CHAPTER 1

Internal Set Theory

Ordinarily in mathematics, when one introduces a new concept one defines it. For example, if this were a book on "blobs" I would begin with a definition of this new predicate: x is a *blob* in case x is a topological space such that no uncountable subset is Hausdorff. Then we would be all set to study blobs. Fortunately, this is not a book about blobs, and I want to do something different. I want to begin by introducing a new predicate "standard" to ordinary mathematics without defining it.

The reason for not defining "standard" is that it plays a syntactical, rather than a semantic, role in the theory. It is similar to the use of "fixed" in informal mathematical discourse. One does not define this notion, nor consider the set of all fixed natural numbers. The statement "there is a natural number bigger than any fixed natural number" does not appear paradoxical. The predicate "standard" will be used in much the same way, so that we shall assert "there is a natural number bigger than any standard natural number." But the predicate "standard"—unlike "fixed"—will be part of the formal language of our theory, and this will allow us to take the further step of saying, "call such a natural number, one that is bigger than any standard natural number, unlimited."

We shall introduce axioms for handling this new predicate "standard" in a consistent way. In doing so, we do not enlarge the world of mathematical objects in any way, we merely construct a richer language to discuss the same objects as before. In this way we construct a theory extending ordinary mathematics, called Internal Set Theory[1] that axiomatizes a portion of Abraham Robinson's nonstandard analysis. In this construction, nothing in ordinary mathematics is changed.

[1]It was first presented in [Ne] Edward Nelson, "Internal set theory: A new approach to nonstandard analysis," *Bulletin American Mathematical Society* **83** (1977), 1165–1198. A good introductory account is [Rt] Alain Robert, "Analyse non standard," Presses polytechniques romandes, EPFL Centre Midi, CH–1015 Lausanne, 1985; translated by the author as "Nonstandard Analysis," Wiley, New York, 1988.

1.1 External concepts

Let us begin by adjoining to ordinary mathematics a new predicate called *standard*. Using this new predicate, we can introduce new notions. For example, we can make the following definitions, in which the variables range over $\mathbf{R}$, the set of all real numbers:

- x is *infinitesimal* in case for all standard $\varepsilon > 0$ we have $|x| \le \varepsilon$,
- x is *limited* in case for some standard r we have $|x| \le r$,
- $x \simeq y$ (x is *infinitely close to* y) in case $x - y$ is infinitesimal,
- $x \ll y$ (x is *strongly less than* y) in case $x < y$ and $x \not\simeq y$,
- $x \gg y$ (x is *strongly greater than* y) in case $y \ll x$,
- $x \simeq \infty$ in case $x \ge 0$ and x is unlimited (i.e., not limited),
- $x \simeq -\infty$ in case $-x \simeq \infty$,
- $x \ll \infty$ in case $x \not\simeq \infty$,
- $x \gg -\infty$ in case $x \not\simeq -\infty$,
- $x \lesssim y$ (x is *nearly less than* y) in case $x \le y + \alpha$ for some infinitesimal α,
- $x \gtrsim y$ (x is *nearly greater than* y) in case $y \lesssim x$.

A formula of ordinary mathematics—one that does not involve the new predicate "standard", even indirectly—is called *internal*; otherwise, a formula is called *external*. The eleven definitions given above are all external formulas.

Notice that the adjectives "internal" and "external" are metamathematical concepts—they are properties of formulas and do not apply to sets.

One of the basic principles of ordinary mathematics, or *internal mathematics* as I shall call it from now on, is the subset axiom. This asserts that if X is a set then there is a set S, denoted by $\{\, x \in X : \mathrm{A} \,\}$, such that for all x we have $x \in S \leftrightarrow x \in X \;\&\; \mathrm{A}$. Here A is a formula of internal mathematics. Usually it will have x as a free variable, but it may have other free variables as well. When we want to emphasize its dependence on x, we write it as $\mathrm{A}(x)$. Nothing in internal mathematics has been changed by introducing the new predicate, so the subset axiom continues to hold. But nothing in internal mathematics refers to the new predicate, so nothing entitles us to apply the subset axiom to external formulas. For example, we cannot prove that there exists a set I such that $x \in I \leftrightarrow x \in \mathbf{R} \;\&\; x$ is infinitesimal. The notation $\{\, x \in \mathbf{R} : x \simeq 0 \,\}$ is not allowed; it is an example of *illegal set formation*. Only internal formulas can be used to define subsets. (Nevertheless, in Chapter 4 we shall find a way to introduce so-called external sets.)

Certain set formations that might at first sight appear to be illegal are perfectly legitimate. For example, suppose that we have an infinitesimal $x > 0$. Then we can form the closed interval $[-x, x]$ consisting of all y such that $-x \leq y \leq x$. This is simply because we already know that for *any* $x > 0$ we can form the set $[-x, x]$.

Exercises for Section 1.1

1. Let n be a nonstandard natural number. Can one form the set of all natural numbers k such that $k \leq n$? Is this set finite?
2. Can one form the set of all limited real numbers?
3. Can one prove that every standard positive real number is limited?
4. Can one form the set of all standard real numbers x such that $x^2 \leq 1$?
5. Assume that 1 is standard. Can one form the set of all limited real numbers x such that $x^2 \leq 1$?
6. Can one prove that the sum of two infinitesimals is infinitesimal?

1.2 The transfer principle

We cannot yet prove anything of interest involving "standard" because we have made no assumptions about it. Our first axiom is the *transfer principle* (T).

The notation $\forall^{\mathrm{st}}$ means "for all standard", and $\exists^{\mathrm{st}}$ means "there exists a standard". Let A be an internal formula whose only free variables are x, t_1, ..., t_{n}. Then the transfer principle is

$$\forall^{\mathrm{st}} t_1 \cdots \forall^{\mathrm{st}} t_{\mathrm{n}} [\forall^{\mathrm{st}} x \mathrm{A} \leftrightarrow \forall x \mathrm{A}]. \tag{1.1}$$

We may think of the t_1, ..., t_{n} as parameters; we are mainly interested in x. Then the transfer principle asserts that if we have an internal formula A, and all the parameters have standard values, and if we know that A holds for all standard x, then it holds for all x. (The converse direction is trivial, and we could have stated (T) with just $\rightarrow$ instead of $\leftrightarrow$.)

The intuition behind (T) is that if something is true for a fixed, but arbitrary, x then it is true for all x.

Notice that two formulas are equivalent if and only if their negations are. But we have $\neg\forall x \neg \mathrm{A} \leftrightarrow \exists x \mathrm{A}$, so if we apply (T) to $\neg\mathrm{A}$ we obtain the *dual form of the transfer principle*:

$$\forall^{\mathrm{st}} t_1 \cdots \forall^{\mathrm{st}} t_{\mathrm{n}} [\exists^{\mathrm{st}} x \mathrm{A} \leftrightarrow \exists x \mathrm{A}]. \tag{1.2}$$

Let us write $\mathrm{A} \equiv \mathrm{B}$ (A is *weakly equivalent* to B) to mean that for all standard values of the free variables in the formulas, we have $\mathrm{A} \leftrightarrow \mathrm{B}$. Then we can rewrite the two forms of (T) as $\forall^{\mathrm{st}} x \mathrm{A} \equiv \forall x \mathrm{A}$ and $\exists^{\mathrm{st}} x \mathrm{A} \equiv \exists x \mathrm{A}$ whenever A is an internal formula. Applying these rules repeatedly, we see that any internal formula A is weakly equivalent to the formula A^{st} obtained by replacing each $\forall$ by $\forall^{\mathrm{st}}$ and $\exists$ by $\exists^{\mathrm{st}}$. Then

$$t_1, \ldots, t_{\mathrm{n}} \text{ are standard } \rightarrow \mathrm{A}^{\mathrm{st}},$$

where t_1, ..., t_{n} are the free variables in A, is called the *relativization of* A *to the standard sets.*

Consider an object, such as the empty set $\emptyset$, the natural numbers $\mathbf{N}$, or the real numbers $\mathbf{R}$, that can be described uniquely within internal mathematics. That is, suppose that there is an internal formula $\mathrm{A}(x)$ whose only free variable is x such that we can prove existence $\exists x \mathrm{A}(x)$ and uniqueness $\mathrm{A}(x_1)\ \&\ \mathrm{A}(x_2) \rightarrow x_1 = x_2$. By the dual form of transfer, $\exists^{\mathrm{st}} x \mathrm{A}(x)$; so by uniqueness, the x such that $\mathrm{A}(x)$ holds is standard. For example, let $\mathrm{A}(x)$ be $\forall y[y \notin x]$. There is a unique set, the empty set $\emptyset$, that satisfies $\mathrm{A}(x)$. Therefore $\emptyset$ is standard. The formulas describing $\mathbf{N}$ and $\mathbf{R}$ are longer, but by the same reasoning, $\mathbf{N}$ and $\mathbf{R}$ are standard. Any object that can be uniquely described within internal mathematics is standard: the real numbers 0, 1, and π, the Hilbert space $L^2(\mathbf{R}, dx)$ where dx is Lebesgue measure, the first uncountable ordinal, and the loop space of the fifteen dimensional sphere are all standard. The real number 10^{-100} is standard, so if x is infinitesimal then $|x| \leq 10^{-100}$.

