi_1}$.

- $\exists a \in A.\ \phi_1(a)$, where A is not type-like. Return $*$.
- $\exists a.\ \phi_1(a)$. Return $*$.
- $\forall a \in A.\ \phi_1(a)$, where A is type-like. Return a function G which takes an element $(t, \mathcal{Q})$ of $Q_{Type(A)}$, applies $\mathcal{P}$ to t and $\mathcal{Q}$ to get a proof R of $\phi_1(t)$, and returns $E(\mathcal{R})$. Without loss of generality, we can assume that t is not type-like. By the inductive hypothesis and Lemma 5.1, $E(\mathcal{R}) : \overline{\phi_1}$, so $G : Q_{Type(A)} \to \overline{\phi_1} = \overline{\forall a \in A.\ \phi_1(a)}$.
- $\forall a \in A.\ \phi_1(a)$, where A is not type-like. Return $*$.
- $\forall a.\ \phi_1(a)$. Return $*$.

5.2. **HOL extraction.**

As in case of IZF, we will show how to do extraction from a subclass of CHOL proofs. The choice of the subclass is largely arbitrary, our choice illustrates the method and can be easily extended.

We say that a CHOL formula is *extractable* if it is generated by the following abstract grammar, where τ varies over pure TT^0 types and $\oplus \in \{\wedge, \vee, \to\}$.

$$\phi ::= \forall x : \tau.\ \phi \mid \exists x : \tau.\ \phi \mid \phi \oplus \phi \mid \bot \mid t = t$$

We will define extraction for CHOL proofs of extractable formulas. By Theorem 3.11, if CHOL $\vdash \phi$, then IZF $\vdash 0 \in \llbracket \phi \rrbracket$. We need to slightly transform this IZF proof in order to come up with a valid input to E from the previous section. To this means, for any extractable $\phi(a_1, \ldots, a_n)$ we define an IZF formula $\phi'(b_1, \ldots, b_n)$ such that IZF $\vdash 0 \in [\![\phi(a_1, \ldots, a_n)]\!]_{\rho[a_1:=b_1,\ldots,a_n:=b_n]} \leftrightarrow \phi'$. The formula ϕ' is essentially ϕ with type membership information replaced by set membership information. We define ϕ' by induction on ϕ, checking the correctness on the way. We work in IZF. Let $\rho' = \rho[a_1 := b_1, \ldots, a_n := b_n]$. Thus we want to show IZF $\vdash 0 \in [\![\phi]\!]_{\rho'} \leftrightarrow \phi'$. Case ϕ of:

- $\bot$. Take $\phi' \equiv 0 \in [\![\bot]\!]_{\rho'}$. The correctness is trivial.
- $t = s$. Take $\phi' \equiv 0 \in [\![t = s]\!]_{\rho'}$. The correctness is trivial.
- $\phi_1 \vee \phi_2$. By the inductive hypothesis we get ϕ'_1 and ϕ'_2 such that $0 \in [\![\phi_1]\!]_{\rho'} \leftrightarrow \phi'_1$ and $0 \in [\![\phi_2]\!]_{\rho'} \leftrightarrow \phi'_2$. Take $\phi' \equiv \phi'_1 \vee \phi'_2$. We have $0 \in [\![\phi_1 \vee \phi_2]\!]_{\rho'}$ iff $0 \in [\![\phi_1]\!]_{\rho'}$ or $0 \in [\![\phi_2]\!]_{\rho'}$ iff $\phi'_1 \vee \phi'_2$, which shows the claim.
- $\phi_1 \wedge \phi_2$. By the inductive hypothesis we get ϕ'_1 and ϕ'_2 such that $0 \in [\![\phi_1]\!]_{\rho'} \leftrightarrow \phi'_1$ and $0 \in [\![\phi_2]\!]_{\rho'} \leftrightarrow \phi'_2$. Set $\phi' \equiv \phi'_1 \wedge \phi'_2$. The correctness follows easily.
- $\phi_1 \to \phi_2$. By the inductive hypothesis we get ϕ'_1 such that $0 \in [\![\phi_1]\!]_{\rho'} \leftrightarrow \phi'_1$ and ϕ'_2 such that $0 \in [\![\phi_2]\!]_{\rho'} \leftrightarrow \phi'_2$. Set $\phi' = \phi'_1 \to \phi'_2$. The correctness follows easily.
- $\forall a : \tau.\ \phi_1(a, a_1, \ldots, a_n)$. By the inductive hypothesis we get $\phi'_1(b, b_1, \ldots, b_n)$ such that $\forall b, b_1, \ldots, b_n,\ 0 \in [\![\phi'_1]\!]_{\rho'[a:=b]} \leftrightarrow \phi'_1$. Set $\phi' \equiv \forall a \in [\![\tau]\!].\ \phi'_1(a, b_1, \ldots, b_n)$. For the correctness, we have $0 \in [\![\forall a : \tau.\ \phi_1(a, a_1, \ldots, a_n)]\!]_{\rho'}$ iff $\forall A \in [\![\tau]\!],\ 0 \in [\![\phi_1]\!]_{\rho'[a:=A]}$. By the inductive hypothesis, this is equivalent to $\forall A \in [\![\tau]\!].\ \phi'_1(A, b_1, \ldots, b_n)$ which is precisely ϕ'_1.
- $\exists a : \tau.\ \phi_1$. By the inductive hypothesis we get $\phi'_1(b, b_1, \ldots, b_n)$ such that
$$\forall b, b_1, \ldots, b_n.\ 0 \in [\![\phi'_1]\!]_{\rho'[a:=b]} \leftrightarrow \phi_1.$$
Set $\phi' \equiv \exists a \in [\![\tau]\!].\ \phi'_1(a, b_1, \ldots, b_n)$. The correctness follows as in the previous case.

