INTERNAL SET THEORY: A NEW APPROACH TO NONSTANDARD ANALYSIS

EDWARD NELSON[1]

1. Internal set theory. We present here a new approach to Abraham Robinson's nonstandard analysis [10] with the aim of making these powerful methods readily available to the working mathematician. This approach to nonstandard analysis is based on a theory which we call *internal set theory* (IST). We start with axiomatic set theory, say ZFC (Zermelo-Fraenkel set theory with the axiom of choice [1]). In addition to the usual undefined binary predicate $\in$ of set theory we adjoin a new undefined unary predicate *standard*. The axioms of IST are the usual axioms of ZFC plus three others, which we will state below.

All theorems of conventional mathematics remain valid. No change in terminology is required. What is new in internal set theory is only an addition, not a change. We choose to call certain sets standard (and we recall that in ZFC every mathematical object–a real number, a function, etc.–is a set), but the theorems of conventional mathematics apply to all sets, nonstandard as well as standard.

In writing formulas we use $\wedge$ for and, $\vee$ for or, $\sim$ for not, $\Rightarrow$ for implies, and $\Leftrightarrow$ for is equivalent to. We call a formula of IST *internal* in case it does not involve the new predicate "standard" (that is, in case it is a formula of ZFC); otherwise we call it *external*. Thus "x standard" is the simplest example of an external formula. To assert that x is a standard set has no meaning within conventional mathematics–it is a new undefined notion.

The fact that we have adjoined "standard" as an undefined predicate (rather than defining it in terms of $\in$ as is the case with all of the predicates of conventional mathematics) requires a readjustment of an engrained habit. We are used to defining subsets by means of predicates. In fact, it follows from the axioms of ZFC that if $A(z)$ is an internal formula then for all sets x there is a set $y = \{z \in x: A(z)\}$ such that for all sets z we have $z \in y \Leftrightarrow z \in x \wedge A(z)$. However, the axioms of ZFC say nothing about external predicates. For example, no axioms allow us to assert that there is a subset S of the set N of all natural numbers such that for all n we have $n \in S \Longleftrightarrow n \in \mathbf{N} \wedge n$ standard. *We may not use external predicates to define subsets.* We call the violation of this rule *illegal set formation.*

We adopt the following abbreviations:

$\forall^{st} x$ for $\forall x(x \text{ standard}) \Rightarrow,$ $\exists^{st} x$ for $\exists x(x \text{ standard}) \wedge$

$\forall^{fin} x$ for $\forall x(x \text{ finite}) \Rightarrow,$ $\exists^{fin} x$ for $\exists x(x \text{ finite}) \wedge$

$\forall^{\text{st fin}} x$ for $\forall^{\text{st}} x(x \text{ finite}) \Rightarrow,$ $\exists^{\text{st fin}} x$ for $\exists^{\text{st}} x(x \text{ finite}) \wedge$

$\forall x \in y$ for $\forall x(x \in y) \Rightarrow,$ $\exists x \in y$ for $\exists x(x \in y) \wedge$.

Here "x finite" has its usual meaning: it is an abbreviation for the internal formula which asserts that there is no bijection of x with a proper subset of itself (or equivalently that there is a bijection of x with $\{m \in \mathbf{N}: m < n\}$ for some natural number n).

The axioms of IST are the axioms of ZFC together with three additional axiom schemes which we call the *transfer principle* (T), the *principle of idealization* (I), and the *principle of standardization* (S). They are as follows.

Let $A(x, t_1, \ldots, t_k)$ be an *internal* formula with free variables $x, t_1, \ldots, t_k$ and *no other free variables*. Then

(T) $\forall^{\text{st}} t_1 \cdots \forall^{\text{st}} t_k (\forall^{\text{st}} x \, A(x, t_1, \ldots, t_k) \Rightarrow \forall x \, A(x, t_1, \ldots, t_k)).$

Let $B(x, y)$ be an *internal* formula with free variables x, y and possibly other free variables. Then

(I) $\forall^{\text{st fin}} z \exists x \forall y \in z \, B(x, y) \Leftrightarrow \exists x \forall^{\text{st}} y \, B(x, y).$

Finally, let $C(z)$ be a formula, internal or external, with free variable z and possibly other free variables. Then

(S) $\forall^{\text{st}} x \exists^{\text{st}} y \forall^{\text{st}} z(z \in y \Leftrightarrow z \in x \wedge C(z)).$

This completes the description of internal set theory.

A *statement* is a formula with no free variables. Let A be an internal statement and let A^{st} be the statement obtained by replacing each occurrence of $\exists x$ by $\exists^{\text{st}} x$ and each occurrence of $\forall x$ by $\forall^{\text{st}} x$, for all variables x occurring in A. We call A^{st} the *relativization of A to the standard sets*. By successive applications of (T) (working from outside in) we see that $A \Leftrightarrow A^{\text{st}}$. Thus all theorems of conventional mathematics also hold when relativized to the standard sets. Conversely, to prove an internal theorem it suffices to prove its relativization to the standard sets.

So far we have made no mention of constants (such as $\varnothing$ for the empty set and $\mathbf{R}$ for the set of all real numbers). Constants are a matter of convenience. If we introduce them into the theory then all of our axioms remain valid for formulas containing constants, except that *the transfer principle is only valid if all of the constants occurring in the formula are standard.*

Before applying transfer to an assertion, we must verify two things: that the assertion is internal and that all parameters in it have standard values. We call the violation of this rule illegal transfer. It is the most insidious pitfall awaiting the mathematician who wants to use nonstandard analysis.

Suppose that there exists a unique x such that $A(x)$, where $A(x)$ is an internal formula whose only free variable is x. Then that x must be standard, since by transfer $\exists x \, A(x) \Rightarrow \exists^{\text{st}} x \, A(x)$. For example, the set $\mathbf{N}$ of all natural numbers, the set $\mathbf{R}$ of all real numbers, the real number π, and the Hilbert space $L^2(\mathbf{R})$ are all standard sets, since they may be uniquely described in conventional mathematical terms. *Every specific object of conventional mathematics is a standard set.* It remains unchanged in the new theory. For example, in internal

set theory there is only one real number system, the system **R** with which we are already familiar.

Let $B(x,y)$ be an internal formula with free variables x and y and possibly others. Then (I) asserts that the relation is simultaneously satisfiable for all standard y if and only if it is simultaneously satisfiable on every standard finite set. Whether or not the latter is true may depend on the other free variables, if any.

THEOREM 1.1. *Let X be a set. Then every element of X is standard if and only if X is a standard finite set.*

PROOF. Let $B(x,y)$ be $x \in X \wedge x \neq y$. Then the right-hand side of (I) is equivalent to $\exists x \in X \sim (x \text{ standard})$. Taking negations, we have

$$\forall x \in X(x \text{ standard}) \Leftrightarrow \exists^{\text{st fin}} z \forall x \exists y \in z(\sim (x \in X) \vee x = y)$$

$$\Leftrightarrow \exists^{\text{st fin}} z(X \subset z).$$

If X is a standard finite set then every element of it is standard, because we may take $z = X$. Conversely, if every element of X is standard then $X \subset z$, where z is a standard finite set. Hence $X \in P(z)$, where $P(z)$, the power set of z, is also a finite set. By what we have already proved, this means that X is standard, and it is finite since it is a subset of a finite set.   Q.E.D.

In particular, every infinite set has a nonstandard element. Thus there exists a nonstandard natural number. On the other hand, by transfer, 0 is a standard natural number, and for all natural numbers n, if n is standard then $n + 1$ is standard. This does not contradict the induction theorem (which says that if S is a subset of **N**, such that $0 \in S$ and such that for all n we have $n \in S \Rightarrow (n + 1) \in S$, then $S = \mathbf{N}$)–it merely shows that there does not exist a subset S of **N** such that a natural number is in S if and only if it is standard.

THEOREM 1.2. *There is a finite set F such that for all standard x we have $x \in F$.*

PROOF. Apply (I) to $B(F,X)$ given by $(x \in F \wedge F \text{ finite})$.   Q.E.D.

Such a set F cannot be standard, for if it were then by transfer it would contain all sets x. Also, there is no smallest such F. If we attempt to define the intersection of all such F we are engaging in illegal set formation, because we may only define intersections of *sets* of sets and we cannot use an external predicate to define the set of sets to be intersected.

Although we cannot use external predicates to define subsets, the principle of standardization provides a substitute. Two sets are equal if they have the same elements. By transfer, two standard sets are equal if they have the same standard elements. Thus the set y given by (S) is unique. We denote it by $^S\{z \in x: C(z)\}$. This may be read as "the standard subset of x whose standard elements are those which satisfy C". Do not read it as "the set of all standard elements in x which satisfy C" because this is illegal set formation. *When a standard set is defined by the standardization principle, the criterion for set membership applies only to standard elements.* The standardization principle does not give a direct criterion for deciding whether a nonstandard element z of x is in $y = {}^S\{z \in x: C(z)\}$ or not. It may happen that $z \in y$ but $\sim C(z)$,

and it may happen that $z \notin y$ but $C(z)$. For example, let n be a nonstandard natural number. We claim that ${}^S\{z \in \mathbf{N}: z < n\} = \mathbf{N}$ (although there are z in $\mathbf{N}$ which do not satisfy $z < n$). To see this, it is enough, by transfer, to show that the two sets have the same standard elements, since both sets are standard. That is, we must show that if z is a standard natural number then $z < n$. But $\{w \in \mathbf{N}: w \leqslant z\}$ is a standard finite set. By Theorem 1.1 every element of it is standard, so that it does not contain n. In the same way we see that ${}^S\{z \in \mathbf{N}: z \geqslant n\} = \varnothing$ (although there are natural numbers z with $z \geqslant n$).

In the above example the predicate $C(z)$ given by $z < n$ is internal, so that we may also form the set $\{z \in \mathbf{N}: z < n\}$. This is a nonstandard set which is a proper subset of $\mathbf{N}$. Someone might object: "How can we form this set if n is nonstandard, since nonstandard is an external notion?" The objection has no merit. The formula $z < n$ is internal. For every natural number n we can form $\{z \in \mathbf{N}: z < n\}$.

The principle of standardization may be used to show the existence of standard functions. In the following theorem $A(x,y)$ is a formula, internal or external, with free variables x and y and possibly others.

THEOREM 1.3. *Let X and Y be standard sets and suppose that for all standard x in X there is a standard y in Y such that $A(x,y)$. Then there is a standard function $\tilde{y}: X \to Y$ such that for all standard x in X we have $A(x,\tilde{y}(x))$.*

PROOF. If for all standard x in X there is a *unique* standard y in Y such that $A(x,y)$, then this is immediate: the standard set ${}^S\{\langle x,y \rangle \in X \times Y: A(x,y)\}$ is, by (T), a function $\tilde{y}: X \to Y$ and, by definition, $A(x,\tilde{y}(x))$ for all standard x in X.

Now consider the general case. Let $P(Y)$ be the power set of Y (the set of all subset of Y). This is a standard set since Y is a standard set. For all x in X there is a unique standard set T in $P(Y)$ such that $T = {}^S\{y \in Y: A(x,y)\}$. By what we have just shown in the previous paragraph, there is a standard function $\tilde{T}: X \to P(Y)$ such that for all standard x in X we have $\tilde{T}(x) = {}^S\{y \in Y: A(x,y)\}$. By hypothesis $\tilde{T}(x) \neq \varnothing$ for all standard x in X. By the axiom of choice relativized to the standard sets, there is a standard function $\tilde{y}: X \to Y$ such that for all standard x in X we have $\tilde{y}(x) \in \tilde{T}(x)$ and, consequently, $A(x,\tilde{y}(x))$.   Q.E.D.

A real number x is called *infinitesimal* in case $|x| \leqslant \varepsilon$ for all standard $\varepsilon > 0$, *limited* in case $|x| \leqslant r$ for some standard r, and *unlimited* in case it is not limited. Notice that 0 is infinitesimal. By (T) it is the only standard infinitesimal, but by (I) there exist nonzero infinitesimals and there exist unlimited real numbers. We emphasize again that we are talking about the ordinary real number system $\mathbf{R}$ with which we are familiar and that everything we know about $\mathbf{R}$ remains valid. For example, if $x \neq 0$ then there is a integer n such that $nx \geqslant 1$. The integer n will be unlimited if x is infinitesimal, but this is an additional piece of knowledge which does not change anything we already know.

Two real numbers x and y are called *infinitely close*, denoted by $x \simeq y$, in case $x - y$ is infinitesimal. (Some people say "infinitesimally close", but they are not saying what they mean. Can you imagine gazing into the eyes of someone you love and saying, "I feel infinitesimally close to you"?)