The same reasoning applies to internal formulas A(x) containing parameters—provided the parameters have standard values. For example, if t is a standard real number then so is $\sin t$ (let $\mathrm{A}(x, t)$ be $x = \sin t$) and if X is a standard Banach space so is its dual.

As an example of transfer, we know that for all real $x > 0$ there is a natural number n such that $nx \geq 1$; therefore, for all standard $x > 0$ there is a standard n such that $nx \geq 1$. But suppose that $x > 0$ is infinitesimal. Do we know that there is a natural number n such that $nx \geq 1$? Of course; we already know this for *all* $x > 0$. But if we try to argue as follows—"there is an n such that $nx \geq 1$; therefore, by the dual form of transfer, there is a standard n such that $nx \geq 1$"—then we have made an error: transfer is only valid for the *standard* values of the parameters (in this case x) in the formula. This is an example of *illegal transfer*. It is the most common error in learning nonstandard analysis. Before applying transfer, one must make sure that any parameters in the formula—even those that may be implicit in the discussion—have standard values.

Another form of illegal transfer is the attempt to apply it to an external formula. For example, consider "for all standard natural numbers n, the number n is limited; by transfer, all natural numbers are limited". This is incorrect. Before applying transfer, one must check two things: that the formula is internal and that all parameters in it have standard values.

Exercises for Section 1.2

1. Can one prove that the sum of two infinitesimals is infinitesimal?
2. If r and s are limited, so are $r+s$ and rs.
3. If $x \simeq 0$ and $|r| \ll \infty$, then $xr \simeq 0$.
4. If $x \neq 0$, then x is infinitesimal if and only if $1/x$ is unlimited.
5. Is it true that the infinitesimals are a maximal ideal in the integral domain of limited real numbers? What does the quotient field look like?
6. Consider the function $f\colon x \mapsto x^2$ on a closed interval. Then f is bounded, and since it is standard it has a standard bound.—Is this reasoning correct?
7. Let x and y be standard with $x \lesssim y$. Prove that $x \leq y$.

1.3 The idealization principle

So far, we have no way to prove that any nonstandard objects exist. Our next assumption is the *idealization principle* (I).

The notation $\forall^{\mathrm{stfin}}$ means "for all standard finite sets", and $\exists^{\mathrm{stfin}}$ means "there exists a standard finite set such that". Also, $\forall x{\in}X$ means "for all x in X", and $\exists x{\in}X$ means "there exists x in X such that". Let A be an internal formula. Then the idealization principle is

$$\forall^{\mathrm{stfin}} x' \exists y \forall x{\in}x' \mathrm{A} \leftrightarrow \exists y \forall^{\mathrm{st}} x \mathrm{A}. \tag{1.3}$$

There are no particular pitfalls connected with this assumption: we must just be sure that A is internal. It can contain free variables in addition to x and y (except for x').

The intuition behind (I) is that we can only fix a finite number of objects at a time. To say that there is a y such that for all fixed x we have A is the same as saying that for any fixed finite set of x's there is a y such that A holds for all of them.

As a first application of the idealization principle, let A be the formula $y \neq x$. Then for every finite set x', and so in particular for every standard finite set x', there is a y such that for all x in x' we have $y \neq x$. Therefore, there exists a nonstandard y. The same argument works when x and y are restricted to range over any infinite set. In other words, every infinite set contains a nonstandard element. In particular, there exists a nonstandard natural number.

By transfer, we know that 0 is standard, and we know that if n is standard then so is $n+1$. Do we have a contradiction here? Why can't we say that by induction, all natural numbers are standard? The induction theorem says this: if S is a subset of $\mathbf{N}$ such that 0 is in S and such that whenever n is in S then $n+1$ is in S, then $S = \mathbf{N}$. So to apply induction to prove that every natural number is standard, we would need a set S such that n is in S if and only if n is standard, and this we don't have.

As a rough rule of thumb, until one feels at ease with nonstandard analysis, it is best to apply the familiar rules of internal mathematics freely to elements, but to be careful when working with sets of elements. (From a foundational point of view, everything in mathematics is a set. For example, a real number is an equivalence class of Cauchy sequences of rational numbers. Even a natural number is a set: the number 0 is the empty set, the number 1 is the set whose only element is 0, the number 2 is the set whose only elements are 0 and 1, etc. When I refer to "elements" or "objects" rather than to sets, only a psychological distinction is intended.)

We can prove that certain subsets of $\mathbf{R}$ and $\mathbf{N}$ do not exist.

Theorem 1. *There does not exist S_1, S_2, S_3, S_4, or S_5 such that, for all n in $\mathbf{N}$ and x in $\mathbf{R}$, we have $n \in S_1 \leftrightarrow n$ is standard, $n \in S_2 \leftrightarrow n$ is nonstandard, $x \in S_3 \leftrightarrow x$ is limited, $x \in S_4 \leftrightarrow x$ is unlimited, or $x \in S_5 \leftrightarrow x$ is infinitesimal.*

Proof. As we have seen, the existence of S_1 would violate the induction theorem. If S_2 existed we could take $S_1 = \mathbf{N} \setminus S_2$, if S_3 existed we could take $S_1 = \mathbf{N} \cap S_3$, if S_4 existed we could take $S_3 = \mathbf{R} \setminus S_4$, and if S_5 existed we could take $S_4 = \{\, x \in \mathbf{R} : 1/x \in S_5 \,\}$. ∎

This may seem like a negative result, but it is frequently used in proofs. Suppose that we have shown that a certain internal property $\mathrm{A}(x)$ holds for every infinitesimal x; then we automatically know that $\mathrm{A}(x)$ holds for some non-infinitesimal x, for otherwise we could let S_5 be the set $\{\, x \in \mathbf{R} : \mathrm{A}(x) \,\}$. This is called *overspill*.

By Theorem 1 there is a non-zero infinitesimal, for otherwise we could let $S_5 = \{0\}$. We can also see this directly from (I). We denote by $\mathbf{R}^+$ the set of all strictly positive real numbers. For every finite subset x' of $\mathbf{R}^+$ there is a y in $\mathbf{R}^+$ such that $y \leq x$ for all x in x'. By (I), there is a y in $\mathbf{R}^+$ such that $y \leq x$ for all standard x in $\mathbf{R}^+$. That is, there exists an infinitesimal $y > 0$.

Notice that in this example, to say that y is smaller than every element of x' is the same as saying that y is smaller than the least element of x', so the finite set really plays no essential role. This situation occurs so frequently that it is worth discussing it in a general context in which the idealization principle takes a simpler form. Recall that a *directed set* is a set D together with a transitive binary relation $\prec$ such that every pair of elements in D has an upper bound. For example, $\mathbf{R}^+$ is a directed set with respect to $\geq$. (In this example, the minimum of two elements is an upper bound for them.) Let A be a formula with the free variable x restricted to range over the directed set D. We say that A *filters* in x (with respect to D and $\prec$) in case one can show that whenever $x \prec z$ and $\mathrm{A}(z)$, then $\mathrm{A}(x)$. (Here $\mathrm{A}(z)$ is the formula obtained by substituting z for each free occurrence of x in $\mathrm{A}(x)$, with the understanding that z is not a bound variable of A.) Suppose that the internal formula A filters in x. Then the idealization principle takes the simpler form

$$\forall^{\mathrm{st}} x \exists y \mathrm{A} \leftrightarrow \exists y \forall^{\mathrm{st}} x \mathrm{A}. \tag{1.4}$$

That is, we can simply interchange the two quantifiers.

The idealization principle also has a dual form. If A is internal, then

$$\exists^{\mathrm{stfin}} x' \forall y \exists x {\in} x' \mathrm{A} \leftrightarrow \forall y \exists^{\mathrm{st}} x \mathrm{A}. \tag{1.5}$$

We say that A *cofilters* in x in case we can show that whenever $x \prec z$ and $\mathrm{A}(x)$, then $\mathrm{A}(z)$. Then (1.5) takes the simpler form:

$$\exists^{\mathrm{st}} x \forall y \mathrm{A} \leftrightarrow \forall y \exists^{\mathrm{st}} x \mathrm{A}. \tag{1.6}$$

In practice, there is no need to make the distinction between filtering and cofiltering, and I may say "filters" when "cofilters" is correct.

As an example of the dual form, suppose that $f\colon \mathbf{R} \to \mathbf{R}$ is such that every value of f is limited. Then f is bounded, and even has a standard bound. To see this, use (1.6), where the filtering relation is $\leq$ on the values of the function. We know that for all y there exists a standard x such that $|f(y)| \leq x$. Hence there is a standard x such that $|f(y)| \leq x$ for all y.

Theorem 2. *Let S be a set. Then S is a standard finite set if and only if every element of S is standard.*

Proof. Let S be a standard finite set, and suppose that it contains a nonstandard y. Then there exists a y in S such that for all standard x we have $y \neq x$, so by (I), for all standard finite sets x', and in particular for S, there is a y in S such that for all x in S we have $y \neq x$. This is a contradiction, so every element of S is standard.