Now we can finally define the extraction process. Suppose CHOL $\vdash \phi$, where ϕ is closed and extractable. Let ρ be the empty environment. Using the soundness theorem, construct an IZF proof P that $0 \in [\![\phi]\!]_\rho$. Use the definition above to get ϕ' such that IZF $\vdash 0 \in [\![\phi]\!]_\rho \leftrightarrow \phi'$ and using P obtain a proof R of ϕ'. Finally, apply the extraction function E to R to get the computational extract.

5.3. **Implementation issues.** The extraction process is parameterized by the implementation of NEP, DP and TEP for IZF. Obviously, searching through all IZF proofs to get a witnessing natural number, term or a disjunct would not be a very effective method. We discuss two alternative approaches.

The first approach is based on realizability. Rathjen defines a realizability relation in [Rat05] for weaker, predicative constructive set theory CZF. For any CZF proof of a formula ϕ, there is a realizer e such that the realizability relation $e \Vdash \phi$ holds, moreover, this realizer can be found constructively from the proof. Realizers provide the information for DP and NEP — which of the disjuncts holds and the witnessing number. They could be implemented using lambda terms. These results have been also recently extended to IZF [Rat06]. The approach has the drawback of not providing the proof of TEP, which would restrict the extraction process from statements of the form $\exists x \in [\![\tau]\!].\ \phi$ to atomic types τ. Moreover, the gap between the existing theoretical result and possible implementation is quite wide.

The second, more direct approach is based on Moczydłowski's proof in [Moc06a] of weak normalization of the lambda calculus λZ corresponding to proofs in IZF. The normalization is used to prove NEP, DP and TEP for the theory and the necessary information is extracted from the normal form of the lambda term corresponding to the IZF proof. Thus in order to provide the implementation of DP, NEP and TEP for IZF, it would suffice to implement λZ, which is specified completely in [Moc06a, Moc06b].

An alternative approach has been presented by Berghofer [Ber04]. He defines extraction for a constructive variant of HOL logic directly in the generic theorem prover Isabelle and uses realizability to justify its correctness. His approach could likely be tailored to our CHOL, so that it would yield extracts equivalent to ours. An exciting project would be to *formalize* IZF and both methods of extraction in Isabelle and show their equivalence and correctness.

6. Conclusion

We have presented a computational semantics for HOL via standard interpretation in intuitionistic set theory. The semantics is clean, simple and agrees with the standard one.

The advantage of this approach is that the extraction mechanism is completely external to Constructive HOL. Using only the semantics, we can take any constructive HOL proof and extract from it computational information. No enrichment of the logic in normalizing proof terms is necessary.