THEOREM 1.4. *Every limited real number is infinitely close to a unique standard real number.*

PROOF. Let x be a limited real number, so that there is a standard real number r with $|x| \leqslant r$. Let $E = {}^S\{t \in \mathbf{R}: t \leqslant x\}$. Then E is by definition a standard set, and for all standard t in E we have $t \leqslant r$. By transfer, for all t in E we have $t \leqslant r$. Therefore E is bounded above. The set E is nonempty since $-r \in E$. Therefore E has a least upper bound a. Since E is standard, a is standard, by transfer. Suppose that $x - a > \varepsilon$ for some standard $\varepsilon > 0$. Then $a + \varepsilon \leqslant x$, and since $a + \varepsilon$ is standard, $a + \varepsilon$ is in E, which contradicts the fact that a is an upper bound for E. Suppose that $a - x > \varepsilon$ for some standard $\varepsilon > 0$. Then $x \leqslant a - \varepsilon$, so that for any standard t in E we have $t \leqslant a - \varepsilon$ and by transfer (since E and $a - \varepsilon$ are standard) for all t in E we have $t \leqslant a - \varepsilon$, which contradicts the fact that a is the least upper bound of E. Consequently $x \simeq a$. The uniqueness is clear from the fact that 0 is the only standard infinitesimal.   Q.E.D.

Let us point out several places where we might have gone wrong in the argument. Suppose we had said that $E = {}^S\{t \in \mathbf{R}: t \leqslant x\}$ is bounded above by x or that E is nonempty since $x \in E$. There is no justification for either of these assertions; they involve misapplication of (S). For example, let x be a nonzero infinitesimal and let $E = {}^S\{t \in \mathbf{R}: t \leqslant x\}$. Then $E = \{t \in \mathbf{R}: t \leqslant 0\}$ if $x > 0$ and $E = \{t \in \mathbf{R}: t < 0\}$ if $x < 0$. In the former case x is not in E and in the latter case x is not an upper bound for E. Suppose we had said that for all standard t in E we have $t \leqslant x$ and therefore for all t in E we have $t \leqslant x$. This is illegal transfer since x is not necessarily standard. *Part of the art of reasoning in nonstandard analysis consists of weakening assertions to a point at which transfer becomes applicable.*

If x is a limited real number, the standard real number which is infinitely close to it is called its *standard part* and is denoted by st x. Notice that if a and b are standard real numbers and x is in the closed interval $[a, b]$, then x is a limited real number and st $x \in [a, b]$. More generally, if E is a standard closed and bounded subset of $\mathbf{R}$ and $x \in E$, then x is limited and st $x \in E$.

One of the chief uses of the principle of standardization is in making definitions. For example, let $f: \mathbf{R} \to \mathbf{R}$ and $x \in \mathbf{R}$. We say that f is *continuous* at x in case (for f and x standard) for all y if $y \simeq x$ then $f(y) \simeq f(x)$. It is understood when we use locutions such as "in case (for f and x standard)" that we are defining a standard relation by means of (S). Let $\mathbf{R}^{\mathbf{R}}$ denote, as usual, the set of all functions from $\mathbf{R}$ to $\mathbf{R}$. This is a standard set. The above definition is the same as saying that $\langle f, x \rangle$ is an element of

$$ {}^S\{\langle f, x \rangle \in \mathbf{R}^{\mathbf{R}} \times \mathbf{R}: \forall y (y \simeq x \Rightarrow f(y) \simeq f(x))\}. $$

This defines, somewhat implicitly, what it means for an arbitrary (not necesarily standard) f to be continuous at an arbitrary x. Similarly, we define f to be *uniformly continuous* on the set E in case (for f and E standard) for all x and y in E, if $y \simeq x$ then $f(y) \simeq f(x)$. We will show in the next section that these definitions are equivalent to the usual definitions. The point is that these external criteria of continuity and uniform continuity are considerably easier to work with than the familiar internal $\varepsilon - \delta$ criteria. *Part of the power of nonstandard analysis is due to the fact that a complicated internal notion is frequently equivalent, on the standard sets, to a simple external notion.*

THEOREM 1.5. *If f is continuous at each point of a closed and bounded subset E of* **R** *then f is uniformly continuous on E.*

PROOF. By transfer, we may assume that f and E are standard. Let x and y be in E with $x \simeq y$. Then st $x \in E$, and $x \simeq$ st x, $y \simeq$ st x so that $f(x) \simeq f(\text{st } x)$, $f(y) \simeq f(\text{st } x)$ and thus $f(x) \simeq f(y)$.   Q.E.D.

Someone might reasonably ask, "How is the transfer principle applicable, since the given definitions of continuity of f at x and uniform continuity of f on E involve external formulas?" The answer is as follows: The transfer principle tells us that

$$\forall^{st} C \forall^{st} U [\forall^{st} f \forall^{st} E (\forall x (x \in E \Rightarrow \langle f, x \rangle \in C) \Rightarrow \langle f, E \rangle \in U)$$

$$\Rightarrow \forall f \forall E (\forall x (x \in E \Rightarrow \langle f, x \rangle \in C) \Rightarrow \langle f, E \rangle \in U)].$$

Now

$$C = {}^{S}\{\langle f, x \rangle \in \mathbf{R}^{\mathbf{R}} \times \mathbf{R} : \forall y (y \simeq x \Rightarrow f(y) \simeq f(x))\},$$

$$U = {}^{S}\{\langle f, E \rangle \in \mathbf{R}^{\mathbf{R}} \times P(\mathbf{R}) : \forall x \in E \forall y \in E (y \simeq x \Rightarrow f(y) \simeq f(x))\}$$

are, by definition, standard sets, so the result for standard f and E implies the result for general f and E. We will not spell out such an argument ever again.

Anticipating the result of the next section that our definitions of continuity and uniform continuity are equivalent to the usual ones, we have in Theorem 1.5 our first example of a proof of an internal theorem by means of IST. The question arises as to whether proofs by means of internal set theory are legitimate. In the Appendix we present the result, due to William C. Powell, that every internal theorem of IST is a theorem of ZFC. *Internal set theory may be used freely in proving conventional theorems.* This result also shows that it is always possible to avoid such methods. One may use them or not as one chooses.

2. A lexicon of nonstandard analysis. Nonstandard analysis involves an interplay of internal and external notions. Some of the theorems which we prove are external. Can we reformulate them so that they become internal? Definitions of standard objects made by means of (S) may involve external notions. Can we find equivalent internal formulations of such definitions? We will show that the answer to both questions is *yes* by exhibiting an algorithm which reduces any external formula of IST to an internal formula, with the same free variables, which is equivalent to it for all standard values of the free variables (subject to a technical qualification mentioned below).

Our point of view is a syntactical one. In IST we have two new quantifiers, $\forall^{st}$ and $\exists^{st}$, which we call *external* quantifiers as contrasted with the *internal* quantifiers $\forall$ and $\exists$, and (I), (S), (T) are essentially rules for handling these quantifiers.

Judging from the author's experience, it will be well to review informally the rules for handling internal quantifiers. We assume that displayed variables do not occur except where displayed (e.g. in (2.3) below, x does not occur in B). The basic rules are:

(2.1) $\sim \forall x\, A(x) \Leftrightarrow \exists x \sim A(x),$

(2.2) $\forall x \forall y\, A(x,y) \Leftrightarrow \forall y \forall x\, A(x,y),$

(2.3) $\forall x\, A(x) \wedge B \Rightarrow \forall x(A(x) \wedge B),$

(2.4) $\exists x\, A(x) \wedge B \Leftrightarrow \exists x(A(x) \wedge B).$

The rules for forming negations induce a duality; for example, corresponding to (2.2) we have $\exists x \exists y\, B(x,y) \Leftrightarrow \exists y \exists x\, B(x,y)$. These rules imply some others, for example:

(2.5) $[\forall x\, A(x) \Rightarrow B] \Leftrightarrow [\exists x(A(x) \Rightarrow B)].$

("All men are dishonest implies Diogenes was right" is equivalent to "There is a man such that if *he* is dishonest then Diogenes was right".)

(2.6) $[A \Rightarrow \forall x\, B(x)] \Leftrightarrow [\forall x(A \Rightarrow B(x))]$

and   similarly $[\exists x\, A(x) \Rightarrow B] \Leftrightarrow [\forall x(A(x) \Rightarrow B)]$ and $[A \Rightarrow \exists x\, B(x)]$ $\Leftrightarrow [\exists x(A \Rightarrow B(x))]$. Also

(2.7)
$$[\forall x\, A(x) \Rightarrow \forall y\, B(y)] \Leftrightarrow [\exists x \forall y(A(x) \Rightarrow B(y))]$$
$$\Leftrightarrow [\forall y \exists x(A(x) \Rightarrow B(y))],$$

(2.8)
$$[\forall x\, A(x) \Leftrightarrow \forall y\, B(y)] \Leftrightarrow [(\forall x\, A(x) \Rightarrow \forall y\, B(y))$$
$$\wedge\ (\forall w\, B(w) \Rightarrow \forall z\, A(z))]$$
$$\Leftrightarrow \exists x \exists w \forall y \forall z[(A(x) \Rightarrow B(y)) \wedge (B(w) \Rightarrow A(z))]$$
$$\Leftrightarrow \forall y \forall z \exists x \exists w[(A(x) \Rightarrow B(y)) \wedge (B(w) \Rightarrow A(z))].$$

These rules may be used to rewrite any formula as an equivalent formula of the form $Q_1 x_1 \cdots Q_n x_n A$, where each Q_i is $\forall$ or $\exists$ and A is a formula without quantifiers.

Rules (2.1)–(2.8) obviously apply to the external quantifiers $\forall^{st}$ and $\exists^{st}$ as well. By Theorem 1.3 we have

(S′) $\forall^{st} x \exists^{st} y\, A(x,y) \Leftrightarrow \exists^{st} \tilde{y} \forall^{st} x\, A(x, \tilde{y}(x)),$

where $A(x,y)$ may be internal or external and may have other free variables and, by duality,

$$\exists^{st} x \forall^{st} y\, A(x,y) \Leftrightarrow \forall^{st} \tilde{y} \exists^{st} x\, A(x, \tilde{y}(x)),$$

with the tacit understanding that x and y range over a standard set V. (This is the technical qualification mentioned before. In practice it is not restrictive because in concrete mathematics we are usually talking about something. To avoid further discussion we make the blanket assumption that whenever (S′) is used the variables in question range over a fixed standard set V.) We remark that

(2.9) $\forall x \forall^{st} y\, A(x,y) \Leftrightarrow \forall^{st} y \forall x\, A(x,y)$

and, by duality, $\exists x \exists^{st} y\, A(x,y) \Leftrightarrow \exists^{st} y \exists x\, A(x,y)$.

We introduce the notation $A \equiv B$ to mean that for all standard values of the free variables in A and B we have $A \Leftrightarrow B$. With this notation we may rewrite (T) as

$$(\text{T}') \qquad\qquad \forall^{st} x\, A(x) \equiv \forall x\, A(x), \qquad A \text{ internal,}$$

and, by duality, $\exists^{st} x\, A(x) \equiv \exists x\, A(x)$, again provided that A is internal. In these formulas, $A(x)$ may have other free variables, but equivalence is only asserted when they have standard values.

Now we can describe the reduction algorithm.

Step 1. Replace all external defined predicates (such as infinitesimal or $\simeq$) by their definitions, until the only remaining external predicate is "standard".

Step 2. If necessary, rewrite the formula so that "standard" appears only in the external quantifiers, replacing (x standard) by $\exists^{st} y\, (y = x)$.

Step 3. Using rules (2.1)–(2.8), rewrite the formula in the form $Q_1 x_1 \cdots Q_n x_n\, A(x_1,\ldots,x_n)$ where $A(x_1,\ldots,x_n)$ is internal and each Q_i is $\forall$, $\exists$, $\forall^{st}$, or $\exists^{st}$. We say that the formula is of *rank j* in case there are j internal quantifiers followed (on the right) by at least one external quantifier.

Step 4. If the rank of the formula is $j > 0$, let Q_i be the rightmost internal quantifier followed by at least one external quantifier. Say Q_i is $\forall$. It is followed by a string of external quantifiers and then an internal formula. If it is followed only by universal external quantifiers, use the trivial rule (2.9) to pull them through to the left, thereby reducing the rank to $j - 1$. If it is followed by both universal and existential external quantifiers, use (S') and (2.9) to pull the universal external quantifiers through to the left. Then it is followed by a string of existential external quantifiers, which may be treated as one by taking an ordered tuple, and an internal formula. Use (I) to pull the existential external quantifier to the left, so that Q_i is followed only by an internal formula and the rank is reduced to $j - 1$. (If Q_i is $\exists$ proceed by duality.)