Conversely, suppose that every element of S is standard. We already know that S must be finite, because every infinite set contains a nonstandard element. Let x and y range over S and apply the dual form of idealization to the formula $x = y$. We know that for all y there is a standard x with $x = y$, so there is a standard finite set x' such that for all y there is an x in x' with $x = y$. Then $x' \supseteq S$; that is, $S \in \wp x'$ where $\wp$ denotes the power set. But x' is standard, so by (T) $\wp x'$ is standard. Also, x' is finite, so $\wp x'$ is a standard finite set. By the forward direction of the theorem established in the preceding paragraph, every element of $\wp x'$ is standard, so S is standard.[2] ∎

This has the following corollary.

Theorem 3. *Let n and k be natural numbers, with n standard and $k \leq n$. Then k is standard.*

Proof. Let $S = \{ k \in \mathbf{N} : k \leq n \}$. By (T), S is standard. It is finite, so all of its elements are standard. ∎

Therefore we can picture the natural numbers as lying on a tape, with the standard numbers to the left and the nonstandard numbers to the right. The demarcation between the two portions is strange: the left portion is not a set, and neither is the right. I want to emphasize that we did not start with the left portion and invent a new right portion to be tacked on to it—we started with the whole tape, the familiar set $\mathbf{N}$ of all natural numbers, and invented a new way of looking at it.

Theorem 4. *There is a finite set that contains every standard object.*

Proof. This is easy. Just apply (I) to the formula $x \in y$ & y is finite. ∎

If we think of "standard" semantically—with the world of mathematical objects spread out before us, some bearing the label "standard" and others not—then these results violate our intuition about finite sets. But recall what it means to say that X is a finite set. This is an internal

[2] I am grateful to Will Schneeberger for pointing out a gap in my first proof of this theorem.

notion, so it means what it has always meant in mathematics. What has it always meant? There are two equivalent characterizations. For n in $\mathbf{N}$, we let $I_n = \{\, k \in \mathbf{N} : k < n \,\}$. Then a set X is finite if and only if there is a bijection of X with I_n for some n. There is also the Dedekind characterization: X is finite if and only if there is no bijection of X with a proper subset of itself. Consider the set I_n where n is a nonstandard natural number. This certainly satisfies the first property. Nothing was said about n being standard; this cannot even be formulated within internal mathematics. But does it satisfy the Dedekind property? Suppose that we send each nonstandard element of I_n to its predecessor and leave the standard elements alone; isn't this a bijection of I_n with I_{n-1}? No, its definition as a function would involve illegal set formation.—Perhaps it is fair to say that "finite" does not mean what we have always thought it to mean. What have we always thought it to mean? I used to think that I knew what I had always thought it to mean, but I no longer think so. In any case, intuition changes with experience. I find it intuitive to think that very, very large natural numbers and very, very small strictly positive real numbers were there all along, and now we have a suitable language for discussing them.

Exercises for Section 1.3

1. A natural number is standard if and only if it is limited.

2. Does there exist an unlimited prime?

3. What does the decimal expansion of an infinitesimal look like?

4. Let H be a Hilbert space. Does there exist a finite dimensional subspace containing all of its standard elements? Is it closed?

5. Suppose that the S_n, for n in $\mathbf{N}$, are a sequence of disjoint sets with union S, and suppose that every element x in S is in S_n for some standard n. Can one show that all but a finite number of them are empty? Can one show that all but a standard finite number of them are empty?

6. Prove *Robinson's lemma*: Let $n \mapsto a_n$ be a sequence such that $a_n \simeq 0$ for all standard n. Then there is an unlimited N such that $a_n \simeq 0$ for all $n \le N$.

7. Let $n \mapsto a_n$ be a sequence such that $a_n \ll \infty$ for all standard n. Is there an unlimited N such that $a_n \ll \infty$ for all $n \le N$?

8. We have been saying "let x range over X" to mean that each $\forall x\mathrm{A}$ should be replaced by $\forall x[x \in X \to \mathrm{A}]$ and each $\exists x\mathrm{A}$ should be replaced by $\exists x[x \in X \;\&\; \mathrm{A}]$. Let $\wp_{\mathrm{fin}} X$ be the set of all non-empty finite subsets of X. Show that if x ranges over X, then (I) can be written as $\forall^{\mathrm{st}} x' \exists y \forall x{\in}x' \mathrm{A} \leftrightarrow \exists y \forall^{\mathrm{st}} x \mathrm{A}$ where x' ranges over $\wp_{\mathrm{fin}} X$.

1.4 The standardization principle

Our final assumption about the new predicate "standard" is the *standardization principle* (S). It states that

$$\forall^{\mathrm{st}} X \exists^{\mathrm{st}} Y \forall^{\mathrm{st}} z [z \in Y \leftrightarrow z \in X \;\&\; \mathrm{A}]. \tag{1.7}$$

Here A can be any formula (not containing Y), external or internal. It may contain parameters (free variables in addition to z and X).

The intuition behind (S) is that if we have a fixed set, then we can specify a fixed subset of it by giving a criterion for judging whether each fixed element is a member of it or not.

Two sets are equal if they have the same elements. By (T), two standard sets are equal if they have the same standard elements. Therefore, the standard set Y given by (S) is unique. It is denoted by ${}^{\mathrm{S}}\{\, z \in X : \mathrm{A} \,\}$, which may be read as "the standard set whose standard elements are those standard elements of X such that A holds". Unfortunately, any shortening of this cumbersome phrase is apt to be misleading. For standard elements z, we have a direct criterion for z to be an element of ${}^{\mathrm{S}}\{\, z \in X : \mathrm{A} \,\}$, namely that $\mathrm{A}(z)$ hold. But for nonstandard elements z, this is not so. It may happen that z is in ${}^{\mathrm{S}}\{\, z \in X : \mathrm{A} \,\}$ but $\mathrm{A}(z)$ does not hold, and conversely $\mathrm{A}(z)$ may hold without z being in ${}^{\mathrm{S}}\{\, z \in X : \mathrm{A} \,\}$. For example, let $X = {}^{\mathrm{S}}\{\, z \in \mathbf{R} : z \simeq 0 \,\}$. Then $X = \{\, 0 \,\}$ since 0 is the only standard infinitesimal. Thus we can have $z \simeq 0$ without z being in X. Let $Y = {}^{\mathrm{S}}\{\, z \in \mathbf{R} : |z| \ll \infty \,\}$. Then $Y = \mathbf{R}$ since every standard number is limited. Thus we can have $z \in Y$ without z being limited.

The standardization principle is useful in making definitions. Let x range over the standard set X. When we make a definition of the form: "for x standard, x is *something-or-other* in case a certain property holds", this is understood to mean the same as "x is *something-or-other* in case $x \in {}^{\mathrm{S}}\{\, x \in X :$ a certain property holds$\,\}$". For example, let $f \colon \mathbf{R} \to \mathbf{R}$ and let us say that for f standard, f is *nice* in case every value of f is limited. Thus f is nice if and only if $f \in N$, where we let $N = {}^{\mathrm{S}}\{\, f \in \mathbf{R}^{\mathbf{R}} :$ every value of f is limited$\,\}$. (The notation X^Y signifies the set of all functions from Y to X.) I claim that f is nice if and only if f is bounded. To prove this, we may, by (T), assume that f is standard. We already saw, in the previous section, that if every value of f is limited, then f is bounded. Conversely, let f be bounded. Then by (T) it has a standard bound, and so is nice. This proves the claim. But, one might object, the first transfer is illegal because "nice" was defined externally. The point is this: being nice is the same as being in the standard set N, so the transfer is legal. In complete detail, (T) tells us that for all standard N we have

$$\forall^{\mathrm{st}} f[f \in N \leftrightarrow f \text{ is bounded}] \rightarrow \forall f[f \in N \leftrightarrow f \text{ is bounded}].$$

In other words, this kind of definition by means of (S) is a way of defining, somewhat implicitly, an *internal* property.

Let x and y range over a set V, let $\tilde{y}$ range over V^V, and let A be internal. Then

$$\forall x \exists y \mathrm{A}(x, y) \leftrightarrow \exists \tilde{y} \forall x \mathrm{A}\big(x, \tilde{y}(x)\big).$$

The forward direction is the axiom of choice, and the backward direction is trivial. If we apply transfer to this, assuming that V is standard, we obtain

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

But we can do much better than this. Let A be any formula, external or internal. Then

$$\forall^{\mathrm{st}} x \exists^{\mathrm{st}} y \mathrm{A}(x, y) \leftrightarrow \exists^{\mathrm{st}} \tilde{y} \forall^{\mathrm{st}} x \mathrm{A}\big(x, \tilde{y}(x)\big). \tag{1.8}$$

The backward direction is trivial, so we need only consider the forward direction. Suppose first that for all standard x there is a unique standard y such that $\mathrm{A}(x, y)$. Then we can let $\tilde{y} = {}^{\mathrm{S}}\{\, \langle x, y \rangle : \mathrm{A}(x, y) \,\}$. In the general case, let

$$\widetilde{Y} = {}^{\mathrm{S}}\{\, \langle x, Y \rangle : Y = {}^{\mathrm{S}}\{\, y : \mathrm{A}(x, y) \,\} \,\}.$$

Then $\widetilde{Y}$ is a standard set-valued function whose values are non-empty sets, so by the axiom of choice relativized to the standard sets, it has a standard cross-section $\tilde{y}$ of the desired form. We call (1.8) the *functional form of standardization* ($\widetilde{\mathrm{S}}$). It has a dual form:

$$\exists^{\mathrm{st}} x \forall^{\mathrm{st}} y \mathrm{A}(x, y) \leftrightarrow \forall^{\mathrm{st}} \tilde{y} \exists^{\mathrm{st}} x \mathrm{A}\big(x, \tilde{y}(x)\big). \tag{1.9}$$

The requirements in (1.8) and (1.9) are that x and y range over some standard set V and that $\tilde{y}$ range over V^V. This includes the case that x and y range over different standard sets X and Y, with $\tilde{y}$ ranging over Y^X, because we can always let $V = X \cup Y$ and include the conditions $x \in X$ and $y \in Y$ in the formula A.