The separation of the extraction mechanism from the logic makes the logic very easily extendable. For example, inductive datatypes and subtyping have clean set-theoretic semantics, so can easily be added to HOL preserving consistency, as witnessed in PVS. As the semantics would work constructively, the extraction mechanisms from section 5 could be easily adapted to incorporate them. Similarly, one could define a set-theoretic semantics for the constructive version of HOL implemented in Isabelle ([Ber04, BN02]) in the same spirit, with the same advantages.

The modularity of our approach and the fact that it is much easier to give set-theoretic semantics for the logic than to prove normalization, could make the development of new trustworthy provers with extraction capabilities much easier and faster.

We would like to thank anonymous reviewers for their helpful comments.

References

[ABC$^+$06] Stuart Allen, Mark Bickford, Robert Constable, Richard Eaton, Christoph Kreitz, Lori Lorigo, and Evan Moran. Innovations in computational type theory using Nuprl. *Journal of Applied Logic*, 4(4):428–469, 2006.

[ACE$^+$00] Stuart Allen, Robert Constable, Richard Eaton, Christoph Kreitz, and Lori Lorigo. The Nuprl open logical environment. In David McAllester, editor, *Proceedings of the 17th International Conference on Automated Deduction*, volume 1831 of *Lecture Notes in Artificial Intelligence*, pages 170–176. Springer Verlag, 2000.

[ACN90] Lennart Augustsson, Thierry Coquand, and Bengt Nordström. A short description of another logical framework. In *Proceedings of the First Annual Workshop on Logical Frameworks*, pages 39–42, Sophia-Antipolis, France, 1990.

[Acz78] Peter Aczel. The type theoretic interpretation of constructive set theory. In A. MacIntyre, L. Pacholski, and J. Paris, editors, *Logic Colloquium '77*, pages 55–66. North Holland, 1978.

[Acz99] Peter Aczel. On relating type theories and set theories. In T. Altenkirch, W. Naraschewski, and B. Reus, editors, *Types for Proofs and Programs: International Workshop, TYPES '98, Kloster Irsee, Germany, March 1998*, volume 1657 of *LNCS*, pages 1–18, 1999.

[BBS$^+$98] H. Benl, U. Berger, H. Schwichtenberg, et al. Proof theory at work: Program development in the Minlog system. In W. Bibel and P. G. Schmitt, editors, *Automated Deduction*, volume II, pages 41–71. Kluwer, 1998.

[BC04] Yves Bertot and Pierre Castéran. *Interactive Theorem Proving and Program Development; Coq'Art: The Calculus of Inductive Constructions*. Texts in Theoretical Computer Science. Springer-Verlag, 2004.

[Bee85] Michael J. Beeson. *Foundations of Constructive Mathematics*. Springer-Verlag, 1985.

[Ber04] Stefan Berghofer. *Proofs, Programs and Executable Specifications in Higher Order Logic*. PhD thesis, Technische Universität München, 2004.

[BN02] Stefan Berghofer and Tobias Nipkow. Executing Higher Order Logic. In P. Callaghan, Z. Luo, J McKinna, and R. Pollack, editors, *Types for Proofs and Programs: TYPES'2000*, volume 2277 of *LNCS*, pages 24–40. Springer-Verlag, 2002.

[C$^+$86] Robert L. Constable et al. *Implementing Mathematics with the Nuprl Proof Development System*. Prentice-Hall, NJ, 1986.

[Chu40] Alonzo Church. A formulation of the simple theory of types. *The Journal of Symbolic Logic*, 5:55–68, 1940.

[CPM90] Thierry Coquand and Christine Paulin-Mohring. Inductively defined types, preliminary version. In *COLOG '88, International Conference on Computer Logic*, volume 417 of *LNCS*, pages 50–66. Springer, Berlin, 1990.

[Fri73] Harvey Friedman. The consistency of classical set theory relative to a set theory with intuitionistic logic. *The Journal of Symbolic Logic*, pages 315–319, 1973.

[GM93] Michael Gordon and Tom Melham. *Introduction to HOL: A Theorem Proving Environment for Higher-Order Logic*. Cambridge University Press, Cambridge, 1993.

[Har96] John Harrison. HOLLight: A tutorial introduction. In *Formal Methods in Computer-Aided Design (FMCAD'96)*, volume 1166 of *LNCS*, pages 265–269. Springer, 1996.