Step 5. Repeat Step 4 until a formula of rank 0 is obtained.

Step 6. Use (T') to replace all external quantifiers by the corresponding internal quantifiers. This gives an internal formula, with the same free variables as the original formula, which is equivalent to the original formula for all standard values of the free variables.

We remark that it is not always necessary to carry out Step 3 fully, and that doing so may introduce needless complications.

In the following lexicon, A and B are internal formulas. We give the reductions of the frequently occurring patterns $\forall \exists^{st} \forall^{st}$, $\forall \exists \forall^{st}$, $\forall(\forall^{st} \Rightarrow \forall^{st})$, $\forall(\forall^{st} \Leftrightarrow \forall^{st})$, and $\forall(\forall^{st} \Rightarrow \exists \forall^{st})$.

Lexicon

$$\forall x \exists^{st} y \forall^{st} z\, A(x,y,z) \equiv \forall \tilde{z} \exists^{fin} y' \forall x \exists y \in y'\, A(x,y,\tilde{z}(y)),$$

$$\forall x \exists y \forall^{st} z\, A(x,y,z) \equiv \forall x \forall^{fin} z' \exists y \forall z \in z'\, A(x,y,z),$$

$$\forall x (\forall^{st} y\, A(x,y) \Rightarrow \forall^{st} z\, B(x,z))$$

$$\equiv \forall z \exists^{fin} y' \forall x (\forall y \in y'\, A(x,y) \Rightarrow B(x,z)),$$

$$\forall t(\forall^{\text{st}} x \, A(t,x) \Leftrightarrow \forall^{\text{st}} y \, B(t,y))$$

$$\equiv \forall y \forall z \exists^{\text{fin}} x' \exists^{\text{fin}} w' \forall t [(\forall x \in x' \, A(t,x) \Rightarrow B(t,y))$$
$$\wedge \, (\forall w \in w' \, B(t,w) \Rightarrow A(t,z))],$$

$$\forall x(\forall^{\text{st}} y \, A(x,y) \Rightarrow \exists z \forall^{\text{st}} w \, B(x,z,w))$$

$$\equiv \forall^{\text{fin}} w' \exists^{\text{fin}} y' \forall x(\forall y \in y' \, A(x,y) \Rightarrow \exists z \forall w \in w' \, B(x,z,w)).$$

Of these, the first four are straightforward applications of the reduction algorithm. Let us reduce the last one, $\forall(\forall^{\text{st}} \Rightarrow \exists\forall^{\text{st}})$, to illustrate the above remark about not carrying out Step 3 fully. By (I), the formula $\exists z \forall^{\text{st}} w \, B(x,z,w)$ is equivalent to $\forall^{\text{st fin}} w' \exists z \forall w \in w' \, B(x,z,w)$. Then $\forall^{\text{st fin}} w'$ pulls through all the way to the left, and $\forall^{\text{st}} y$ may be pulled out of the parentheses as $\exists^{\text{st}} y$. Therefore the formula is equivalent to

$$\forall^{\text{st fin}} w' \forall x \exists^{\text{st}} y(A(x,y) \Rightarrow \exists z \forall w \in w' \, B(x,z,w)).$$

By (I) again this is equivalent to

$$\forall^{\text{st fin}} w' \exists^{\text{st fin}} y' \forall x \exists y \in y'(A(x,y) \Rightarrow \exists z \forall w \in w' \, B(x,z,w)).$$

We choose to push $\exists y \in y'$ back inside the parentheses, where it becomes $\forall y \in y'$. Then by (T') we have the asserted result.

We will give several illustrations here, and there will be more in §4. Consider the formula $\forall y(y \simeq x \Rightarrow f(y) \simeq f(x))$. Here the free variables are a function $f: \mathbf{R} \to \mathbf{R}$ and a point x in $\mathbf{R}$, and the bound variable y ranges over $\mathbf{R}$. Replacing $\simeq$ by its definition, we see that this is equivalent to

$$\forall y(\forall^{\text{st}} \delta \, |y - x| \leqslant \delta \Rightarrow \forall^{\text{st}} \varepsilon \, |f(y) - f(x)| \leqslant \varepsilon)$$

where δ and ε range over the strictly positive real numbers. By $\forall(\forall^{\text{st}} \Rightarrow \forall^{\text{st}})$ in the lexicon, this is equivalent (for f and x standard) to

$$\forall \varepsilon \exists^{\text{fin}} \delta' \forall y(\forall \delta \in \delta' \, |y - x| \leqslant \delta \Rightarrow |f(y) - f(x)| \leqslant \varepsilon).$$

For δ' a finite set, $\forall \delta \in \delta' |y - x| \leqslant \delta$ is the same as $|y - x| \leqslant \delta$ for $\delta = \min \delta'$, and so our formula is equivalent (for f and x standard) to

$$\forall \varepsilon \exists \delta \forall y(|y - x| \leqslant \delta \Rightarrow |f(y) - f(x)| \leqslant \varepsilon).$$

Thus the definition given in §1 of f being continuous at x is equivalent to the usual one, and similarly for f being uniformly continuous on E.

We say that a function $f: \mathbf{R} \to \mathbf{R}$ is *S-continuous* at the point x in $\mathbf{R}$ in case $\forall y(y \simeq x \Rightarrow f(y) \simeq f(x))$. If f and x are standard, then f is S-continuous at x if and only if f is continuous at x. This is not true in general, however. Let $f(x) = x^2$ for all x in $\mathbf{R}$. Then f is continuous at all x in $\mathbf{R}$. Let x be an unlimited real number and let $y = x + x^{-1}$. Then $y \simeq x$ but $y^2 = x^2 + 2 + x^{-2}$ is not infinitely close to x^2, so that f is not S-continuous at x. Again, let $f(x) = \alpha/(\alpha^2 + x^2)$ where $\alpha \neq 0$. The function f is continuous at all real x. However, if α is infinitesimal then f is not S-continuous at 0.

If we apply the reduction algorithm to Theorem 1.2 we obtain the following

triviality: For all finite sets x' there exists a finite set F such that for all x in x' we have $x \in F$. This helps remove any aura of paradox surrounding this theorem.

Theorem 1.4 implies that every limited real number is infinitely close to a standard real number:

$$\forall x (\exists^{st} r \ |x| \leqslant r \Rightarrow \exists^{st} a \forall^{st} \varepsilon \ |x - a| \leqslant \varepsilon).$$

Here ε is restricted to strictly positive values. This is equivalent to $\forall^{st} r \forall x \exists^{st} a \forall^{st} \varepsilon (|x| \leqslant r \Rightarrow |x - a| \leqslant \varepsilon)$. Treat r as a standard parameter and refer to $\forall \exists^{st} \forall^{st}$ in the lexicon. Thus we obtain

$$\forall r \forall \tilde{\varepsilon} \exists^{fin} a' \forall x \exists a \in a' (|x| \leqslant r \Rightarrow |x - a| \leqslant \tilde{\varepsilon}(a)),$$

which is the Heine-Borel theorem for intervals $[-r, r]$.

There are two general principles which we can deduce from the form of the reduction algorithm.

Notice that the reduction of $\forall x \exists y \forall^{st} z \ A(x, y, z)$ is the same as the reduction of $\forall^{st} x \exists y \forall^{st} z \ A(x, y, z)$. Therefore from the latter we may infer the former, even though the transfer principle is not applicable since $\exists y \forall^{st} z \ A(x, y, z)$ is an external formula. We call a formula *universally semi-internal* in case, when it is written in the form of Step 3 above, the only external quantifiers are $\forall^{st}$, and we define *existentially semi-internal* similarly.

THEOREM 2.1 (GENERAL TRANSFER PRINCIPLE). *Let $A(x, t_1, \ldots, t_k)$ be a universally semi-internal formula with free variables $x, t_1, \ldots, t_k$ and no other free variables. Then*

$$(2.10) \qquad \forall^{st} t_1 \cdots \forall^{st} t_k (\forall^{st} x \ A(x, t_1, \ldots, t_k) \Rightarrow \forall x \ A(x, t_1, \ldots, t_k)).$$

PROOF. When we apply the reduction algorithm to $A(x, t_1, \ldots, t_k)$ we get a string of quantifiers $\forall^{st}$ followed by an internal formula. We may push $\forall^{st} x$ past them, change it to $\forall x$ by (T), and then pull it back to the left. (Notice that (S′) is not used in the reduction, so we do not need to worry about relativization.) Q.E.D.

By duality, if $A(x)$ is an existentially semi-internal formula then $\exists x \ A(x) \equiv \exists^{st} x \ A(x)$.

We pointed out in §1 that, because of the transfer principle, if $A(x)$ is an internal formula whose only free variable is x and there exists a unique x such that $A(x)$ then that x is standard. Next we show that this remains true even if $A(x)$ is an external formula.

THEOREM 2.2 (UNIQUENESS PRINCIPLE). *Let $A(x)$ be a formula of IST (relativized to a standard set V, which we treat as a constant), internal or external, whose only free variable is x. If there exists a unique x such that $A(x)$ then there exists a standard x such that $A(x)$.*

PROOF. Apply the reduction algorithm to $A(x)$ through Step 5 but omitting the last step. By repeated applications of (S′) and taking tuples, we may write $A(x)$ in the form $\forall^{st} u \exists^{st} v \ B(x, u, v)$ where $B(x, u, v)$ is an internal formula

whose only free variables are x, u, and v. Suppose that there is a unique x such that $A(x)$, and let x be the set such that $A(x)$. Then $\forall^{st} u \exists^{st} v\ B(x, u, v)$ or, equivalently,

$$(2.11) \qquad \exists^{st} \tilde{v} \forall^{st} u\ B(x, u, \tilde{v}(u)).$$

Also $\forall y (\forall^{st} u \exists^{st} v\ B(y, u, v) \Rightarrow y = x)$, which reduces to

$$(2.12) \qquad \forall^{st} \tilde{v} \exists^{\text{st fin}} u' \forall y (\forall u \in u'\ B(y, u, \tilde{v}(u)) \Rightarrow y = x).$$

Let $\tilde{v}$ be a standard function such that

$$(2.13) \qquad \forall^{st} u\ B(x, u, \tilde{v}(u)).$$

This is possible by (2.11). Let u' be a standard finite set such that

$$(2.14) \qquad \forall y (\forall u \in u'\ B(y, u, \tilde{v}(u)) \Rightarrow y = x).$$

This is possible by (2.12). By Theorem 1.1 every element u of u' is standard, so by (2.13) we have $\forall u \in u'\ B(x, u, \tilde{v}(u))$, which implies that

$$(2.15) \qquad \exists y \forall u \in u'\ B(y, u, \tilde{v}(u)).$$

But $\forall u \in u'\ B(y, u, \tilde{v}(u))$ is an internal formula whose only free variables are y, u', and $\tilde{v}$, and since u' and $\tilde{v}$ have standard values we may apply the transfer principle to conclude that

$$(2.16) \qquad \exists^{st} y \forall u \in u'\ B(y, u, \tilde{v}(u)).$$

By (2.16) and (2.14) we have $\exists^{st} y\ (y = x)$, and so x is standard. Q.E.D.

If $A(x)$ contains other free variables then the theorem remains true for all standard values of them by the same proof (or equivalently we may allow $A(x)$ to contain standard constants).

3. External sets. The usual presentations of nonstandard analysis have an advantage which internal set theory lacks; namely, that external predicates may be used to define subsets. It is an easily proved theorem of IST that for all limited real numbers x and y and all infinitesimals α and β we have that $x + y$ and xy are limited, $\alpha + \beta$ and $x\alpha$ are infinitesimal, and if x is not infinitesimal, then x^{-1} is limited. How much simpler it would be to say that the limited real numbers are an integral domain in which the infinitesimals are a maximal ideal, but this is illegal set formation and cannot be formulated within IST.

There is a simple way in which we may achieve the freedom to use such language. Powell showed that IST is a conservative extension of ZFC (see the Appendix). If we assume that ZFC is consistent we conclude that IST is consistent, and so by Gödel's completeness theorem [1], IST has a model.

Consider a model of IST. We want to continue referring to the entities of IST as sets, and we want to continue to denote the elementhood relation in IST by $\in$. To avoid confusion, we will refer to the "sets" of the set theory in which the model is being considered as *external sets*, and we will denote the elementhood relation between external sets by $\mathbf{\in}$. Also, we will use bold-face type when forming ordered pairs, etc. of external sets.