If X is a set, contained in some standard set V, we let

$${}^{\mathrm{S}}X = {}^{\mathrm{S}}\{\, x \in V : x \in X \,\}.$$

This clearly does not depend on the choice of V, and in practice the requirement that X be contained in some standard set is not restrictive. Then ${}^{\mathrm{S}}X$ is the unique standard set having the same standard elements as X. It is easy to see that if f is a function then so is ${}^{\mathrm{S}}f$.

The theory obtained by adjoining (I), (S), and (T) to internal mathematics is *Internal Set Theory* (IST).

So far, we have not proved any theorems of internal mathematics by these new methods. Here is a first example of that, a theorem of de Brujn and Erdős[3] on the coloring of infinite graphs.

By a *graph* I mean a set G together with a subset R of $G \times G$. We define a *k-coloring* of G, where k is a natural number, to be a function $g\colon G \to \{1, \ldots, k\}$ such that $g(x) \neq g(y)$ whenever $\langle x, y \rangle \in R$. An example is $G = \mathbf{R}^2$ and $R = \{\, \langle x, y \rangle : |x - y| = 1 \,\}$ where $|x - y|$ is the Euclidean metric on $\mathbf{R}^2$. (This graph is known to have a 7-coloring but no 3-coloring.) The theorem asserts that if every finite subgraph of G has a k-coloring, so does G. This is not trivial, because if we color a finite subgraph, we may be forced to go back and change its coloring to color a larger finite subgraph containing it. To prove the theorem, we assume, by (T), that G and k are standard. By (I), there is a finite subgraph F of G containing all its standard elements, and by hypothesis F has a k-coloring f. Let $g = {}^{\mathrm{S}}f$. Since f takes only standard values, every standard element of G is in the domain of g. By (T), every element of G is in the domain of g. To verify that g is a k-coloring, it suffices, by (T), to examine the standard elements, where it agrees with f. This concludes the proof.

Exercises for Section 1.4

1. Show that if f is a function, so is ${}^{\mathrm{S}}f$. What can one say about its domain and range?

2. Deduce (S) from ($\widetilde{\mathrm{S}}$).

3. Define f by $f(x) = t/\pi(t^2 + x^2)$ where $t > 0$ is infinitesimal. What is $\int_{-\infty}^{\infty} f(x)\,dx$? What is ${}^{\mathrm{S}}f$?

4. Establish *external induction:* Let $\mathrm{A}(n)$ be any formula, external or internal, containing the free variable n and possibly other parameters. Suppose that A holds for 0, and that whenever it holds for a standard n it also holds for $n + 1$. Then $\mathrm{A}(n)$ holds for all standard n.

5. Let the Greek variables range over ordinals. Establish *external transfinite induction*: $\exists^{\mathrm{st}}\alpha \mathrm{A}(\alpha) \to \exists^{\mathrm{st}}\beta\big[\mathrm{A}(\beta) \;\&\; \forall^{\mathrm{st}}\gamma[\mathrm{A}(\gamma) \to \beta \le \gamma]\big]$.

6. Deduce (I) from the forward direction of (I).

[3] N. G. de Brujn and P. Erdős, "A colour problem for infinite graphs and a problem in the theory of relations," *Indagationes Mathematicae* **13** (1951), 371–373.

1.5 Elementary topology

In this section I shall illustrate IST with some familiar material. In the following definitions by (S), we assume that $E \subseteq \mathbf{R}$, that $f\colon E \to \mathbf{R}$, and that $x, y \in E$.

- For f and x standard, f is *continuous at* x in case whenever $y \simeq x$ we have $f(y) \simeq f(x)$.
- For f and E standard, f is *continuous on* E in case for all standard x, whenever $y \simeq x$ we have $f(y) \simeq f(x)$.
- For f and E standard, f is *uniformly continuous on* E in case whenever $y \simeq x$ we have $f(y) \simeq f(x)$.
- For E standard, E is *compact* in case for all x in E there is a standard x_0 in E with $x \simeq x_0$.
- For E standard, E is *open* in case for all standard x_0 in E and all z in $\mathbf{R}$, if $z \simeq x_0$ then z is in E.
- For E standard, E is *closed* in case for all x in E and standard x_0 in $\mathbf{R}$, if $x \simeq x_0$ then x_0 is in E.
- For E standard, E is *bounded* in case each of its elements is limited.

It can be shown that these definitions are equivalent to the usual definitions, but rather than worry about that now, let us develop a direct intuition for these formulations by examining them in various familiar situations.

Let us prove that $[0,1]$ is compact. Notice that it is standard. Let x be in $[0,1]$. Then it can be written as $x = \sum_{n=1}^{\infty} a_n 2^{-n}$ where each a_n is 0 or 1, and conversely each number of this form is in $[0,1]$. This binary expansion is determined by a function $a\colon \mathbf{N}^+ \to \{0,1\}$. Let $b = {}^{\mathrm{S}}a$. Then $b_n = a_n$ for all standard n, so if we let $x_0 = \sum_{n=1}^{\infty} b_n 2^{-n}$ then x_0 is standard, $x_0 \simeq x$, and x_0 is in $[0,1]$. Therefore $[0,1]$ is compact. This has the following corollary.

Theorem 5. *Each limited real number is infinitely close to a unique standard real number.*

Proof. Let x be limited. Then $[x]$ is a limited, and therefore standard, integer. Since $x - [x]$ is in the standard compact set $[0,1]$, there is a standard y_0 infinitely close to it, so if we let $x_0 = y_0 + [x]$ then x_0 is standard and infinitely close to x. The uniqueness is clear, since 0 is the only standard infinitesimal. ∎

If x is limited, the standard number that is infinitely close to it is called the *standard part* of x, and is denoted by $\operatorname{st} x$.

Now let us prove that a continuous function f on a compact set E is bounded. By (T), we assume them to be standard. Let $K \simeq \infty$, and let x be in E. It suffices to show that $f(x) \leq K$. There is a standard x_0 in E with $x \simeq x_0$, and since $f(x_0)$ is standard we have $f(x_0) \ll K$. But by the continuity of the standard function f we have $f(x) \simeq f(x_0)$, so $f(x) \leq K$, which concludes the proof. (It may seem like cheating to produce an unlimited bound. By (T), if a standard function f is bounded, then it has a standard bound. But this is a distinction that can be made only in nonstandard analysis.)

Somewhat more ambitiously, let us show that a continuous function f on a compact set E achieves its maximum. Again, we assume them to be standard. By (I), there is a finite subset F of E that contains all the standard points, and the restriction of f to the finite set F certainly achieves its maximum on F at some point x. By compactness and continuity, there is a standard x_0 in E with $x_0 \simeq x$ and $f(x_0) \simeq f(x)$. Therefore $f(x_0) \gtrsim f(y)$ for all y in F. Since every standard y is in F, we have $f(x_0) \gtrsim f(y)$ whenever y is standard, but since both numbers are standard we must have $f(x_0) \geq f(y)$ whenever y is standard. By (T), this holds for all y, and the proof is complete.

The same device can be used to prove that if f is continuous on $[0,1]$ with $f(0) \leq 0$ and $f(1) \geq 0$, then $f(x) = 0$ for some x in $[0,1]$.

The proof that a continuous function on a compact set is uniformly continuous requires no thought; it is a simple verification from the definitions. It is equally easy to prove that a subset of $\mathbf{R}$ is compact if and only if it is closed and bounded (use Theorem 5 for the backward direction).

All of this extends to $\mathbf{R}^n$. Notice that $\mathbf{R}^n$ is standard if and only if n is, so in the definitions by (S), include “for n standard”. For a general metric space $\langle X, d\rangle$, where d is the metric, we define $x \simeq y$ to mean that $d(x,y)$ is infinitesimal. Much of what was done above extends to metric spaces. In definitions by (S), include “for X and d standard”. For X and d standard, a metric space $\langle X, d\rangle$ is *complete* in case for all x, if for all standard $\varepsilon > 0$ there is a standard y with $d(x,y) \leq \varepsilon$, then there is a standard x_0 with $x \simeq x_0$. This is equivalent to the usual definition.

I shall sketch a proof of the Baire category theorem to illustrate an important point about nonstandard analysis. Let $\langle X, d\rangle$ be a complete non-empty metric space, and let the U_n be a sequence of open dense sets with intersection U. We want to show that U is non-empty. We denote the ε-neighborhood of x by $N(\varepsilon, x)$. There are x_1 and $0 < \varepsilon_1 < 1/2$

such that $N(\varepsilon_1, x_1) \subset U_1$. Since U_2 is open and dense, there are x_2 and $0 < \varepsilon_2 < 1/2^2$ such that $N(\varepsilon_2, x_2) \subset U_2 \cap N(\varepsilon_1/2, x_1)$. Continue in this way by induction, constructing a Cauchy sequence x_n that converges to some point x, since X is complete. This x is in each U_n, and so is in U.

