2

The virtue of simplicity

Edward Nelson*

Part I. Technical

It is known that IST (internal set theory) is a conservative extension of ZFC (Zermelo-Fraenkel set theory with the axiom of choice); see for example the appendix to [2] for a proof using ultrapowers and ultralimits. But these semantic constructions leave one wondering what actually makes the theory work—what are the inner mechanisms of Abraham Robinson's new logic. Let us examine the question syntactically.

Notational conventions: we use x to stand for a variable and other lowercase letters to stand for a sequence of zero or more variables; variables with a prime ′ range over finite sets; variables with a tilde ˜ range over functions.

We take as the axioms of IST the axioms of ZFC together with the following, in which A is an internal formula:

(T) $\forall^{\mathrm{st}}\mathrm{t}\,[\forall^{\mathrm{st}}\mathrm{x}\mathrm{A} \to \forall\mathrm{x}\mathrm{A}]$, where A has free variables x and the variables of t,

(I) $\forall^{\mathrm{st}}\mathrm{y}'\exists\mathrm{x}\forall\mathrm{y}{\in}\mathrm{y}'\,\mathrm{A} \leftrightarrow \exists\mathrm{x}\forall^{\mathrm{st}}\mathrm{y}\,\mathrm{A}$,

(S) $\forall^{\mathrm{st}}\mathrm{x}\exists^{\mathrm{st}}\mathrm{y}\,\mathrm{A}(\mathrm{x},\mathrm{y}) \to \exists^{\mathrm{st}}\tilde{\mathrm{y}}\forall^{\mathrm{st}}\mathrm{x}\,\mathrm{A}\big(\mathrm{x},\tilde{\mathrm{y}}(\mathrm{x})\big)$.

We have written the standardization principle (S) in functional form and required A to be internal; we call this the *restricted* standardization principle. It can be shown that the general standardization principle is a consequence.

All functions must have a domain. There is a neat way, using the reflection principle of set theory, to ensure that $\tilde{\mathrm{y}}$ has a domain, but let me avoid discussion of this point.

We do not take the predicate symbol *standard* as basic, but introduce it by

$$x \text{ is standard } \leftrightarrow \exists^{\mathrm{st}} y[y = x].$$

In this way $\forall^{\mathrm{st}}$ and $\exists^{\mathrm{st}}$ are new *logical* symbols and (I), (S), (T) are *logical* axioms of Abraham Robinson's new logic.

*Department of Mathematics, Princeton University.
`nelson@math.princeton.edu`

For any formula A of IST we define a formula A^+, called the *partial reduction* of A. It will always be of the form $\forall^{st}u\exists^{st}vA^\bullet$ where $A^\bullet$ is internal. It is defined recursively as follows:

$$\begin{array}{rll} \text{if A is internal, } A^+ & \text{is} & A \\ (\neg A)^+ & \text{is} & \forall^{st}\tilde{v}\exists^{st}u\neg A^\bullet\big(u, \tilde{v}(u)\big) \\ (A_1 \vee A_2)^+ & \text{is} & \forall^{st}u_1u_2\exists^{st}v_1v_2[A_1^\bullet \vee A_2^\bullet] \\ (\forall xA)^+ & \text{is} & \forall^{st}u\exists^{st}v'\forall x\exists v{\in}v' A^\bullet \\ (\forall^{st}xA)^+ & \text{is} & \forall^{st}xu\forall^{st}vA^\bullet. \end{array}$$

(We take $\neg$, $\vee$, and $\forall$ as the basic logical operators—the others can be defined in terms of them.) It is understood when forming $(A_1 \vee A_2)^+$ that a variant may be taken (bound variables changed) to avoid colliding variables. If z are the free variables of A, then the *reduction* of A, denoted by A°, is the internal formula

$$\forall u\exists v'\forall z\exists v{\in}v' A^\bullet.$$

This is the same as the partial reduction of the closure of A with $\forall^{st}$ and $\exists^{st}$ replaced by $\forall$ and $\exists$.