Thus a model of IST consists of an external set **I**, an external subset **S** of **I**, and an external subset **T** of **I** $\times$ **I** such that with the interpretations

$$\langle x, y \rangle \in T \quad \text{for } x \in y, \qquad x \in S \quad \text{for } x \text{ standard,}$$

all of the axioms of IST hold for **I**.

If $x \in$ **I** we let

$$^*x = \left\{ x \in I : y \in x \right\}$$

so that for all $y \in$ **I** we have $y \in {}^*x \Leftrightarrow y \in x$. An external set which is of the form $^*x$ for some $x \in$ **S** is called *standard*, an external set which is of the form $^*x$ for some $x \in$ **I** is called *internal*, and an external set which is not internal is called *strictly external*.

Let **R** be the real numbers, and consider a model of IST. Then *R is a standard external set. We define the external sets M_0, M_1, and R_{st} by

$$M_0 = \left\{ x \in {}^*R : x \text{ is limited} \right\},$$

$$M_1 = \left\{ x \in {}^*R : x \text{ is infinitesimal} \right\},$$

$$R_{st} = \left\{ x \in {}^*R : x \text{ is standard} \right\}.$$

With the zero element 0, unity 1, addition defined by $\langle x, y \rangle \mapsto x + y$, multiplication defined by $\langle x, y \rangle \mapsto xy$, and the order relation $\{\langle x, y \rangle \in {}^*R \times {}^*R : x < y\}$, it is clear that *R is an ordered field. It is not an Archimedean ordered field, for if x is any infinitesimal with $x > 0$, then for every external natural number n we have that $nx < 1$, as is easily seen by induction. The external set M_0 is strictly external, for if it were of the form $^*X$ then X would be a proper subring of **R** containing the unit interval. If M_1 were internal then M_0 would be internal, since

$$M_0 = \left\{ x \in {}^*R : \forall y \in M_1 xy \in M_1 \right\},$$

and if R_{st} were internal then M_0 would be internal, since

$$M_0 = \left\{ x \in {}^*R : \exists r \in R_{st} |x| < r \right\}.$$

Thus M_0, M_1, and R_{st} are all strictly external. It is easily seen that M_0 is an integral domain in which M_1 is a maximal ideal, and that the quotient field M_0/M_1 is isomorphic to R_{st} under the isomorphism $x \mapsto \text{st } x$.

Model theory is a powerful subject which has had stunning applications to algebra and number theory. In this paper we are concerned with those aspects of nonstandard analysis which do not involve model theory. We will work within IST but we shall on occasion make use of the extra freedom and simplicity which the language of external sets affords. We will be pedantic about the distinction between sets and external sets, for it must be borne in mind that external sets are not entities of IST.

4. A grab-bag of nonstandard analysis. This section consists of miscellaneous examples which illustrate the use of (I), (S), (T), the reduction of external definitions and theorems, and the notion of an external set. Since the reader may wish to work some things out, we have broken the discussions in two. The conclusions will be found at the end of this section, with the same numbering.

EXAMPLE 1. THEOREM. *Let* $f: E \to R$ *be continuous, where E is a closed and bounded subset of* **R**. *Then f achieves its maximum.* PROOF. We may assume that f and E are standard. Let F be a finite subset of E containing all the standard

points in E, and let x be an element of F at which the restriction of f to F achieves its maximum ....

EXAMPLE 2. THEOREM. *Let f: $[a, b] \to \mathbf{R}$ be continuous, with $f(a) < 0$ and $f(b) > 0$. Then there is a c in $[a, b]$ with $f(c) = 0$.* PROOF. We may assume that f, a, and b are standard. Let F be a finite subset of $[a, b]$ containing all the standard points, and let x be the least element of F such that $f(x) \geqslant 0$ ....

EXAMPLE 3. Which is more intuitive, the conventional definition of continuity or the nonstandard definition? Here is an unfamiliar notion with the same level of difficulty as the notion of continuity. We present it in two equivalent forms: (i) A function f: $\mathbf{R} \to \mathbf{R}$ is *suounitnoc* at the point x in case for all ε there is a δ such that for all y, if $|y - x| \leqslant \varepsilon$ then $|f(y) - f(x)| \leqslant \delta$. (ii) A function f: $\mathbf{R} \to \mathbf{R}$ is *suounitnoc* at the point x in case (for f and x standard) for all y, if $y - x$ is limited then $f(y) - f(x)$ is limited. Using either definition, show that if a function is suounitnoc at one point then it is suounitnoc at all points ....

EXAMPLE 4. A *graph* is a set G and a subset R of $G \times G$. If k is a natural number, a *k-coloring* of the graph $\langle G, R \rangle$ is a function f: $G \to \{1, \ldots, k\}$ such that for all $\langle x, y \rangle \in R$ we have $f(x) \neq f(y)$. THEOREM. *If every finite subset of a graph has a k-coloring then the graph has a k-coloring.* PROOF. We may assume that G, R, and k are standard. Let F be a finite subset of G containing all the standard points ....

EXAMPLE 5. The prefix S- is used in defining an external relation which for all standard values of the arguments agrees with the familiar internal relation which is being generalized. An example of this is the definition given in §2 of what it means for f to be S-continuous at x. Let "nice" denote some familiar internal n-ary relation. Whenever we define $\langle x_1, \ldots, x_n \rangle$ to be S-nice in case $A(x_1, \ldots, x_n)$ we have the moral obligation to prove that for all standard $x_1, \ldots, x_n$ we have $A(x_1, \ldots, x_n)$ if and only if $\langle x_1, \ldots, x_n \rangle$ is nice. DEFINITIONS. Let s: $\mathbf{N} \to \mathbf{R}$ and let a be in $\mathbf{R}$. Then the sequence s *S-converges* to a in case $s_n \simeq a$ for all unlimited n, and a is limited. The sequence s is *S-Cauchy* in case $s_n \simeq s_m$ for all unlimited natural numbers n and m ....

EXAMPLE 6. THEOREM. *If S is any subset of $\mathbf{N}$ which contains all unlimited natural numbers then the complement of S is a standard finite set* ....

EXAMPLE 7. The sequence $s_n = 1/n$ (for $n > 0$) is infinitesimal for all unlimited natural numbers n. May we conclude from Example 6 above that $1/n$ is infinitesimal except on a standard finite set?

EXAMPLE 8. THEOREM. *Let s: $\mathbf{N} \to \mathbf{R}$ and suppose that s_n is infinitesimal for all limited natural numbers n. Then there exists an unlimited natural number ν such that s_n is infinitesimal for all $n \leqslant \nu$.* PROOF. Apply the reduction algorithm to the conclusion

$$\exists \nu \{ (\forall^{\mathrm{st}} m (\nu \geqslant m)) \wedge \forall n (n \leqslant \nu \Rightarrow \forall^{\mathrm{st}} \varepsilon > 0 \, |s_n| \leqslant \varepsilon) \} \ldots .$$

EXAMPLE 9. Let K be a standard ordered field. Then the definitions of infinitesimal, limited element of K, unlimited element of K, and infinitely close make sense in K. THEOREM. *A standard ordered field K is isomorphic to the field $\mathbf{R}$ of real numbers if and only if every limited element of K is infinitely close to a standard element of K.* PROOF. Necessity was proved in Theorem 1.4. To prove sufficiency, we need only show that the least upper bound axiom holds. Let E

be a nonempty subset of K which is bounded above . . . .

EXAMPLE 10. Let $A(n)$ be a formula, internal or external, with free variable n and possibly others. THEOREM (EXTERNAL INDUCTION). *Suppose that $A(0)$ and for all standard natural numbers n, if $A(n)$ then $A(n + 1)$. Then for all standard natural numbers n we have $A(n)$* . . . .

EXAMPLE 11. Let X be a standard set, let Y be a subset of X, and let $Z = {}^S\{z \in X: z \in Y\}$. Then Z is a standard subset of X such that for all standard z in X we have $z \in Z \Leftrightarrow z \in Y$. Thus we have the following

$$\forall^{st} X \forall Y \in P(X) \exists^{st} Z \in P(X) \forall^{st} z \in X(z \in Z \Leftrightarrow z \in Y).$$

If we refer to $\forall \exists^{st} \forall^{st}$ in the lexicon of §2, treating X as a standard parameter, we obtain the following internal theorem: Let X be a set and let $f: P(X) \rightarrow X$. Then there is a finite number $Z_1, \ldots, Z_n$ of subsets of X such that for all subsets Y of X we have

$$(f(Z_1) \in Y \Leftrightarrow f(Z_1) \in Z_1) \vee \cdots \vee (f(Z_n) \in Y \Leftrightarrow f(Z_n) \in Z_n).$$

It is amusing to find a conventional proof of this . . . .

EXAMPLE 12. Let X be a set. A *filter* on X is a family $\mathcal{F}$ of subsets of X such that $\varnothing \notin \mathcal{F}$, $X \in \mathcal{F}$, and for all subsets E and F of X we have $E \in \mathcal{F} \wedge F \in \mathcal{F} \Rightarrow E \cap F \in \mathcal{F}$ and $E \in \mathcal{F} \wedge E \subset F \Rightarrow F \in \mathcal{F}$. A filter $\mathcal{F}$ is called *principal* in case it contains a minimal element M, in which case $\mathcal{F}$ consists of all supersets of M. Let X be a standard set and let $\mathcal{F}$ be a standard filter on X. Then the *monad* $\mu(\mathcal{F})$ is the external set

$$\mu(\mathcal{F}) = \bigcap_{\substack{E \in \mathcal{F} \\ E \text{ standard}}} {}^*E.$$

An element F of $\mathcal{F}$ is called *infinitesimal* in case ${}^*F \subset \mu(\mathcal{F})$. THEOREM. *Every standard filter has infinitesimal elements* . . . .

EXAMPLE 13. THEOREM. *Let $\mathcal{F}$ be a standard filter on the standard set X. If the monad of $\mathcal{F}$ is internal then it is standard, and $\mathcal{F}$ is principal.* PROOF. Suppose that $\mu(\mathcal{F}) = {}^*M$ for some subset M of X. By Example 12 above, there is an F in $\mathcal{F}$ with ${}^*F \subset {}^*M$, so that $F \subset M$ and thus $M \in \mathcal{F}$. Then M may be characterized as the largest infinitesimal element of $\mathcal{F}$ . . . .

EXAMPLE 14. THEOREM. *Let $\mathcal{F}$ be a standard filter on the standard set X. If an internal set intersects ${}^*E$ for every standard element E of $\mathcal{F}$ then it intersects $\mu(\mathcal{F})$.* PROOF. Formulate this as a statement of IST and refer to the lexicon . . . .

EXAMPLE 15. If x is a point in the topological space X then the family $\mathfrak{N}_x$ of all neighborhoods of x is a filter on X, and it is standard if x and X are standard, in which case $\mu(\mathfrak{N}_x)$ is called the *monad* of x and is denoted by $\mu(x)$. If E is a subset of X, we define the *standard part* of E, denoted st E, by st $E = {}^S\{y \in X: {}^*E \cap \mu(y) \neq \varnothing\}$. THEOREM. *Let E be a subset of the standard topological space X. Then st E is closed* . . . .

EXAMPLE 16. What do you get if you apply the reduction algorithm to the theorem of Example 15 above? . . .

CONCLUSIONS . . . 1. Then st x is in E and consequently in F. By continuity, $f(x) \simeq f(\text{st } x)$. Therefore, for all standard y in E we have $f(\text{st } x) \geq f(y) + \alpha$ where $\alpha \simeq 0$, since $f(x) \geq f(y)$. Since both $f(\text{st } x)$ and $f(y)$ are standard, we must have $f(\text{st } x) \geq f(y)$. By transfer, $f(\text{st } x) \geq f(y)$ for all y in E. Q.E.D.

...2. Then st x is in $[a, b]$ and by continuity $f(\text{st } x) \simeq f(x)$, so that $f(\text{st } x) \geqslant \alpha$ where $\alpha \simeq 0$. Since $f(\text{st } x)$ is standard, $f(\text{st } x) \geqslant 0$. Let y be the element of F preceding x. We must have $y \simeq x$ since F contains all standard elements of $[a, b]$, and consequently $f(y) \simeq f(\text{st } x)$. But $f(y) < 0$ and so we conclude that $f(\text{st } x) \leqslant 0$. Therefore $f(\text{st } x) = 0$.   Q.E.D.

...3. We use (ii). Since x and $f(x)$ are standard, they are limited. Thus to say that $y - x$ is limited is the same as saying that y is limited, and to say that $f(y) - f(x)$ is limited is the same as saying that $f(y)$ is limited. Thus the standard function f is suounitnoc at the standard point x if and only if it is limited at all limited points, a characterization which does not involve the point x. By transfer, if a function is suounitnoc at one point it is suonitnoc at all points.   Q.E.D.