This proof is just the usual proof, and that is the important point. Nonstandard analysis is not an alternative to internal mathematics, it is an addition. By the nature of this book, almost all of the material consists of nonstandard analysis, but I emphatically want to avoid giving the impression that it should somehow be separated from internal mathematics. Nonstandard analysis supplements, but does not replace, internal mathematics.

Let X be a standard topological space, and let x be a standard point in it. Then we define the relation $y \simeq x$ to mean that y is in every standard neighborhood of x. The external discussion of continuity and compactness given above for $\mathbf{R}$ extends to this setting.

Let E be a subset of the standard topological space X. Then we define the *shadow* of E, denoted by $^\circ E$, as follows:

$$^\circ E = {}^{\mathrm{S}}\{\, x \in X : y \simeq x \text{ for some } y \text{ in } E \,\}. \tag{1.10}$$

Theorem 6. *Let E be a subset of the standard topological space X. Then the shadow of E is closed.*

Proof. Let z be a standard point of X in the closure of $^\circ E$. Then every open neighborhood of z contains a point of $^\circ E$, so by (T) every standard open neighborhood U of z contains a standard point x of $^\circ E$. But for a standard point x of $^\circ E$, there is a y in E with $y \simeq x$, so y is in U. That is to say, $\forall^{\mathrm{st}} U \exists y [y \in E \cap U]$. The open neighborhoods of z are a directed set under inclusion and this formula filters in U, so by the simplified version (1.4) of (I) we have $\exists y \forall^{\mathrm{st}} U [y \in E \cap U]$; that is, $\exists y {\in} E [y \simeq z]$. Thus z is in $^\circ E$. We have shown that every standard point z in the closure of $^\circ E$ is in $^\circ E$, so by (T) every point z in the closure of $^\circ E$ is in $^\circ E$. Thus $^\circ E$ is closed. ∎

There is a beautiful nonstandard proof of the Tychonov theorem. Let T be a set, let X_t for each t in T be a compact topological space, and let Ω be the Cartesian product $\Omega = \prod_{t \in T} X_t$ with the product topology. We want to show that Ω is compact. By (T), we assume that $t \mapsto X_t$ is standard, so that Ω is also standard. Let ω be in Ω. For all standard t there is a standard point y in X_t such that $y \simeq \omega(t)$, so by $(\widetilde{\mathrm{S}})$ there is a standard η in Ω such that for all standard t we have $\eta(t) \simeq \omega(t)$. By the definition of the product topology, $\eta \simeq \omega$, so Ω is compact.

Exercises for Section 1.5

1. Precisely how is $(\widetilde{\mathrm{S}})$ used in the proof of the Tychonov theorem?

2. Let X be a compact topological space, f a continuous mapping of X onto Y. Show that Y is compact.

3. Is the shadow of a connected set connected?

4. Let x be in $\mathbf{R}$. What is the shadow of $\{x\}$?

5. Let $f(x) = t/\pi(t^2 + x^2)$, where $t > 0$ is infinitesimal. What is the shadow of the graph of f?

6. Let E be a regular polygon of n sides inscribed in the unit circle. What is the shadow of E?

1.6 Reduction of external formulas

It turns out that (with a proviso which I shall discuss shortly) every external formula can be reduced to a weakly equivalent internal formula. This is accomplished by a kind of formal, almost algebraic, manipulation of formulas, using $(\widetilde{\mathrm{S}})$ and (I) to push the external quantifiers $\forall^{\mathrm{st}}$ and $\exists^{\mathrm{st}}$ to the left of the internal quantifiers $\forall$ and $\exists$, and then using (T) to get rid of them entirely. In this way we can show that the definitions made using (S) in the previous section are equivalent to the usual ones. More interestingly, we can reduce the rather weird external theorems that we have proved to equivalent internal form. The proviso, which has only nuisance value, is that whenever we use $(\widetilde{\mathrm{S}})$ to introduce a standard function $\tilde{y}(x)$, then x and y must be restricted to range over a standard set, to give the function $\tilde{y}$ a domain. (Actually, it suffices to make this restriction on x alone.) I will not make this explicit all of the time.

To reduce a formula, first eliminate all external predicates, replacing them by their definitions until only "standard" is left. Even this should be eliminated, replacing "x is standard" by $\exists^{\mathrm{st}}y[y = x]$.

Second, look for an internal quantifier that has some external quantifiers in its scope. (If this never happens, we are ready to apply (T) to obtain a weakly equivalent internal formula.) If the internal quantifier is $\forall$, use $(\widetilde{\mathrm{S}})$ to pull the $\forall^{\mathrm{st}}$'s to the left of the $\exists^{\mathrm{st}}$'s (if necessary, thereby introducing standard functions), where they can then be pulled to the left of $\forall$. If the internal quantifier is $\exists$, proceed by duality.

Third, whenever $\exists y \forall^{\mathrm{st}} x \mathrm{A}$ (or its dual) occurs with A internal, use (I), taking advantage of its simplification (1.4) whenever A filters in x.

One thing to remember in using this reduction algorithm is that the implication $\exists^{\mathrm{st}} x \mathrm{A}(x) \to \mathrm{B}$ is equivalent to $\forall^{\mathrm{st}} x[\mathrm{A}(x) \to \mathrm{B}]$, and dually $\forall^{\mathrm{st}} x \mathrm{A}(x) \to \mathrm{B}$ is equivalent to $\exists^{\mathrm{st}} x[\mathrm{A}(x) \to \mathrm{B}]$, provided in both cases that x does not occur free in B. (This has nothing to do with the superscript "st"; it is a general fact about quantifiers in the hypothesis of an implication. For example, let A(s) be "s is an odd perfect number" and let B be the Riemann hypothesis. Then $\exists s \mathrm{A}(s) \to \mathrm{B}$ and $\forall s[\mathrm{A}(s) \to \mathrm{B}]$ say the same thing.) If B begins with an external quantifier, it comes out of the implication unchanged, but we have our choice of which quantifier to take out first. The idea is to do this in such a way as to introduce as few functions as possible. If we have an equivalence, we must rewrite it as the conjunction of two implications and rename the bound variables, since they come out of the implications in different ways. For example, the formula $\forall^{\mathrm{st}} x \mathrm{A}(x) \leftrightarrow \forall^{\mathrm{st}} y \mathrm{B}(y)$ is equivalent to

$$[\forall^{\mathrm{st}} x \mathrm{A}(x) \to \forall^{\mathrm{st}} y \mathrm{B}(y)] \mathbin{\&} [\forall^{\mathrm{st}} u \mathrm{B}(u) \to \forall^{\mathrm{st}} v \mathrm{A}(v)],$$

which in turn is equivalent to

$$\exists^{\mathrm{st}} x \exists^{\mathrm{st}} u \forall^{\mathrm{st}} y \forall^{\mathrm{st}} v \big[[\mathrm{A}(x) \to \mathrm{B}(y)] \mathbin{\&} [\mathrm{B}(u) \to \mathrm{A}(v))\big]$$

and also to

$$\forall^{\mathrm{st}} y \forall^{\mathrm{st}} v \exists^{\mathrm{st}} x \exists^{\mathrm{st}} u \big[[\mathrm{A}(x) \to \mathrm{B}(y)] \mathbin{\&} [\mathrm{B}(u) \to \mathrm{A}(v))\big].$$

If all of this is preceded by $\forall t$, where t is a free variable in A or B, the second form is advantageous; if it is preceded by $\exists t$, one should use the first form.

Let us illustrate the reduction algorithm with the definitions by (S) in the previous section of "continuous at x", "continuous", and "uniformly continuous". The formulas in question are respectively

$$\begin{aligned} \forall y\big[y \simeq x \to f(y) \simeq f(x)\big], \\ \forall^{\mathrm{st}} x \forall y\big[y \simeq x \to f(y) \simeq f(x)\big], \\ \forall x \forall y\big[y \simeq x \to f(y) \simeq f(x)\big]. \end{aligned}$$

We must eliminate $\simeq$. With ε and δ ranging over $\mathbf{R}^+$, we obtain

$$\forall y\big[\forall^{\mathrm{st}}\delta[|y-x| \le \delta] \to \forall^{\mathrm{st}}\varepsilon[|f(y)-f(x)| \le \varepsilon]\big], \tag{1.11}$$

$$\forall^{\mathrm{st}} x \forall y\big[\forall^{\mathrm{st}}\delta[|y-x| \le \delta] \to \forall^{\mathrm{st}}\varepsilon[|f(y)-f(x)| \le \varepsilon]\big], \tag{1.12}$$

$$\forall x \forall y\big[\forall^{\mathrm{st}}\delta[|y-x| \le \delta] \to \forall^{\mathrm{st}}\varepsilon[|f(y)-f(x)| \le \varepsilon]\big]. \tag{1.13}$$

In all of these formulas, f is a standard parameter, and x is a standard parameter in (1.11). First bring $\forall^{\mathrm{st}}\varepsilon$ out; it goes all the way to the left.