We need only show that if A is an axiom of IST, then A° is a theorem, and that for every rule of inference with premise A_1 (or premises A_1 and A_2) and conclusion B, if A_1° is a theorem (or A_1° and A_2° are theorems), then B° is a theorem. This turns out to be quite straightforward in the main, but there is one exception. When I spoke in Aveiro I thought I could present a truly simple syntactical proof of conservativity, but I was mistaken. This remains a desirable goal. So the first part of this paper celebrates the virtue of simplicity by its absence.

The complication lies with the rule of detachment, or modus ponens. First we need a purely internal lemma.

Lemma 1 *(Cross-section) Let A be internal. Then*

$$\exists\widetilde{v}'\forall u'\exists z\forall v{\in}\widetilde{v}'(u)\, A(u, v, z) \leftrightarrow \exists\tilde{v}\forall u'\exists z\forall u{\in}u'\, A\big(u, \tilde{v}(u), z\big).$$

Proof. The backward direction is trivial: let $\widetilde{v}'(u) = \{\tilde{v}(u)\}$. To prove the forward direction, fix $\widetilde{v}'$ and let

$$\Omega = \prod_u \widetilde{v}'(u).$$

Then Ω is the set of all cross-sections of $\widetilde{v}'$. Each $\widetilde{v}'(u)$ is a finite set; give it the discrete topology, so it is compact. Give Ω the product topology, so it is compact by Tychonov's theorem.

By hypothesis, for each u' there exists an element $\tilde{v}_{u'}$ of Ω such that we have $\exists z \forall u {\in} u' A\big(u, \tilde{v}(u), z\big)$ (let $\tilde{v}_{u'}$ be arbitrary outside u'). The u' are a directed set under inclusion, so $u' \mapsto \tilde{v}_{u'}$ is a net in Ω. Since Ω is compact, this net has a limit point $\tilde{v}$, which has the desired property. □

Corollary 2 *(Dual form of cross-section) Again let A be internal. Then*

$$\forall \widetilde{v'} \exists u' \forall z \exists v {\in} \widetilde{v'}(u)\, A(u, v, z) \leftrightarrow \forall \tilde{v} \exists u' \forall z \exists u {\in} u'\, A\big(u, \tilde{v}(u), z\big).$$

Theorem 3 *(Detachment) If* A° *and* $(A \to B)^\circ$ *are theorems, so is* B°.

Proof. Let y be the free variables common to A and B, let w be the remaining free variables of A, and let z be the remaining free variables of B. We shall derive a contradiction from A°, $(A \to B)^\circ$, and $\neg(B^\circ)$. These formulas are

(1) $\forall u_0 \exists v'_0 \forall w_0 y_0 \exists v_0 {\in} v'_0\, A^\bullet(u_0, v_0, w_0, y_0)$

(2) $\forall \tilde{v}_1 r_1 \exists u'_1 s'_1 \forall w_1 y_1 z_1 \exists u_1 {\in} u'_1 s_1 {\in} s'_1\, \big[\neg A^\bullet\big(u_1, \tilde{v}_1(u_1), w_1, y1\big) \vee B^\bullet(r_1, s_1, y_1, z_1)\big]$

(3) $\exists r_2 \forall s'_2 \exists y_2 z_2 \forall s_2 {\in} s'_2\, \neg B^\bullet(r_2, s_2, y_2, z_2)$.

Fix r_2 and let $r_1 = r_2$. (That is, delete $\exists r_2$ in (3), replace the variables r_2 by constants also denoted by r_2, delete $\forall r_1$ in (2) and replace each occurrence of r_1 by r_2.) Now apply choice to (1) to pull out v'_0 as an existentially quantified function $\widetilde{v'_0}$ of u_0. Notice that (2) has the form of the right hand side of the dual form of the cross-section lemma, so replace it by the left hand side. In this way we obtain

(1′) $\exists \widetilde{v'_0} \forall u_0 z_0 \exists v_0 {\in} \widetilde{v'_0}(u_0)\, A^\bullet(u_0, v_0, w_0, y_0)$

(2′) $\forall \widetilde{v'_1} \exists u'_1 s'_1 \forall z_1 \exists u_i {\in} u'_1 s_1 {\in} s'_1 \forall v_1 {\in} \widetilde{v'_1}(u_1, s_1)\, \big[\neg A^\bullet(u_1, v_1, w_1, y_1) \vee B^\bullet(r_2, s_1, y_1, z_1)\big]$

(3′) $\forall s'_2 \exists z_2 \forall s_2 {\in} s'_2\, \neg B^\bullet(r_2, s_2, y_2, z_2)$.