If we apply the reduction algorithm to the assertion that f is limited at all limited points, we see a function is suounitnoc if and only if it is bounded on all bounded sets. The reduction algorithm also shows that (i) and (ii) are equivalent.

...4. By hypothesis, F has a k-coloring $f: F \to \{1, \ldots, k\}$. By (S) there is a standard function $g: G \to \{1, \ldots, k\}$ such that $\forall^{\text{st}} x \, g(x) = f(x)$. By transfer, g is a k-coloring.   Q.E.D.

...5. The sequence s S-converges to a if and only if

$$\forall n [(\forall^{\text{st}} m \, n \geqslant m) \Rightarrow \forall^{\text{st}} \varepsilon > 0 \, |s_n - a| \leqslant \varepsilon].$$

Pulling out the quantifiers, we obtain $\forall^{\text{st}} \varepsilon > 0 \, \forall n \exists^{\text{st}} m$, and applying (I) to $\forall n \exists^{\text{st}} m$ and then (T) (for s and a standard) we obtain the usual definition. The reduction of the notion S-Cauchy follows the same pattern.

...6. This follows immediately from Theorem 1.1, since every element of the complement of S is standard.

...7. No. If we try to form "the set of all natural numbers $n > 0$ such that $1/n$ is infinitesimal" we commit illegal set formation.

...8. We obtain, using (I) and the facts that the smallest ε of a finite set and the largest m of a finite set are just as forceful as the whole finite set,

$$\forall^{\text{st}} \varepsilon > 0 \forall^{\text{st}} m \exists \nu \{(\nu \geqslant m) \wedge \forall n (n \leqslant \nu \Rightarrow |s_n| \leqslant \varepsilon)\}.$$

But this is true by hypothesis (take $\nu = m$).   Q.E.D.

...9. We may assume that E is standard. Let F be a finite subset of E containing all the standard points, and let x be the largest element of F. Since E is standard and bounded above, it is bounded above by a standard element r, by (T). Since E is standard and nonempty, it contains a standard element s, by (T). Therefore $s \leqslant x \leqslant t$, so that x is a limited element. By hypothesis, x is infinitely close to a standard element a. One verifies easily that a is the least upper bound of E.   Q.E.D.

...10. This is an immediate consequence of the induction theorem applied to $^S\{n \in \mathbf{N}: A(n)\}$.

...11. We can, in fact, choose $n = 2$. If we can find Z_1 and Z_2 with $f(Z_1) = f(Z_2)$ and $f(Z_1) \in Z_1$ but $f(Z_2) \notin Z_2$, then we are done, because if $f(Z_1) \in Y$ then $f(Z_1) \in Y \Leftrightarrow f(Z_1) \in Z_1$, while if $f(Z_1) = f(Z_2) \notin Y$ then $f(Z_2) \in Y \Leftrightarrow f(Z_2) \in Z_2$. Suppose that there do not exist such sets Z_1 and

Z_2. Then the range of f must be the disjoint union of two sets, say V and W, such that for all subsets Y of X we have $f(Y) \in V \Rightarrow f(Y) \in Y$ and $f(Y) \in W \Rightarrow f(Y) \notin Y$. But then $f(W) \in W \Rightarrow f(W) \notin W$, so we must have $f(W) \notin W$, so that $f(W) \in V$, which implies that $f(W) \in W$, which is a contradiction. Notice the similarity to Russell's paradox.

... 12. To say that $*F \subset \mu(\mathcal{F})$ is the same as saying that $\forall^{st} E \in \mathcal{F}(F \subset E)$, and since $\mathcal{F}$ is closed under the formation of finite intersections it is an immediate consequence of (I) that every standard filter has infinitesimal elements.

... 13. By the uniqueness principle (Theorem 2.2), M is standard. Now we may use (T) and conclude that $\forall E \in \mathcal{F}(M \subset E)$, so that $\mathcal{F}$ is principal. Q.E.D. The uniqueness principle may also be used to show that M_0, M_1, and R_{st} of §3 are strictly external (because they are not standard, but if of the form $*E$ then E is uniquely describable within IST).

... 14. Letting E and F range over $\mathcal{F}$ and letting Y and Z range over $P(X)$, we may reformulate the theorem as

$$\forall Y(\forall^{st} E(E \cap Y \neq \varnothing) \Rightarrow \exists Z \forall^{st} F(Z \subset F \wedge Y \cap Z \neq \varnothing)).$$

This is of the form $\forall(\forall^{st} \Rightarrow \exists \forall^{st})$ in the lexicon of §2, and so is equivalent to the assertion that if $F_1, \ldots, F_n$ are in $\mathcal{F}$ then there exist $E_1, \ldots, E_m$ in $\mathcal{F}$ such that for all subsets Y of X we have

$$Y \cap E_1 \neq \varnothing \wedge \cdots \wedge Y \cap E_m \neq \varnothing$$
$$\Rightarrow \exists Z(Z \subset F_1 \wedge \cdots \wedge Z \subset F_n \wedge Y \cap Z \neq \varnothing).$$

But this is a true statement: take $m = 1$ and take $E_1 = Z = F_1 \cap \cdots \cap F_n$. Q.E.D.

... 15. PROOF. Let z be a standard point in the closure of st E. Let U be a standard open neighborhood of z. Then U intersects st E, so by (T) there is a standard point y in $U \cap$ st E. By definition of st E we have that $*E$ intersects $\mu(y)$, and $\mu(y) \subset *U$ so that $*E$ intersects $*U$. Since this is true for all standard open neighborhoods U of z, we have, by Example 14 above, that $*E$ intersects $\mu(z)$, so that by definition of st E we have $z \in$ st E. Thus every standard point in the closure of st E is in st E, and by transfer every point in the closure of st E is in st E, so that st E is closed. Q.E.D.

... 16. Letting capital letters range over subsets of the standard topological space X and small letters range over points of X, we may formulate the theorem of Example 15 as $\forall E \forall^{st} F\{[F = \text{st } E] \Rightarrow F \text{ closed}\}$; that is, as

$$\forall E \forall^{st} F\{[\forall^{st} y(y \in F \Leftrightarrow \exists z$$
$$(\forall^{st} N(N \text{ neighborhood of } y \Rightarrow z \in N) \wedge z \in E))] \Rightarrow F \text{ closed}\}.$$

If you apply the reduction algorithm to this you get a mess.

5. Infinitesimal calculus. The purpose of this section is to show how simple the calculus becomes when one uses nonstandard analysis. It should be read only by mathematicians who have struggled to do an honest job of teaching the mean value theorem to freshmen, for only they appreciate the difficulties

of motivating and explaining the traditional approach to the calculus. One technical point should be commented on before we begin: Newton had an awkward notion of derivative. For him the derivative of f at x is the limiting value of the slope of a line through $\langle x, f(x) \rangle$ and a nearby point on the graph. A much better definition is that the derivative of f at x is the limiting value of the slope of a line through two nearby points (neither of which has to be held fixed at $\langle x, f(x) \rangle$). This more restrictive notion of smoothness of a function at a point together with the basic notions of nonstandard analysis make possible an approach to the calculus in which intuition and rigor are fully compatible. To avoid misunderstanding, we declare that the following exposition is not intended for freshmen.

Let $f: I \to \mathbf{R}$, where I is an interval, and let $x \in I$. We say that f is *derivable* at x in case (for f and x standard) there is a standard number $f'(x)$ such that whenever x_1 and x_2 are distinct points in I infinitely close to x we have

$$(5.1) \qquad \frac{f(x_2) - f(x_1)}{x_2 - x_1} \simeq f'(x).$$

Letting $x_2 = x_1 + h$, we may restate this as: whenever $x_1 \simeq x$ and $h \simeq 0$, with x_1 and $x_1 + h$ in I, we have

$$(5.2) \qquad f(x_1 + h) = f(x_1) + f'(x)h + \alpha h$$

where $\alpha \simeq 0$. The number $f'(x)$ is unique; it is called the *derivative* of f at x. If we let $x_1 = x$ in (5.2) we see that if f is derivable at x then it is continuous at x. We say that f is *derivable* in case it is derivable at all points in I.

THEOREM 5.1. *Let f and g be derivable at x. Then so are $f + g$ and fg, with*

$$(5.3) \qquad (f + g)'(x) = f'(x) + g'(x),$$

$$(5.4) \qquad (fg)'(x) = f'(x)g(x) + f(x)g'(x),$$

and so is f/g provided that $g(x) \neq 0$, with

$$(5.5) \qquad \left(\frac{f}{g}\right)'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{g(x)^2}.$$

Let g be derivable at x and let f be derivable at $g(x)$. Then $f \circ g$ is derivable at x, with

$$(5.6) \qquad (f \circ g)'(x) = f'(g(x))g'(x).$$

PROOF. We may assume that f, g, and x are standard. Let $x_1 \simeq x$ and $h \simeq 0$. Then

$$f(x_1 + h) = f(x_1) + f'(x)h + \alpha h,$$

$$g(x_1 + h) = g(x_1) + g'(x)h + \beta h,$$

with $\alpha \simeq 0$ and $\beta \simeq 0$. Therefore

$$f(x_1 + h) + g(x_1 + h) = f(x_1) + g(x_1) + [f'(x) + g'(x)]h + \{\alpha + \beta\}h.$$

Since $\alpha + \beta$ is infinitesimal, $f + g$ is derivable at x and (5.3) holds. Also

$$f(x_1 + h)g(x_1 + h) = f(x_1)g(x_1) + [f'(x)g(x_1) + f(x_1)g'(x)]h$$
$$+ \{f(x_1)\beta + f'(x)g'(x)h + \alpha g(x_1)\}h.$$

The term in brackets is infinitely close to $f'(x)g(x) + f(x)g'(x)$, by the continuity of f and g at x, and the term in braces is infinitesimal, so that fg is deribable at x and (5.4) holds. Suppose that x_1 and x_2 are distinct points infinitely close to x, and suppose that $g(x) \neq 0$. Since $g(x)$ is standard and $g(x_1) \simeq g(x) \simeq g(x_2)$, we have that $g(x_1) \neq 0$, $g(x_2) \neq 0$ and

$$\frac{[1/g(x_2)] - [1/g(x_1)]}{x_2 - x_1} = \frac{1}{g(x_2)g(x_1)} \frac{g(x_1) - g(x_2)}{x_2 - x_1} \simeq -\frac{1}{g(x)^2}g'(x).$$

Thus $1/g$ is derivable at x with derivative $-g'(x)/g(x)^2$. By (5.4), $f/g = f(1/g)$ is derivable at x with derivative given by (5.5). Finally, suppose that g is derivable at x, f is derivable at $g(x)$, $x_1 \simeq x$, and $h \simeq 0$. Then there are infinitesimals α and β such that

$$f(g(x_1 + h)) = f(g(x_1) + [g'(x)h + \beta h])$$
$$= f(g(x_1)) + f'(g(x))[g'(x)h + \beta h] + \alpha[g'(x)h + \beta h]$$
$$= f(g(x_1)) + f'(g(x))g'(x)h + \{f'(g(x))\beta + \alpha g'(x) + \alpha\beta\}h.$$

Since the term in braces is infinitesimal, $f \circ g$ is derivable at x and (5.6) holds. Q.E.D.