Then bring $\forall^{\mathrm{st}}\delta$ out; it changes to $\exists^{\mathrm{st}}\delta$, and since the formula filters in δ, it goes to the left of $\forall y$, and also of $\forall x$ in (1.13). This yields

$$\begin{aligned} \forall^{\mathrm{st}}\varepsilon\exists^{\mathrm{st}}\delta\forall y\big[|y-x|\le\delta\to|f(y)-f(x)|\le\varepsilon\big],\\ \forall^{\mathrm{st}}\varepsilon\forall^{\mathrm{st}}x\exists^{\mathrm{st}}\delta\forall y\big[|y-x|\le\delta\to|f(y)-f(x)|\le\varepsilon\big],\\ \forall^{\mathrm{st}}\varepsilon\exists^{\mathrm{st}}\delta\forall x\forall y\big[|y-x|\le\delta\to|f(y)-f(x)|\le\varepsilon\big]. \end{aligned}$$

Now apply transfer; this simply removes the superscripts "st" and shows that our definitions are equivalent to the usual ones.

Now consider $\forall x\exists^{\mathrm{st}}x_0[x\simeq x_0]$; that is,

$$\forall x\exists^{\mathrm{st}}x_0\forall^{\mathrm{st}}\varepsilon\big[|x-x_0|\le\varepsilon\big]. \tag{1.14}$$

It is understood that x and x_0 range over the standard set E; this was our definition by (S) of E being compact. First we move $\forall^{\mathrm{st}}\varepsilon$ to the left, by $(\widetilde{\mathrm{S}})$, introducing the standard function $\tilde\varepsilon\colon E\to\mathbf{R}^+$. Thus (1.14) becomes

$$\forall^{\mathrm{st}}\tilde\varepsilon\forall x\exists^{\mathrm{st}}x_0\big[|x-x_0|\le\tilde\varepsilon(x_0)\big]. \tag{1.15}$$

Now use (I). There is no filtering in (1.15), so it becomes

$$\forall^{\mathrm{st}}\tilde\varepsilon\exists^{\mathrm{stfin}}x_0'\forall x\exists x_0{\in}x_0'\big[|x-x_0|\le\tilde\varepsilon(x_0)\big].$$

Now apply (T). We obtain

$$\forall\tilde\varepsilon\exists^{\mathrm{fin}}x_0'\forall x\exists x_0{\in}x_0'\big[|x-x_0|\le\tilde\varepsilon(x_0)\big],$$

where $\exists^{\mathrm{fin}}$ means "there exists a finite set such that" (and similarly $\forall^{\mathrm{fin}}$ means "for all finite sets"). But this is mathematically equivalent to the usual definition of a set being compact.

Definitions by (S) are an external way of characterizing internal notions, but they simultaneously suggest new external notions. These are often indicated by the prefix S-. Thus we say that f is S-*continuous at* x in case whenever $y\simeq x$ we have $f(y)\simeq f(x)$. If both f and x are standard, this is the same as saying that f is continuous at x. But let $t>0$ be infinitesimal, and let $f(x)=t/\pi(t^2+x^2)$. Then f is continuous at 0 (this internal property is true for any $t>0$), but it is not S-continuous at 0. Let $g(x)=t$ for $x\ne 0$ and $g(0)=0$. Then g is discontinuous at 0 but is S-continuous at 0. Let $h(x)=x^2$. Then h is continuous at $1/t$ but is not S-continuous at $1/t$. Similarly, we say that f is *S-uniformly continuous* in case whenever $y\simeq x$ we have $f(y)\simeq f(x)$.

I am not very fond of the S-notation. It seems to imply that to every internal notion there is a unique corresponding external S-notion, but this is not so. There may be distinct external notions that are equivalent for standard values of the parameters.

Now let us apply the reduction algorithm to some of our external theorems. Let us split Theorem 2 into three different statements.

(i) *If every element of a set is standard, then the set is finite.* That is,

$$\forall S\big[\forall y{\in}S\exists^{\mathrm{st}}x[y=x]\to S\text{ is finite}\big],$$
$$\forall S\exists y{\in}S\forall^{\mathrm{st}}x\big[y=x\to S\text{ is finite}\big],$$
$$\forall^{\mathrm{stfin}}x'\forall S\exists y{\in}S\forall x{\in}x'\big[y=x\to S\text{ is finite}\big], \tag{1.16}$$
$$\forall^{\mathrm{fin}}x'\forall S\exists y{\in}S\forall x{\in}x'\big[y=x\to S\text{ is finite}\big], \tag{1.17}$$
$$\forall^{\mathrm{fin}}x'\forall S\big[S\subseteq x'\to S\text{ is finite}\big].$$

In other words, (i) is equivalent to (i′): *every subset of a finite set is finite.* We used (I) for (1.16), (T) for (1.17), and then pushed quantifiers back inside the implication as far as possible to make the result more readable.

(ii) *If every element of a set is standard, then the set is standard.* That is,

$$\forall S\big[\forall y{\in}S\exists^{\mathrm{st}}x[y=x]\to\exists^{\mathrm{st}}T[S=T]\big]. \tag{1.18}$$

Apply (I) to the hypothesis:

$$\forall S\big[\exists^{\mathrm{stfin}}x'\forall y{\in}S\exists x{\in}x'[y=x]\to\exists^{\mathrm{st}}T[S=T]\big];$$

simplify:

$$\forall S\big[\exists^{\mathrm{stfin}}x'[S\subseteq x']\to\exists^{\mathrm{st}}T[S=T]\big];$$

pull out the quantifiers and apply (I):

$$\forall^{\mathrm{stfin}}x'\exists^{\mathrm{stfin}}T'\forall S\exists T{\in}T'[S\subseteq x'\to S=T];$$

simplify and apply (T):

$$\forall^{\mathrm{fin}}x'\exists^{\mathrm{fin}}T'\forall S[S\subseteq x'\to S\in T'].$$

This is true for $T'=\wp x'$, where $\wp X$ denotes the power set of X (the set of all subsets of X). In other words, (ii) is equivalent to (ii′): *the power set of a finite set is finite.* When reducing an external formula, it is a good idea to attack subformulas first. For example, had we pulled the quantifiers out of (1.18) directly, we would have obtained

$$\forall S \exists^{\mathrm{st}} T \exists y{\in}S \forall^{\mathrm{st}} x[y = x \to S = T],$$

which after reduction and simplification gives:

$$\forall \widetilde{x'} \exists^{\mathrm{fin}} T' \forall S \Big[S \subseteq \bigcap_{T \in T'} \widetilde{x'}(T) \to S \in T' \Big]. \tag{1.19}$$

Every function must have a domain, so the use of ($\widetilde{\mathrm{S}}$) to produce the finite-set valued function $\widetilde{x'}$ was illegitimate; to be honest, we should have introduced $\forall^{\mathrm{st}} V$ with all of the variables restricted to lie in V, so that after transfer the formula begins $\forall V$ with all of the variables restricted to lie in V. Consider this done. Here is an internal proof of (1.19). Choose any T_0 and let

$$T' = \wp \left(\widetilde{x'}(T_0) \right) \cup \{ T_0 \}.$$

If $S \subseteq \bigcap_{T \in T'} \widetilde{x'}(T)$, then in particular $S \subseteq \widetilde{x'}(T_0)$ and so $S \in T'$. Thus (1.19) is an obscure way of saying that the power set of a finite set is finite.

(iii) *Every element of a standard finite set is standard.* That is,

$$\forall^{\mathrm{st}} S \big[S \text{ is finite} \to \forall y{\in}S \exists^{\mathrm{st}} x[y = x] \big].$$

This reduces easily to

$$\forall S \exists^{\mathrm{fin}} x' [S \text{ is finite} \to S \subseteq x'].$$

In other words, (iii) is equivalent to (iii$'$): *every finite set is a subset of some finite set.*

With the variables ranging over $\mathbf{N}$, Theorem 3 is

$$\forall n \forall k \big[k \le n \to \exists^{\mathrm{st}} j [j = k] \big],$$

which reduces easily to

$$\forall n \exists^{\mathrm{fin}} j' \forall k [k \le n \to k \in j'].$$

Theorem 4, that there is a finite set containing every standard object, reduces immediately to the following triviality: for all finite sets x' there is a finite set F such that every element of x' is an element of F.

Theorem 4 is shocking, but the informal statement "there is a finite set containing any fixed element" appears quite reasonable. Similarly, with the variables ranging over $\mathbf{R}^+$, one is accustomed to saying "there is an x less than any fixed ε." Nonstandard analysis takes the further step of saying "call such an x an infinitesimal." I want to emphasize again that the predicate "standard" has no semantic content in IST; it is a kind of syntactical place-holder signifying that the object in question is to be held fixed. With many objects in play at once, some depending on others, the syntax of being held fixed becomes complicated, and the rules for handling the idea correctly are (I), (S), and (T). What Abraham Robinson invented is nothing less than a new logic. He was explicit about this in the last paragraph of his epoch-making book:[4]

> Returning now to the theory of this book, we observe that it is presented, naturally, within the framework of contemporary Mathematics, and thus appears to affirm the existence of all sorts of infinitary entities. However, from a formalist point of view we may look at our theory syntactically and may consider that what we have done is to introduce *new deductive procedures* rather than new mathematical entities.