Fix $\widetilde{v'_0}$; let $\widetilde{v'_1}$ be defined by $\widetilde{v'_1}(u, s) = \widetilde{v'_0}(u)$ for all u and s; fix u'_1 and s'_1; let $s'_2 = s'_1$; fix y_2 and z_2; let $y_1 = y_2$ and $z_1 = z_2$, and let w_1 be arbitrary; let $w_0 = w_1$ and $z_0 = z_2$; fix u_1 and s_1; let $u_0 = u_1$ and $s_2 = s_1$; fix v_0; let $v_1 = v_0$. Then we have

(1″) $A^\bullet(u_1, v_0, w_1, y_2)$

(2″) $\neg A^\bullet(u_1, v_0, w_1, y_2) \vee B^\bullet(r_2, s_1, y_2, z_2)$

(3″) $\neg B^\bullet(r_2, s_1, y_2, z_2)$,

which is a contradiction. □

I have sketched the main step in a syntactical proof of the conservativity of IST over ZFC. But a better argument is needed, one that gives a practical method for converting external proofs into internal proofs. This should be possible. Whenever one uses an ideal object, such as an infinitesimal or a finite set of unlimited cardinal, it depends on the free variables in only a finite way. I expect it to be possible to develop a syntactical procedure that examines the external proof and establishes this dependence in an internal fashion.

Part II. General

Much of mathematics is intrinsically complex, and there is delight to be found in mastering complexity. But there can also be an extrinsic complexity arising from unnecessarily complicated ways of expressing intuitive mathematical ideas. Heretofore nonstandard analysis has been used primarily to simplify proofs of *theorems*. But it can also be used to simplify *theories*. There are several reasons for doing this. First and foremost is the aesthetic impulse, to create beauty. Second and very important is our obligation to the larger scientific community, to make our theories more accessible to those who need to use them. To simplify theories we need to have the courage to leave results in simple, external form—fully to embrace nonstandard analysis as a new paradigm for mathematics.

Much can be done with what may be called *minimal nonstandard analysis*. Introduce a new predicate symbol *standard* applying *only to natural numbers*, with the axioms:

(1) 0 is standard,
(2) if n is standard then $n + 1$ is standard,
(3) there exists a nonstandard number,
(4) if $\mathrm{A}(0)$ and if for all standard n whenever $\mathrm{A}(n)$ then $\mathrm{A}(n+1)$, then for all standard n, $\mathrm{A}(n)$.

A prime example of unnecessary complication in mathematics is, in my opinion, Kolmogorov's foundational work on probability expressed in terms of Cantor's set theory and Lebesgue's measure theory. A beautiful treatise using these methods is [1], but some probabilists find the alternate treatment in [3] more transparent. Please do not misunderstand what I am saying; these remarks are not polemical. Simplicity is not the only virtue in mathematics and I wish in no way to discount other approaches to the use of nonstandard analysis in probability. I just want to encourage a few others to explore the possibility of using minimal nonstandard analysis in probability theory, functional analysis, differential geometry, or whatever field engages your passion.

In this spirit I shall give a few examples from [3]. A *finite probability space* is a finite set Ω and a strictly positive function pr on Ω such that

$$\sum_{\omega\in\Omega} \mathrm{pr}(\omega) = 1.$$

(The set Ω is finite but we do not require its cardinal to be standard.) An *event* is a subset M of Ω, and its *probability* is

$$\Pr(M) = \sum_{\omega\in M} \mathrm{pr}(\omega).$$

A *random variable* is a function $x : \Omega \to \mathbb{R}$, and its *expectation* is

$$\mathbf{E}x = \sum_{\omega\in\Omega} x(\omega)\mathrm{pr}(\omega).$$

If $a \in \mathbb{R}$, we define

$$x^{(a)}(\omega) = \begin{cases} x(\omega), & |x(\omega)| \le a \\ 0, & \textit{otherwise.} \end{cases}$$