THEOREM 5.2. *Let f, with domain $[a, b]$, be derivable.*

(i) *If $f'(x) > 0$ for all x in $[a, b]$ then f is a strictly increasing function from $[a, b]$ onto $[f(a), f(b)]$, and the inverse function is derivable.*

(ii) *If $f'(x) \geqslant 0$ for all x in $[a, b]$ then f is an increasing function from $[a, b]$ onto $[f(a), f(b)]$.*

(iii) *If $f'(x) = 0$ for all x in $[a, b]$ then f is a constant function on $[a, b]$.*

PROOF. To prove (i) we may assume that f, a, and b are standard. Let n be an unlimited natural number and divide $[a, b]$ into n equal subintervals, so that the division points x_j are given by $x_j = a + jh$ for $0 \leqslant j \leqslant n$ where $h = (b - a)/n$. Then we have

$$f(x_{j+1}) = f(x_j) + f'(\text{st } x_j)h + \alpha_j h, \qquad 0 \leqslant j \leqslant n - 1,$$

where $\alpha_j \simeq 0$. Then $|\alpha_j| < f'(\text{st } x_j)$, since $f'(\text{st } x_j) > 0$ is standard, and so $f(x_j) < f(x_{j+1})$. By induction, $f(a) < f(b)$. By the same argument applied to any closed subinterval, we see that f is strictly increasing. To show that f maps $[a, b]$ onto $[f(a), f(b)]$, let c be any standard number with $f(a) < c < f(b)$. We need only produce an x in $[a, b]$ with $f(x) = c$. Let x_j be the largest division point with $f(x_j) \leqslant c$. Then $f(x_{j+1}) > c$. We have $f(\text{st } x_j) \simeq f(x_j)$ and $f(\text{st } x_j) \simeq f(x_{j+1})$, so that $f(\text{st } x_j) \simeq c$ and, consequently, $f(\text{st } x_j) = c$ since both numbers are standard. Now, changing notation, let x_1 and x_2 be any two distinct points in $[a, b]$ and let $y_1 = f(x_1)$, $y_2 = f(x_2)$. If x_1 and x_2 are not infinitely close, there are two distinct standard numbers between them and f

gives two distinct standard numbers between y_1 and y_2. That is, if $y_1 \simeq y_2$ then $x_1 \simeq x_2$. Suppose that $y_1 \simeq y_2$ and let $x = \mathrm{st}\, x_1$. Then $(y_2 - y_1)/(x_2 - x_1) \simeq f'(x)$ so that $(x_2 - x_1)/(y_2 - y_1) \simeq 1/f'(x)$. This proves that the inverse function g is derivable, with $g'(y) = 1/f'(g(y))$, and (i) is proved.

To prove (ii), again assume that f, a, and b are standard. Let α be a strictly positive infinitesimal and let $F(x) = f(x) + \alpha x$. Then F satisfies the hypotheses of (i), and using the fact that $\alpha \simeq 0$ we immediately derive (ii) for f.

Finally, (iii) is an immediate consequence of (ii) applied to f and $-f$. Q.E.D.

We define $f^{(0)} = f, f^{(1)} = f'$ if f is derivable, and by induction $f^{(n+1)} = f^{(n)\prime}$ if $f^{(n)}$ is derivable. Recall (Example 1 of §4) that a continuous function on a closed interval attains its maximum.

Theorem 5.3. *Let $f: I \to \mathbf{R}$, where I is a closed interval, be such that $f^{(n)}$ is derivable, and let $a \in I$. Define $R_{n,a}(x)$ for x in I by*

$$f(x) = f(a) + f'(a)(x - a) + \cdots + \frac{1}{n!}f^{(n)}(a)(x - a)^n + R_{n,a}(x).$$

Then

$$(5.7) \qquad |R_{n,a}x)| \leqslant \frac{1}{(n + 1)!} \max_{y \in I}|f^{(n+1)}(y)|\,|x - a|^{n+1}.$$

Proof. Let $p(x)$ be the polynomial $f(x) - R_{n,a}(x)$, let

$$c = \max_{y \in I}|f^{(n+1)}(y)|,$$

and let

$$d(x) = \frac{1}{(n + 1)!}c(x - a)^{n+1} + p(x) - f(x) = \frac{1}{(n + 1)!}c(x - a)^{n+1} - R_{n,a}(x).$$

Then $d(a) = d'(a) = \cdots = d^{(n)}(a) = 0$ and $d^{(n+1)}(x) = c - f^{(n+1)}(x) \geqslant 0$. Since $d^{(n+1)} \geqslant 0$ and $d^{(n)}(a) = 0$, by Theorem 5.2 we have $d^{(n)} \geqslant 0$ to the right of a. But $d^{(n-1)}(a) = 0$, so by the same argument $d^{(n-1)}(a) \geqslant 0$ to the right of a. By induction, $d \geqslant 0$ to the right of a. A similar argument works to the left of a, which establishes the required upper bound for $R_{n,a}(x)$, and a similar argument establishes the lower bound, so that (5.7) holds. Q.E.D.

It often happens that for unlimited degree n the remainder term $R_{n,a}(x)$ is infinitesimal, in which case the standard part of the polynomial $p(x)$ (of unlimited degree n) gives the exact value of $f(x)$.

Theorem 5.4. *Let $f: I \to \mathbf{R}$, where I is an interval, be continuous, and let $a \in I$. Then there is a unique derivable function F on I, which we denote by*

$$(5.8) \qquad F(x) = \int_a^x f(t)\,dt,$$

such that $F(a) = 0$ and $F'(x) = f(x)$ for all x in I. If f, a, and x are standard, with $a < x$ and $x \in I$, and $h > 0$ is infinitesimal, then

$$(5.9) \qquad \int_a^x f(t)\,dt = \mathrm{st} \sum_{j:a \leqslant jh < x} f(jh)h.$$

PROOF. Let f, a, and x be standard with $x \in I$. For brevity we consider only the case $a < x$. Let h be a strictly positive infinitesimal. Then the sum in (5.9) is limited, since it is bounded by $(x - a)\max\{|f(y)|: y \in [a, b]\}$, and so it has a standard part. We take (5.9) as the definition of the left-hand side (for f, a, and x standard), and define F by (5.8). If x_1 and x_2 are standard points in I with $a < x_1 < x_2$ then

$$(5.10) \qquad\qquad F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(t)\, dt,$$

$$(5.11) \quad (x_2 - x_1)\min_{x_1 \leqslant t \leqslant x_2} f(t) \leqslant \int_{x_1}^{x_2} f(t)\, dt \leqslant (x_2 - x_1)\max_{x_1 \leqslant t \leqslant x_2} f(t),$$

so by transfer these relations hold for all x_1 and x_2 in I with $a < x_1 < x_2$. Now let $x_1 \simeq x_2 \simeq x$ where x is standard. By (5.10) and (5.11),

$$(F(x_2) - F(x_1))/(x_2 - x_1) \simeq f(x),$$

so that F is derivable with derivative f, and clearly $F(a) = 0$. The uniqueness of such a function follows from (iii) of Theorem 5.2, and, consequently, the construction is independent of the choice of h.

THEOREM 5.5. *Let f: $\mathbf{R} \to \mathbf{R}$ be bounded and continuous. For all x_0 in $\mathbf{R}$ there is a derivable function x: $\mathbf{R} \to \mathbf{R}$ such that $x'(t) = f(x(t))$ for all t in $\mathbf{R}$ and $x(0) = x_0$.*

PROOF. We may suppose that f and x_0 are standard. Let $h > 0$ be infinitesimal, and define ξ inductively for integral multiples of h by

$$\xi(0) = x_0, \quad \xi((n + 1)h) = \xi(nh) + f(\xi(nh))h, \qquad n > 0,$$

and similarly for $n < 0$. For t not an integral multiple of h let $\xi(t) = \xi([t/h]h)$. Let c be a standard bound for f. Then $|\xi((n + 1)h)| \leqslant |\xi(nh)| + ch$, so that $|\xi(t)| \leqslant c(|t| + 1) + |x_0|$. Therefore $\xi(t)$ is limited for limited t (that is, ξ is suounitnoc). Using (S), let x be the standard function which for standard t is given by $x(t) = \operatorname{st} \xi(t)$. We have the estimate $|\xi(t) - \xi(s)| \leqslant c|t - s| + \alpha$ where $\alpha \simeq 0$, so that for standard t and s (and, consequently, for all t and s) we have $|x(t) - x(s)| \leqslant c|t - s|$, and x is continuous. For s and t standard we have

$$x(t) - x(s) \simeq \xi(t) - \xi(s) = \sum f(\xi(nh))h \simeq \sum f(x(nh))h$$

by the continuity of f, where the sums are over those n such that nh lies between s and t. Taking standard parts we have

$$x(t) - x(s) = \int_s^t f(x(r))\, dr$$

so that x is derivable with $x'(t) = f(x(t))$, and $x(0) = x_0$. Q.E.D.

The same proof works for time-dependent systems of ordinary differential equations. Notice that Ascoli's lemma, which is the difficult part of the conventional proof, is completely by-passed in the nonstandard approach. The widely held belief that one cannot get something for nothing is a superstition.

We conclude this section with an argument which illustrates the use of the generalized transfer principle (Theorem 2.1).

THEOREM 5.6. *Let* $f: I \to \mathbf{R}$ *where* I *is an interval. If* f *is derivable on* I *then* f' *is continuous on* I.

PROOF. We may assume that f is standard. Let x_1, x_2, and x range over I. We know that

$$\forall^{st} x \forall x_1 \forall x_2 \left\{ x_1 \simeq x \wedge x_2 \simeq x \wedge x_1 \neq x_2 \Rightarrow \frac{f(x_2) - f(x_1)}{x_2 - x_1} \simeq f'(x) \right\}.$$

This implies the weaker statement,

(5.12)

$$\forall^{st} x \exists \delta > 0 \ \forall x_1 \forall x_2$$

$$\left\{ |x_1 - x| \leqslant \delta \wedge |x_2 - x| \leqslant \delta \wedge x_1 \neq x_2 \Rightarrow \frac{f(x_2) - f(x_1)}{x_2 - x_1} \simeq f'(x) \right\},$$

since any infinitesimal $\delta > 0$ will do. But the formula following $\forall^{st} x$ in (5.12) is universally semi-internal (recall that $a \simeq b$ means $\forall^{st} \varepsilon > 0 \ |a - b| \leqslant \varepsilon$) and f and f' are standard, so by the generalized transfer principle we may replace $\forall^{st} x$ by $\forall x$ in (5.12). Now let x_0 be a standard point in I, let $x \simeq x_0$, and let x_1 and x_2 be distinct points within δ of x and infinitely close to x_0. Then $(f(x_2) - f(x_1))/(x_2 - x_1)$ is infinitely close both to $f'(x)$ and to $f'(x_0)$, so that $f'(x) \simeq f'(x_0)$. Therefore f' is continuous.   Q.E.D.

6. Near-standard points. Let X be a standard topological space. A point x in X is called *near-standard* in case there is a standard point y in X such that x is in the monad of y; that is, such that $x \in U$ for all standard neighborhoods U of y. If X is Hausdorff and x is a near-standard point of X then there is a unique standard point y such that $x \in \mu(y)$–we call y the *standard part* of x and denote it by st x. If y is standard and $x \in \mu(y)$ we also say that x is *infinitely close* to y.

THEOREM 6.1. *Let* X *be a standard topological space. Then* X *is compact if and only if every point of* X *is near-standard.*

PROOF. Look up $\forall x \exists^{st} y \forall^{st} U$ (U neighborhood of $y \Rightarrow x \in U$) in the lexicon of §2.   Q.E.D.

THEOREM 6.2. *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. Then* Ω *is compact.*

PROOF. We may assume that $t \mapsto X_t$ is standard. Let $\omega \in \Omega$. By Theorem 6.1 we need only show that ω is near standard. For all standard t there is a standard point $\eta(t)$ in X_t such that $\omega(t) \in \mu(\eta(t))$, so by Theorem 1.3 there is a standard element η of Ω such that for all standard t we have $\omega(t) \in \mu(\eta(t))$. By definition of the product topology, $\omega \in \mu(\eta)$.   Q.E.D.

The external characterization of compactness given by Theorem 6.1 is the basis for many applications of nonstandard analysis. We have seen a number of examples of this already.

Frequently in analysis we wish to construct an object x satisfying certain conditions. Using nonstandard analysis, it is often easy to modify the conditions infinitesimally and to show the existence of an object x which satisfies the modified conditions. However, unless we can show that x is near-standard but is not infinitely close to a trivial object, we have not achieved much of interest. We will exemplify these general remarks by discussing a problem of interest in quantum mechanics. We omit some details of a functional analytic nature–for a full discussion in conventional terms see the original accounts by Charles N. Friedman [3], [4].

Let Δ be the Laplace operator on $L^2(\mathbf{R}^n)$ with the usual domain which makes it a selfadjoint operator and let $H_0 = -\Delta$. The one-parameter unitary group $t \mapsto e^{-itH_0}$ describes the motion of a free Schrödinger particle in $\mathbf{R}^n$. The question is whether a Schrödinger particle can feel the effect of a force concentrated at a single point, the origin. Let V be a bounded real measurable function (whose bound may be unlimited) which vanishes outside an infinitesimal neighborhood of the origin. We will also use the letter V to denote the bounded selfadjoint operator of multiplication by the function V. Then $H = H_0 + V$ is also a selfadjoint operator and $t \mapsto e^{-itH}$ is a one-parameter unitary group describing the motion of a Schrödinger particle which feels the effect of a force concentrated in an infinitesimal neighborhood of the origin. However, it may be near-trivial. We define a topology on the set of all (possibly unbounded) selfadjoint operators H on a Hilbert space $\mathcal{K}$ by requiring (for nets) that $H_\alpha \to H$ if and only if for all t in $\mathbf{R}$ and ψ in $\mathcal{K}$ we have $e^{-itH_\alpha}\psi \to e^{-itH}\psi$. Then it may happen that $H = H_0 + V$ is infinitely close to H_0. In fact, this happens for $n \geqslant 4$, the reason being that the set of ψ in the domain of H_0 which vanish in a neighborhood of the origin is a domain of essential selfadjointness for H_0.