One technical point is worth commenting on. It is essential to the success of the reduction algorithm that the idealization principle hold for internal formulas with free variables; this feature was not present in Robinson's notion of enlargement [Ro, §2.9].

Exercises for Section 1.6

1. Find a function f that is S-uniformly continuous without being uniformly continuous, and vice versa.

2. What is the reduction of the formula $x \simeq 0$?

3. What is the reduction of Theorem 5?

4. Show that the definitions by (S) of open, closed, and bounded in the previous section are equivalent to the usual ones.

5. Show that the definition by (S) of a complete metric space in the previous section is equivalent to the usual one.

6. Say that a sequence $a\colon \mathbf{N} \to \mathbf{R}$ is *S-Cauchy* in case for all unlimited n and m we have $a_n \simeq a_m$. Show that a standard sequence is S-Cauchy if and only if it is Cauchy. Say that a is *of limited fluctuation* in case for all standard $\varepsilon > 0$ and all k, if $n_1 < \cdots < n_k$ and $|a_{n_1} - a_{n_2}| \geq \varepsilon$, $\ldots$, $|a_{n_{k-1}} - a_{n_k}| \geq \varepsilon$, then k is limited. Show that a standard sequence is of limited fluctuation if and only if it is Cauchy. Can a sequence be S-Cauchy without being of limited fluctuation, or vice versa?

7. Is the unit ball of $\mathbf{R}^n$, where n is unlimited, compact?

[4][Ro] Abraham Robinson, "Non-Standard Analysis," Revised Edition, American Elsevier, New York, 1974.

8? Can the reduction of Theorem 6 be made intelligible? (The ? means that I don't know the answer.)

1.7 Answers to the exercises

Section 1.1

1. Yes and yes; this set can be formed for any natural number, and it is a finite set.
2. No.
3. Let x be standard and positive. Then $x \leq x$, so x is limited.
4. No.
5. At first sight this looks like illegal set formation, but let S be the set of all real numbers x such that $x^2 \leq 1$. Assuming that 1 is standard, we see that every x in S is limited, so S is the set in question.
6. Not yet; so far, nothing guarantees that if ε is standard then so is $\varepsilon/2$.

Section 1.2

1. Yes. Let x and y be infinitesimal and let $\varepsilon > 0$ be standard. By transfer, $\varepsilon/2$ is standard, so that $|x| \leq \varepsilon/2$ and $|y| \leq \varepsilon/2$. Then $|x+y| \leq \varepsilon$, and since ε was an arbitrary strictly positive standard number, $x+y$ is infinitesimal.
2. There are standard numbers R and S such that $|r| \leq R$ and $|s| \leq S$. By (T), $R+S$ and RS are standard.
3. Let $\varepsilon > 0$ be standard. There is a standard R, which we may take to be non-zero, such that $|r| \leq R$. By (T), ε/R is standard, so $|x| \leq \varepsilon/R$. Hence $|xr| \leq \varepsilon$.
4. For $\varepsilon > 0$ and $r = 1/\varepsilon$, ε is standard if and only if r is standard, by the transfer principle.
5. We cannot speak of "the infinitesimals" as an ideal, or of "the integral domain of limited real numbers", without illegal set formation. If we avoid these illegal set formations, the preceding exercises essentially give an affirmative answer to the first question. But to talk about "the quotient field" we really need sets. The discussion must be postponed to Chapter 4, where we shall have external sets at our disposal.
6. No. If the closed interval is $[0, b]$ where b is unlimited, the function has no standard bound. The transfer was illegal because the closed interval was a parameter.
7. We have $x \leq y + \alpha$ where α is infinitesimal and x and y are standard. Then $x \leq y + \varepsilon$ for all standard $\varepsilon > 0$, so by (T) we have $x \leq y + \varepsilon$ for all $\varepsilon > 0$. Hence $x \leq y$.

Section 1.3

1. This is just a restatement of Theorem 3.

2. Yes. Euclid showed that for any natural number n there is a prime greater than n.

3. Let $x > 0$ be infinitesimal. Then $0 < x < 1$, so x is of the form $x = \sum_{n=1}^{\infty} a_n 10^{-n}$ where each a_n is one of 0, ..., 9. Since 10^{-n} is standard if n is, by (T), each a_n with n standard is 0. But $x > 0$, so not all of its decimal digits are 0. As we already know, any $x > 0$ has a first non-zero decimal digit. What about the number whose decimal digits are 0 for all standard n and 7 for all nonstandard n? The question makes so sense. The decimal digits form a sequence; a sequence is a set; and I committed illegal set formation in posing the question.

4. Yes; by (I), there is a finite subset of H containing all of its standard points, so take its span. (This is by no means unique, and we cannot form the smallest such space without illegal set formation.) Any finite dimensional subspace of a Hilbert space is closed.

5. We have $\forall x \exists^{\mathrm{st}} n [x \in S_n]$, where x ranges over S. By (I), we have that $\exists^{\mathrm{stfin}} n' \forall x \exists n {\in} n' [x \in S_n]$, and since the sets are disjoint, all but a standard finite number of them are empty.

6. The set of all N such that $|a_n| \le 1/n$ for all $n \le N$ contains all standard N, so by overspill it contains some unlimited N.

7. Not necessarily; consider the identity function.

8. We can assume that X is non-empty, since otherwise both sides are vacuously true. The backward direction holds by (I). In the forward direction, by (I) we have some finite standard x' that works, and we can take it to be non-empty. But by Theorem 2, $x' \cap X$ is also a standard finite set, and it is an element of $\wp_{\mathrm{fin}} X$.

Section 1.4

1. Let $f: X \to Y$, and assume that X and Y are contained in some standard set. Let $g = {}^{\mathrm{S}}f$. For all z, if $z \in g$ then $z \in {}^{\mathrm{S}}X \times {}^{\mathrm{S}}Y$; for all x in ${}^{\mathrm{S}}X$ and y_1 and y_2 in ${}^{\mathrm{S}}Y$, if $\langle x, y_1 \rangle$ and $\langle x, y_2 \rangle$ are in g, then $y_1 = y_2$. These statements hold by definition for the standard elements, and since the sets in question are standard, they hold by (T) for all elements. Thus ${}^{\mathrm{S}}f$ is a function from a subset of ${}^{\mathrm{S}}X$ into ${}^{\mathrm{S}}Y$.

2. Replace Y by its characteristic function. Let X be a standard set, let z range over X, let y range over $\{0,1\}$, and let $\tilde{y}$ range over $\{0,1\}^X$. We clearly have $\forall^{\mathrm{st}} z \exists^{\mathrm{st}} y [y = 1 \leftrightarrow \mathrm{A}]$, so by $(\widetilde{\mathrm{S}})$ we have $\exists^{\mathrm{st}} \tilde{y} \forall^{\mathrm{st}} z [\tilde{y}(z) = 1 \leftrightarrow \mathrm{A}]$. Then we let $Y = \{ z \in X : \tilde{y}(z) = 1 \}$.

3. By elementary calculus, this integral is equal to 1 for any $t > 0$. There are no standard pairs $\langle x, f(x) \rangle$, so ${}^{\mathrm{S}}f$ is the empty set.

4. Let $S = {}^{\mathrm{S}}\{ n \in \mathbf{N} : \mathrm{A}(n) \}$. Then 0 is in S. By assumption, for all standard n, if n is in S then $n+1$ is in S, so by (T) this is true for all n. By induction, $S = \mathbf{N}$. In particular, every standard n is in S, so $\mathrm{A}(n)$ holds for every standard n.

5. Let
$$S = {}^{\mathrm{S}}\{\, \gamma \le \alpha : \mathrm{A}(\gamma) \,\}.$$
Then every standard element of S is an ordinal, so the same is true by (T) for all elements of S. Since S is non-empty (it contains α), it contains a least β, which is standard by (T). Then $\beta \le \alpha$. Consider a standard γ such that $\mathrm{A}(\gamma)$. If $\alpha < \gamma$, then certainly $\beta \le \gamma$; if $\gamma \le \alpha$ then $\gamma \in S$, so again $\beta \le \gamma$. This concludes the proof.

This is often used to prove $\forall^{\mathrm{st}}\alpha \mathrm{B}(\alpha)$. Argue indirectly. If not, there is a least standard α such that $\mathrm{A}(\alpha)$ where A is $\neg$B. If we can show that α cannot be 0, a successor, or a limit ordinal, then the proof will be complete.

6. To establish (I), we need only show (without using the backward direction of the idealization principle) that every element of a standard finite set x' is standard. We do this by external induction on the cardinality n of x'. By transfer, n is standard. The statement is vacuously true for $n = 0$. For $n > 0$, the set x' contains an element, so by (T) it contains a standard element. Delete this element; what remains is a set of cardinality $n - 1$ and by (T) it is a standard set. By the external induction hypothesis, all of the remaining elements are also standard.

Section 1.5

1. An element η of Ω is a function from T to $\bigcup_{t\in T} X_t$ such that $\eta(t) \in X_t$ for all t in T. So if we know that for all standard t there is a y in X_t with $y \simeq \omega(t)$, then $(\widetilde{\mathrm{S}})$ tells us that there is a standard η (a $\tilde{y}$) in Ω such that $\eta(t) \simeq \omega(t)$ for all standard t. By the definition of the product topology and (T), a standard basic neighborhood of η is given by a standard finite set of t_i in T and a standard finite set of neighborhoods U_i of the $\eta(t_i)$, so $\eta \simeq \omega$.