A random variable x *is* L^1 in case

$$\mathbf{E}|x - x^{(a)}| \simeq 0 \text{ for all } a \simeq \infty.$$

Theorem 4 *(Radon-Nikodym) A random variable x is L^1 if and only if we have $\mathbf{E}|x| \ll \infty$ and for all events M with $\Pr(M) \simeq 0$ we have $\mathbf{E}|x|\chi_M \simeq 0$.*

Proof. Suppose that x is L^1. We have $\mathbf{E}|x - x^{(a)}| \le 1$ for all $a \simeq \infty$, so by overspill this is true for some $a \ll \infty$. Then $\mathbf{E}|x| \le \mathbf{E}|x - x^{(a)}| + \mathbf{E}|x^{(a)}| \le 1 + a \ll \infty$. Now let $\Pr(M) \simeq 0$. Let $a \simeq \infty$ be such that $a\Pr(M) \simeq 0$—for example, let $a = 1/\sqrt{\Pr(M)}$. Then

$$\mathbf{E}|x|\chi_M \le \mathbf{E}|x^{(a)}|\chi_M + \mathbf{E}|x - x^{(a)}|\chi_M \le a\Pr(M) + \mathbf{E}|x - x^{(a)}| \simeq 0.$$

Conversely, suppose that $\mathbf{E}|x| \ll \infty$ and that for all M with $\Pr(M) \simeq 0$ we have $\mathbf{E}|x|\chi_M \simeq 0$. Let $a \simeq \infty$ and let $M = \{|x| > a\}$. Then we have $\Pr(M) \le \mathbf{E}|x|/a \simeq 0$, so that $\mathbf{E}|x|\chi_M \simeq 0$; that is, $\mathbf{E}|x - x^{(a)}| \simeq 0$. □

A property holds *almost everywhere* (a.e.) in case for all $\varepsilon \gg 0$ there in an event N with $\Pr(N) \le \varepsilon$ such that the property holds everywhere except possibly on N.

Theorem 5 *(Lebesgue) If x and y are L^1 and $x \simeq y$ a.e., then $\mathbf{E}x \simeq \mathbf{E}y$.*

Proof. Let $z = x - y$. Then $z \simeq 0$ a.e. For all $\lambda \gg 0$ we have $\Pr(\{|z| \geq \lambda\}) \leq \lambda$, so by overspill this holds for some infinitesimal λ. But then

$$|z| \leq |z|\chi_{\{|z| \geq \lambda\}} + \lambda$$

and since z is L^1, $\mathbf{E}|z| \simeq 0$ by the previous theorem. Hence $\mathbf{E}x \simeq \mathbf{E}y$. □

One final example, useful in probability theory but more general. Let I be a finite subset of $[0, 1]$ of the form

$$0 = t_0 < t_1 \cdots < t_{\nu-1} < t_\nu = 1$$

such that $t_\mu \simeq t_{\mu+1}$ for all $0 \leq \mu < \nu$. To the naked eye, I looks just like $[0, 1]$. Although I is finite, it is "uncountable" in the following sense:

Theorem 6 *(Cantor) For any sequence $x : \mathbb{N} \to I$ there exists $t \in I$ such that t is not infinitely close to any x_n with n standard.*

Proof. Construct t_0 by changing the nth decimal digit of x_n, so that $|t_0 - x_n| \geq 10^{-n}$ for all n. Let t be the greatest element of I that is less than t_0; then t is in I and has the desired property. □

References

[1] PATRICK BILLINGSLEY, *Convergence of Probability Measures*, second edition, Wiley & Sons, Inc., New York (1999).

[2] EDWARD NELSON, "Internal set theory: A new approach to nonstandard analysis", Bull. Amer. Math. Soc., **83** (1987) 1165–1198.

[3] EDWARD NELSON, *Radically Elementary Probability Theory*, Annals of Mathematics Studies #117, Princeton University Press, Princeton, New Jersey (1987). `http://www.math.princeton.edu/~nelson/books.html`