In three dimensions the situation is different. Let $V(x) = 3/4\pi\varepsilon^2$ for $|x| \leqslant \varepsilon$ and $V(x) = 0$ otherwise, where $\varepsilon > 0$ is infinitesimal. Notice that as a distribution $\varepsilon^{-1}V$ is infinitely close to the Dirac δ on $\mathbf{R}^3$, so that as a distribution V itself is infinitely close to 0. Despite this, certain multiples of V have a profoundly disturbing effect on the motion of a Schrödinger particle.

THEOREM 6.3. *Let H_0 and V be as above on $L^2(\mathbf{R}^3)$ and let $H(\alpha) = H + \alpha V$ with α standard. Then the selfadjoint operator $H(\alpha)$ is near-standard, and it is infinitely close to H_0 if and only if α is not of the form $-\pi^3(2n + 1)^2/3$ for some integer n.*

PROOF. On the orthogonal complement of the radial functions in $L^2(\mathbf{R}^3)$, the set of all ψ in the domain of H_0 which vanish in a neighborhood of the origin is a domain of essential selfadjointness for H_0, and we need only consider the subspace of radial functions in $L^2(\mathbf{R}^3)$. But this space is unitarily equivalent to $L^2(0, \infty)$ in such a way that H_0 corresponds to $-d^2/dr^2$ on the domain of all ψ in $L^2(0, \infty)$ which are such that the distribution $-d^2\psi/dr^2$ is in $L^2(0, \infty)$ and such that $\psi(0) = 0$, and V corresponds to multiplication by $(3/4\pi\varepsilon^2)\chi_{[0,\varepsilon]}$. Let $\beta^2 = -3\alpha/4\pi$, so that β is also standard. We may write our operator as $-d^2/dr^2 - (\beta^2/\varepsilon^2)\chi_{[0,\varepsilon]}$. Let $\lambda = \pm i$ and let $\sqrt{\lambda}$ have Re $\sqrt{\lambda} > 0$. Let G_λ be the Green's function; that is, $G_\lambda(r, s)$ is the kernel of the integral operator $(-\lambda - d^2/dr^2 - (\beta^2/\varepsilon^2)\chi_{[0,\varepsilon]})^{-1}$. It is possible, but tedious and

unnecessary, to compute G_λ explicitly. Fix s to be noninfinitesimal. On $[s, \infty)$, $G_\lambda(r, s)$ must be of the form $C \exp(-\sqrt{\lambda}\, r)$, on $[\varepsilon, s]$ it must be of the form $A \exp(\sqrt{\lambda}\, r) + B \exp(-\sqrt{\lambda}\, r)$, and on $[0, \varepsilon]$ it must be of the form $a \sin \sqrt{\beta^2/\varepsilon^2 - \lambda}\, r$ because of the zero boundary condition at the origin. The Green's function is continuous in r, and continuously differentiable except at $r = s$. It is easy to see that A, B, and C must be limited since the norm of the resolvent is 1. Therefore $G_\lambda(\varepsilon, s)$ and $G'_\lambda(\varepsilon, s)$, where the prime denotes differentiation with respect to the first variable, must be limited. That is, $a \sin \sqrt{\beta^2/\varepsilon^2 - \lambda}\, \varepsilon$ and $a\sqrt{\beta^2/\varepsilon^2 - \lambda} \cos \sqrt{\beta^2/\varepsilon^2 - \lambda}\, \varepsilon$ must be limited, and so $a \sin \beta$ and $a\beta\varepsilon^{-1} \cos \beta$ must be limited. But $\beta\varepsilon^{-1} \cos \beta$ is unlimited unless $\cos \beta = 0$; that is, unless α is $-\pi^3/3$ times the square of an odd integer. If $\cos \beta \neq 0$ then a must be infinitesimal, and G_λ is infinitely close to the Green's function for the unperturbed operator, so that $H(\alpha)$ is infinitely close to H_0. If $\cos \beta = 0$ then $G'_\lambda(\varepsilon, s) = 0$ and so G_λ is infinitely close to the Green's function for $-d^2/dr^2$ with the boundary condition $\psi'(0) = 0$ instead of $\psi(0) = 0$.   Q.E.D.

The study of partial differential equations with singular coefficients is an important but relatively unexplored field. Distributions may not be the appropriate coefficients to consider. Using nonstandard analysis, we may allow the coefficients to be qualitatively as smooth as we wish but not require them to be standard. For a well-posed boundary value problem with standard data, when is the solution near-standard? When does a slight change in the coefficients produce an abrupt change in the solution?

7. Finite probability spaces. A *finite probability space* is a finite set Ω and a function pr: $\Omega \to [0, 1]$ such that

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

Let $\langle \Omega, \mathrm{pr} \rangle$ be a finite probability space. An *event* is a subset A of Ω, the *probability* of an event A is

$$\mathrm{Pr}\, A = \sum_{\omega \in A} \mathrm{pr}(\omega),$$

a *random variable* is a function $X: \Omega \to \mathbf{R}$, the *expectation* or *mean* of a random variable X is

$$EX = \sum_{\omega \in \Omega} X(\omega)\,\mathrm{pr}(\omega),$$

and the *variance* of a random variable X is

$$\mathrm{Var}\, X = E(X - EX)^2.$$

If $A(\omega)$ is an internal formula we will frequently abbreviate $\{\omega \in \Omega: A(\omega)\}$ by $\{A\}$. The random variables $X_1, \ldots, X_n$ are called *independent* in case for all real $\lambda_1, \ldots, \lambda_n$ we have

$$\mathrm{Pr}\{X_1 = \lambda_1 \wedge \cdots \wedge X_n = \lambda_n\} = \mathrm{Pr}\{X_1 = \lambda_1\} \cdots \mathrm{Pr}\{X_n = \lambda_n\}.$$

It is obvious that if $f_i: \mathbf{R} \to \mathbf{R}$ and the X_i are independent then the $f_i(X_i)$ are independent. More generally, if $f_i: \mathbf{R}^{k_i} \to \mathbf{R}$ and the X_j are independent then

the random variables $f_1(X_1, \ldots, X_{k_1}), f_2(X_{k_1+1}, \ldots, X_{k_1+k_2}), \ldots$ are independent. The *probability distribution f_X* of a random variable X is the function on **R** defined by

$$f_X(\lambda) = \Pr\{X = \lambda\}.$$

Clearly f_X is 0 except on a finite set and $\sum f_X(\lambda) = 1$. If two random variables have the same probability distribution f then they have the same mean,

$$\mu = \sum \lambda f(\lambda),$$

and variance,

$$\sigma^2 = \sum (\lambda - \mu)^2 f(\lambda).$$

Finite probability spaces are usually used in the discussion of elementary combinatorial problems, such as drawing balls of various colors from urns. We shall attempt to show that if we use nonstandard analysis then the deeper aspects of probability theory may also be formulated in terms of finite probability spaces. The key notion is the following. Let $A(\omega)$ be a formula, internal or external, with ω as free variable and possibly with other free variables. We say that $A(\omega)$ holds *nearly everywhere* (n.e.) on the finite probability space $\langle \Omega, \text{pr} \rangle$ in case for all standard $\varepsilon > 0$ there is an event N with $\Pr(N) \leqslant \varepsilon$ such that for all ω in $\Omega \backslash N$ we have $A(\omega)$. The intuitive meaning of the statement that $A(\omega)$ holds nearly everywhere is that $A(\omega)$ is certain to hold (well, nearly certain).

Let $X_1, \ldots, X_\nu$ be a finite sequence of random variables on a finite probability space. We say that the weak law of large numbers holds in case for all unlimited natural numbers $n \leqslant \nu$ we have

$$(X_1 + \cdots + X_n)/n \simeq E(X_1 + \cdots + X_n)/n \quad \text{n.e.}$$

We say that the *strong law of large numbers* holds in case n.e. for all unlimited $n \leqslant \nu$ we have

$$(X_1 + \cdots + X_n)/n \simeq E(X_1 + \cdots + X_n)/n.$$

Thus for the weak law the exceptional set is allowed to depend on n whereas for the strong law it is not. It is the difference between saying that on any specific date in the future it is very unlikely that there will be an earthquake in San Francisco, and saying that it is very unlikely that there will ever be an earthquake in San Francisco.

THEOREM 7.1. *Let $X_1, \ldots, X_\nu$ be independent random variables on a finite probability space, with the same probability distribution, such that their variance σ^2 is limited. Then the weak law of large numbers holds.*

PROOF. The proof is based on Chebyshev's lemma, which asserts that if X is a random variable and $\lambda > 0$ then

$$\Pr\{X \geqslant \lambda\} \leqslant EX^2/\lambda^2.$$

To see this, observe that

$$EX^2 = \sum_\omega X(\omega)^2 \mathrm{pr}(\omega) \geqslant \sum_{\omega:|X(\omega)|\geqslant\lambda} X(\omega)^2 \mathrm{pr}(\omega) \geqslant \lambda^2 \mathrm{Pr}\{X \geqslant \lambda\}.$$

To apply this, let n be unlimited with $n \leqslant \nu$, let $\lambda > 0$ be standard, let $S_n = X_1 + \cdots + X_n$, and let μ be the mean of the X_i. Then

$$\mathrm{Pr}\left\{\left|\frac{S_n}{n} - \mu\right| \geqslant \lambda\right\} \leqslant \frac{E(S_n/n - \mu)^2}{\lambda^2} = \frac{\sigma^2}{n\lambda^2} \simeq 0.$$

Let $\varepsilon > 0$ be standard and let Λ be the set of all $\lambda > 0$ such that $\mathrm{Pr}\{|S_n/n - \mu| \geqslant \lambda\} \leqslant \varepsilon$. Since Λ contains all standard $\lambda > 0$ it contains an infinitesimal $\lambda > 0$ by (I). By definition, $S_n/n \simeq \mu$ n.e. Q.E.D.

The argument used at the end of the proof is worth recording: if for each standard $\lambda > 0$ we have $\mathrm{Pr}\{|X| \geqslant \lambda\} \simeq 0$ then $X \simeq 0$ n.e. (and conversely).

Recall from Example 5 of §4 the notion of S-convergence. This notion also applies to a finite sequence $x_1, \ldots, x_\nu$: we say that it S-converges to a in case $x_n \simeq a$ for all unlimited $n \leqslant \nu$, and a is limited. With a slight abuse of language we say that a finite series $\sum_{i=1}^\nu x_i$ S-converges to a in case the sequence of partial sums $s_n = x_1 + \cdots + x_n$ S-converges to a. These notions are trivial unless ν is unlimited. Notice that the strong law of large numbers holds for $X_1, \ldots, X_\nu$ if and only if it holds for $X_1 - EX_1, \ldots, X_\nu - EX_\nu$ and that it holds for $X_1, \ldots, X_\nu$ with means 0 if and only if the sequence S_n/n S-converges to 0 n.e., where $S_n = X_1 + \cdots + X_n$. The next theorem gives a criterion for a sequence of random variables to S-converge to 0 n.e.

THEOREM 7.2. *Let $X_1, \ldots, X_\nu$ be random variables on a finite probability space. Then they S-converge to 0 n.e. if and only if for all standard $\lambda > 0$ and unlimited $n \geqslant \nu$ we have*

$$\mathrm{Pr}\left\{\max_{n\leqslant i\leqslant\nu}|X_i| \geqslant \lambda\right\} \simeq 0.$$

PROOF. Suppose that $X_1, \ldots, X_\nu$ S-converges to 0 n.e. Then for all standard $\varepsilon > 0$ there is an event N with $\mathrm{Pr}(N) \leqslant \varepsilon$ such that for all ω in $\Omega\backslash N$, if n is unlimited with $n \leqslant \nu$ then $X_n(\omega) \simeq 0$. If $\lambda > 0$ and $n \leqslant \nu$ let

$$M(n,\lambda) = \left\{\max_{n\leqslant i\leqslant\nu}|X_i| \geqslant \lambda\right\}.$$

Then $M(n,\lambda) \subset N$ if λ is standard and n is unlimited, so that $\mathrm{Pr}\, M(n,\lambda) \leqslant \varepsilon$. Since ε is an arbitrary standard strictly positive number, $\mathrm{Pr}\, M(n,\lambda) \simeq 0$.