2. By (T), assume that f, X, and Y are standard. Let y be in Y. Then there is an x in X with $f(x) = y$ and a standard x_0 in X with $x_0 \simeq x$. By (T), $f(x_0)$ is standard, and by continuity, $f(x_0) \simeq f(x)$. Hence Y is compact.

3. Not necessarily. In the plane, consider the x-axis and the parallel line one unit above it together with the vertical interval of length 1 with x unlimited joining them.

4. If x is limited, then ${}^\circ\{x\}$ is $\{\mathrm{st}\, x\}$; otherwise it is the empty set.

5. If we graph f, what we see is the union of the x-axis and the positive half of the y-axis, and this is the shadow of the graph of f. Let $x \ne 0$ be standard; then $\langle x, f(x)\rangle \simeq \langle x, 0\rangle$. Let $y > 0$ be standard; then the positive solution x of $f(x) = y$ is infinitesimal, so $\langle x, f(x)\rangle \simeq \langle 0, y\rangle$. Theorem 6 tells us that the origin too must be in the shadow. To see this directly, choose $x = t^{1/4}$, for example; then $\langle x, f(x)\rangle \simeq \langle 0, 0\rangle$.

6. Suppose that n is standard. If one, and hence all, of the vertices is standard, then E is standard and since it is closed ${}^\circ E = E$; otherwise ${}^\circ E$ is obtained by an infinitesimal rotation of E. If n is nonstandard, ${}^\circ E$ is the unit circle.

Section 1.6

1. Let $t > 0$ be infinitesimal, and let $f(x) = t/\pi(t^2 + x^2)$. Then f is uniformly continuous but not S-uniformly continuous. For an example in the other direction, let $g(x) = t$ for $x \neq 0$ and $g(0) = 0$.

2. The formula $x \simeq 0$, i.e., $\forall^{\mathrm{st}}\varepsilon[|x| \leq \varepsilon]$ where ε ranges over $\mathbf{R}^+$, is weakly equivalent to the internal formula $\forall\varepsilon[|x| \leq \varepsilon]$, which is equivalent to $x = 0$. The only standard infinitesimal is 0.

3. We have

$$\forall x[\exists^{\mathrm{st}}r[|x| \leq r] \to \exists^{\mathrm{st}}x_0\forall^{\mathrm{st}}\varepsilon[|x - x_0| \leq \varepsilon]],$$
$$\forall^{\mathrm{st}}r\forall x\exists^{\mathrm{st}}x_0\forall^{\mathrm{st}}\varepsilon[|x| \leq r \to |x - x_0| \leq \varepsilon],$$
$$\forall^{\mathrm{st}}r\forall^{\mathrm{st}}\tilde{\varepsilon}\exists^{\mathrm{stfin}}x_0'\forall x\exists x{\in}x_0'[|x| \leq r \to |x - x_0| \leq \tilde{\varepsilon}(x_0)];$$

then remove the superscripts "st". This says that every interval $[-r, r]$ is compact.

4. For "open" we have

$$\forall^{\mathrm{st}}x_0{\in}E\,\forall z{\in}\mathbf{R}[\forall^{\mathrm{st}}\varepsilon[|z - x_0| \leq \varepsilon] \to z \in E].$$

Bring $\forall^{\mathrm{st}}\varepsilon$ out as $\exists^{\mathrm{st}}\varepsilon$, which filters past $\forall z{\in}\mathbf{R}$. Then use (T) to obtain

$$\forall x_0\exists\varepsilon\forall z{\in}\mathbf{R}[|z - x_0| \leq \varepsilon \to z \in E].$$

Similarly, the formula for "closed" reduces to

$$\forall x_0{\in}\mathbf{R}\exists\varepsilon\forall x{\in}E[|x - x_0| \leq \varepsilon \to x_0 \in E],$$

which is perhaps more readable if we push $\forall x{\in}E$ inside the parentheses as $\exists x{\in}E$. For "bounded", use the filtering form of (I) and (T) to obtain the usual formulation $\exists r\forall x{\in}E[|x| \leq r]$.

5. Let $\langle X, d\rangle$ be a standard metric space. To avoid confusion, let us temporarily call it *nice* in case for all x, if for all standard $\varepsilon > 0$ there is a standard y with $d(x, y) \leq \varepsilon$, then there is a standard x_0 with $x \simeq x_0$. Suppose that X is incomplete. By (T), there is a standard Cauchy sequence x_n with no limit. Let $\nu \simeq \infty$ and let $\varepsilon > 0$ be standard. Consider the set S of all n such that $d(x_n, x_\nu) \leq \varepsilon$. This set contains all unlimited n, so by overspill it contains some limited n, for which x_n is standard. But there is no standard x_0 with $x_\nu \simeq x_0$, for then it would be the limit of the x_n. Thus an incomplete standard metric space is not nice. Conversely, let X be complete and let x be a point in X such that for all standard $\varepsilon > 0$ there is a standard y with $d(x, y) \leq \varepsilon$. Then for all standard n there is a standard y with $d(x, y) \leq 1/n$, so by $(\widetilde{\mathrm{S}})$ there is a standard sequence y_n such that $d(x, y_n) \leq 1/n$ for all standard n. By the triangle inequality and (T), y_n is a Cauchy sequence and so has a limit x_0, which is standard by (T), and $x \simeq x_0$. Thus a complete

standard metric space is nice, and our definition by (S) of a metric space being complete is equivalent to the usual one.

It is also interesting to approach this problem via the reduction algorithm. We have

$$\forall x[\forall^{\mathrm{st}}\varepsilon\exists^{\mathrm{st}}y[d(x,y)\le\varepsilon]\rightarrow\exists^{\mathrm{st}}x_0\forall^{\mathrm{st}}\delta[d(x,x_0)\le\delta]].$$

Use $(\widetilde{\mathrm{S}})$ before pulling anything out, to avoid spurious arguments of the functions, and then pull out the resulting functions. We obtain

$$\forall^{\mathrm{st}}\tilde{\delta}\forall^{\mathrm{st}}\tilde{y}\forall x\exists^{\mathrm{st}}x_0\exists^{\mathrm{st}}\varepsilon\big[d(x,\tilde{y}(\varepsilon))\le\varepsilon\rightarrow d(x,x_0)\le\tilde{\delta}(x_0)\big],$$

which by (I) and (T) is equivalent to

$$\forall\tilde{\delta}\forall\tilde{y}\exists^{\mathrm{fin}}\varepsilon'\exists^{\mathrm{fin}}x_0'\forall x\exists x_0{\in}x_0'\big[\forall\varepsilon{\in}\varepsilon'\big(d(x,\tilde{y}(\varepsilon))\le\varepsilon\big)\rightarrow d(x,x_0)\le\tilde{\delta}(x_0)\big],$$

where $\exists\varepsilon{\in}\varepsilon'$ was pushed back inside the implication. Now we have the internal problem of seeing that this is equivalent to completeness of the metric space. Only the Cauchy $\tilde{y}$ are relevant (to obtain a more customary notation, we could let $y_n = \tilde{y}(1/n)$), for otherwise we can find a two-point set ε' that violates the hypothesis of the implication, by the triangle inequality. If the space X is complete, just let x_0' be the singleton consisting of the limit as $\varepsilon\to 0$ of $\tilde{y}(\varepsilon)$. To construct a counterexample when X is not complete, let $\tilde{y}$ be Cauchy without a limit, and let $\tilde{\delta}(x_0)$ be the distance from x_0 to the limit (in the completion) of $\tilde{y}(\varepsilon)$.

6. The S-Cauchy condition is

$$\forall n\forall m[\forall^{\mathrm{st}}r[n\ge r\ \&\ m\ge r]\rightarrow\forall^{\mathrm{st}}\varepsilon[|a_n-a_m|\le\varepsilon]].$$

The r filters out to give the usual definition of a Cauchy sequence. The limited fluctuation condition is

$$\forall^{\mathrm{st}}\varepsilon\forall k[\mathrm{A}(k,\varepsilon,a)\rightarrow\exists^{\mathrm{st}}r[k\le r]],$$

where $\mathrm{A}(k,\varepsilon,a)$ is an abbreviation for the assertion that the sequence a contains k ε-fluctuations. Again the r filters out, and a standard sequence is of limited fluctuation if and only if for all $\varepsilon>0$ there is a bound r on the number of ε-fluctuations, which is the same as being Cauchy. Now let a be an S-Cauchy sequence, not necessarily standard, and let $\varepsilon>0$ be standard. Consider the set of all n such that for all $m>n$ we have $|a_n-a_m|\le\varepsilon$; this set contains all unlimited n, so by overspill it contains some limited n. Consequently, an S-Cauchy sequence is of limited fluctuation. But let $\nu\simeq\infty$ and let $a_n=0$ for $n<\nu$ and $a_n=1$ for $n\ge\nu$. This sequence is of limited fluctuation but is not S-Cauchy. Thus we have two distinct external concepts that agree on standard objects.

7. The unit ball of any Euclidean space is compact.