[HNC$^+$03] Jason Hickey, Aleksey Nogin, Robert L. Constable, Brian E. Aydemir, Eli Barzilay, Yegor Bryukhov, Richard Eaton, Adam Granicz, Alexei Kopylov, Christoph Kreitz, Vladimir N. Krupski, Lori Lorigo, Stephan Schmitt, Carl Witty, and Xin Yu. MetaPRL — A modular logical environment. In David Basin and Burkhart Wolff, editors, *Proceedings of the 16th International Conference on Theorem Proving in Higher Order Logics (TPHOLs 2003)*, volume 2758 of *LNCS*, pages 287–303. Springer-Verlag, 2003.

[How96] Douglas J. Howe. Semantic foundations for embedding HOL in Nuprl. In Martin Wirsing and Maurice Nivat, editors, *Algebraic Methodology and Software Technology*, volume 1101 of *LNCS*, pages 85–101. Springer-Verlag, Berlin, 1996.

[How98] Douglas J. Howe. Toward sharing libraries of mathematics between theorem provers. In *Frontiers of Combining Systems, FroCoS'98, ILLC*. Kluwer Academic Publishers, 1998.

[Lei94] Daniel Leivant. Higher order logic. In D. M. Gabbay, C. J. Hogger, and J. A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 2: Deduction Methodologies*, pages 229–321. Clarendon Press, Oxford, 1994.

[LS86] J. Lambek and P. J. Scott. *Introduction to Higher-Order Categorical Logic*, volume 7 of *Cambridge Studies in Advanced Mathematics*, page ? Cambridge University Press, Cambridge, UK, 1986.

[McC86] David McCarty. Realizability and recursive set theory. *Journal of Pure and Applied Logic*, 32:153–183, 1986.

[ML82] Per Martin-Löf. Constructive mathematics and computer programming. In *Proceedings of the Sixth International Congress for Logic, Methodology, and Philosophy of Science*, pages 153–175, Amsterdam, 1982. North Holland.

[Moc06a] Wojciech Moczydłowski. Normalization of IZF with Replacement. In *Proc. 15th Ann. Conf. of the EACSL (CSL 2006)*, volume 4207 of *LNCS*. Springer, 2006.

[Moc06b] Wojciech Moczydłowski. A Normalizing Intuitionistic Set Theory with Inaccessible Sets. Technical Report TR2006-2051, Cornell University, 2006.

[Moc07] Wojciech Moczydłowski. A Normalizing Intuitionistic Set Theory with Inaccessible Sets. *Logical Methods in Computer Science*, 3, 2007.

[Myh73] John Myhill. Some properties of intuitionistic Zermelo-Fraenkel set theory. In *Cambridge Summer School in Mathematical Logic*, volume 29, pages 206–231. Springer, 1973.

[NPS90] Bengt Nordström, Kent Petersson, and Jan M. Smith. *Programming in Martin-Löf's Type Theory.* Oxford Sciences Publication, Oxford, 1990.

[ORS92] S. Owre, J. M. Rushby, and N. Shankar. PVS: A prototype verification system. In Deepak Kapur, editor, *Proceedings of the 11th International Conference on Automated Deduction*, volume 607 of *Lecture Notes in Artificial Intelligence*, pages 748–752, Saratoga, NY, June 1992. Springer-Verlag.

[Rat05] Michael Rathjen. The disjunction and related properties for constructive Zermelo-Fraenkel set theory. *Journal of Symbolic Logic*, 70:1233–1254, 2005.

[Rat06] Michael Rathjen. Metamathematical properties of intuitionistic set theories with choice principles. 2006. Manuscript, available from the web page of the author.

[RS63] Helena Rasiowa and Roman Sikorski. *The Mathematics of Metamathematics*. Number 41 in Monogrfie Matematyczne. Polish Scientific Publishers, 1963.

[The04] The Coq Development Team. *The Coq Proof Assistant Reference Manual – Version V8.0*, April 2004.

[Š85] Andrej Ščedrov. Intuitionistic set theory. In Morley, Sčedrov, Harrington and Simpson, editors, *Harvey Friedman's Research on the Foundations of Mathematics*, pages 257–284. North-Holland, 1985.