Conversely, suppose that $\mathrm{Pr}\, M(n,\lambda) \simeq 0$ for n unlimited and λ standard. Let $\varepsilon > 0$ be standard. Then for all standard natural numbers $j > 0$ we have

(7.1) $$\mathrm{Pr}\, M(n, 1/j) \leqslant \varepsilon/2^j$$

for all unlimited n. Let n_j be the least natural number for which (7.1) holds (note that (7.1) is internal). Thus n_j is standard if j is. Let

$$N = \bigcup_{j=1}^\infty M\left(n_j, \frac{1}{j}\right).$$

Then $\Pr N \leqslant \varepsilon$, and for all ω in $\Omega \backslash N$ and unlimited $n \leqslant \nu$ we have $X_n(\omega) \simeq 0$.   Q.E.D.

Our goal is to prove that the strong law of large numbers holds under hypotheses weaker than those of Theorem 7.1. We will need some preliminary results, the first of which is Kolmogorov's inequality:

THEOREM 7.3. *Let* $X_1, \ldots, X_n$ *be independent random variables on a finite probability space with means* 0 *and variances* σ_i^2. *Then for all* $\lambda > 0$ *we have*

$$(7.2) \qquad \Pr\left\{\max_{i \leqslant n}|X_1 + \cdots + X_i| \geqslant \lambda\right\} \leqslant \sum_{i=1}^{n} \sigma_i^2 \Big/ \lambda^2.$$

PROOF. Let $S_i = X_1 + \cdots + X_i$, let $A = \{\max_{i \leqslant n}|S_i| \geqslant \lambda\}$, and let $A_j = \{|S_i| < \lambda$ for all $i < j$ and $|S_j| \geqslant \lambda\}$ for $j = 1, \ldots, n$. Then A is the disjoint union of the A_j. Since $EX_i = 0$ we have

$$\sum_{i=1}^{n} \sigma_i^2 = ES_n^2 \geqslant E(S_n^2 \chi_A) = \sum_{j=1}^{n} E(S_n^2 \chi_{A_j})$$

$$= \sum_{j=1}^{n} E((S_j + (S_n - S_j))^2 \chi_{A_j})$$

$$= \sum_{j=1}^{n} [E(S_j^2 \chi_{A_j}) + 2E((S_n - S_j)S_j \chi_{A_j}) + E((S_n - S_j)^2 \chi_{A_j})].$$

The third term in the brackets is positive. The second term in the brackets is 0, since $(S_n - S_j)$ and $S_j \chi_{A_j}$ are independent (because they are functions of $X_{j+1}, \ldots, X_n$ and of $X_1, \ldots, X_j$, respectively) so that

$$E((S_n - S_j)S_j \chi_{A_j}) = E(S_n - S_j)E(S_j \chi_{A_j}) = 0\, E(S_j \chi_{A_j}) = 0.$$

The first term in the brackets is $\geqslant \lambda^2 E\chi_{A_j}$. Therefore

$$\sum_{i=1}^{n} \sigma_i^2 \geqslant \sum_{j=1}^{n} \lambda^2 E\chi_{A_j} = \lambda^2 \Pr A,$$

which is (7.2).   Q.E.D.

THEOREM 7.4. *Let* $X_1, \ldots, X_\nu$ *be independent random variables on a finite probability space with means* 0 *and variances* σ_i^2. *If* $\sum_{i=1}^{\nu} \sigma_i^2$ *S-converges then* $\sum_{i=1}^{\nu} X_i$ *S-converges n.e.*

PROOF. Let $\lambda > 0$ be standard and let $n \leqslant \nu$ be unlimited. By Kolmogorov's inequality

$$\Pr\left\{\max_{n \leqslant k \leqslant \nu}\left|\sum_{i=k}^{\nu} X_i\right| \geqslant \lambda\right\} \leqslant \sum_{i=n}^{\nu} \sigma_i^2 \Big/ \lambda^2 \simeq 0.$$

By Theorem 7.2 the sequence $S_\nu - S_n$, where $S_n = X_1 + \cdots + X_n$ as usual, S-converges to 0 n.e., so that the series $\sum_{i=1}^{\nu} X_i$ S-converges n.e. (It is easy to see that the sum is limited n.e.)   Q.E.D.

THEOREM 7.5. *Let* $X_1, \ldots, X_\nu$ *be independent random variables on a finite*

probability space with means 0 *and variances* σ_i^2. *If* $\sum_{i=1}^{\nu} \sigma_i^2/i^2$ *S-converges then the strong law of large numbers holds.*

PROOF. Let $Y_i = X_i/i$ and let $T_j = Y_1 + \cdots + Y_j$, so that

$$(7.3) \qquad \frac{X_1 + \cdots + X_n}{n} = -\frac{1}{n} \sum_{j=1}^{n} T_j + \frac{n+1}{n} T_n.$$

By Theorem 7.4 the sequence T_n S-converges n.e. Let λ be positive and unlimited. By Kolmogorov's inequality

$$\Pr\left\{ \max_{j \leqslant \nu} |T_j| \geqslant \lambda \right\} \leqslant \frac{\sum_{i=1}^{\nu} \sigma_i^2/i^2}{\lambda^2} \simeq 0,$$

and from this it follows by a now-familiar argument that nearly everywhere the T_j are all limited. But if a sequence of limited real numbers S-converges to a real number t, it is easy to see that the sequence of averages also S-converges to t. Therefore the right-hand side of (7.3) S-converges to 0 n.e.   Q.E.D.

Notice that if $X_1, \ldots, X_\nu$ satisfy the hypotheses of Theorem 7.1 they also satisfy the hypotheses of Theorem 7.5 (after subtracting the mean), so that actually the strong law of large numbers holds under the hypotheses of Theorem 7.1. We can, in fact, weaken the hypotheses, as follows.

If X is a radom variable on a finite probability space $\langle \Omega, \mathrm{pr} \rangle$ and a is a positive real number, we define the truncated random variable $X^{(a)}$ by

$$X^{(a)}(\omega) = \begin{cases} X(\omega), & |X(\omega)| \leqslant a, \\ 0, & |X(\omega)| > a, \end{cases}$$

and we let $\mathbf{L}^1$ be the external set of all random variables X on $\langle \Omega, \mathrm{pr} \rangle$ such that $E|X|$ is limited and $E|X - X^{(a)}| \simeq 0$ for all unlimited $a > 0$. In terms of the probability distribution f_X, this is the same as requiring that the sequence $\sum_{|\lambda| \leqslant n} |\lambda| f_X(\lambda)$ S-converge.

THEOREM 7.6. *Let* $X_1, \ldots, X_\nu$ *be independent random variables on a finite probability space which are* $\in \mathbf{L}^1$ *and have a common probability distribution* f. *Then the strong law of large numbers holds.*

PROOF. Let $\mu_i = EX_i^{(i)}$ and write X_i as

$$(7.4) \qquad X_i = \mu_i + (X_i - X_i^{(i)}) + (X_i^{(i)} - \mu_i).$$

Then $\mu_i = \sum_{|\lambda| \leqslant i} \lambda f(\lambda)$, so that the μ_i S-converge to $\mu = \sum \lambda f(\lambda)$, and since they are limited (being bounded in absolute value by $\sum |\lambda| f(\lambda)$), their averages also S-converge to μ. We need only show that for each of the remaining two terms on the right-hand side of (7.4) the averages S-converge to 0.

For any $n \leqslant \nu$ we have

$$\Pr\{X_i - X_i^{(i)} \neq 0 \text{ for some } i \text{ with } n \leqslant i \leqslant \nu\} \leqslant \sum_{i=n}^{\nu} \Pr\{|X_i| > i\}$$

$$= \sum_{i=n}^{\nu} \sum_{|\lambda| > i} f(\lambda) \leqslant \sum_{|\lambda| > n} |\lambda| f(\lambda).$$

Let $\varepsilon > 0$ be standard. Then there is a limited n such that this is $\leqslant \varepsilon$. It is easy to see that for n limited $X_i - X_i^{(i)}$ are limited for all $i = 1, \ldots, n$ nearly everywhere. Therefore the averages of the $X_i - X_i^{(i)}$ S-converge to 0 n.e.

By Theorem 7.5 we need only show that $\sum E(X_i^{(i)} - \mu_i)^2/i^2$ S-converges. But

$$\sum_{i=n}^{\nu} \frac{1}{i^2} E(X_i^{(i)} - \mu_i)^2 \leqslant \sum_{i=n}^{\nu} \frac{1}{i^2} E(X_i^{(i)})^2$$

$$= \sum_{i=n}^{\nu} \frac{1}{i^2} \sum_{|\lambda| \leqslant i} \lambda^2 f(\lambda) \leqslant \sum_{\lambda} \lambda^2 f(\lambda) \sum_{\substack{i \geqslant |\lambda| \\ i \geqslant n}} \frac{1}{i^2}$$

$$= \sum_{|\lambda| < n^{1/2}} \lambda^2 f(\lambda) \sum_{i \geqslant n} \frac{1}{i^2} + \sum_{|\lambda| > n^{1/2}} \lambda^2 f(\lambda) \sum_{i \geqslant |\lambda|} \frac{1}{i^2}$$

$$\leqslant \sum_{\lambda} n^{1/2} |\lambda| f(\lambda) \frac{\pi^2}{6} \frac{1}{n} + \sum_{|\lambda| > n^{1/2}} |\lambda| f(\lambda) \frac{\pi^2}{6}.$$

This is limited for $n = 1$ and infinitesimal for n unlimited, so that the series S-converges.   Q.E.D.

Once I went with a colleague to give an oral examination in probability theory to a graduate student in engineering. My colleague's first question was, "What is a random variable?" to which the student replied: "Given a set Ω, given a family $\mathfrak{S}$ of subsets of Ω containing Ω itself and closed under the formation of countable unions and complementation, and given a function μ from $\mathfrak{S}$ to the positive reals such that $\mu(\Omega) = 1$ and whenever E in $\mathfrak{S}$ is the disjoint union of a sequence of sets E_n in $\mathfrak{S}$ then $\mu(E)$ is the sum of the series $\sum \mu(E_n)$, then a random variable is a function from Ω to the real numbers such that the inverse image of any interval is an element of $\mathfrak{S}$". The approach to probability theory which we initiated above is simpler, and the scientific content of the theorems we have proved is the same as that of their conventional forms. One can do a vast amount of probability theory within this framework, and one can if one wishes deduce the conventional forms of the theorems from their analogues on finite probability spaces, but we will not pursue the topic further here.

8. Appendix: The conservation theorem. This appendix is devoted to showing that IST is a conservative extension of ZFC; that is, every internal statement which can be proved in IST can be proved in ZFC.

We recall that a filter $\mathfrak{U}$ on a set I is a family of subsets of I not containing $\varnothing$, containing I, closed under the formation of finite intersections, and containing every superset of every set in it. A filter $\mathfrak{U}$ is called an *ultrafilter* in case for all subsets E of I either $E \in \mathfrak{U}$ or $I \backslash E \in \mathfrak{U}$.

Let V be a set and let $\mathfrak{U}$ be an ultrafilter on a set I. Two functions, $i \mapsto x_i$ and $i \mapsto y_i$, in V^I are called *equivalent modulo* $\mathfrak{U}$ in case $\{i \in I: x_i = y_i\} \in \mathfrak{U}$. It is easily seen that this is an equivalence relation. We define the *ultrapower* $^*V = V^I/\mathfrak{U}$ to be the set of all equivalence classes modulo $\mathfrak{U}$ of functions from I to V. The constant functions give a natural injection of V into V^I and this induces a natural injection of V into $^*V$. If $E \subset V$ and $i \mapsto x_i$, $i \mapsto y_i$ are equivalent modulo $\mathfrak{U}$ then $\{i \in I: x_i \in E\} \in \mathfrak{U}$ if and only

if $\{i \in I: y_i \in E\} \in \mathfrak{U}$. We define $*E$ to be the set of all equivalence classes modulo $\mathfrak{U}$ of functions $i \mapsto x_i$ such that $\{i \in I: x_i \in E\} \in \mathfrak{U}$. We call $*E$ the *extension* of E.

THEOREM 8.1. *Let $*V$ be an ultrapower of V. Then the mapping $E \mapsto *E$ of $P(V) \to P(*V)$ is an isomorphism of the Boolean algebra $P(V)$ onto a Boolean subalgebra of $P(*V)$.*

PROOF. Clearly the empty set goes to the empty set and V goes to $*V$. We have that $\{i \in I: x_i \in E \cap F\} \in \mathfrak{U}$ if and only if $\{i \in I: x_i \in E\} \in \mathfrak{U}$ and $\{i \in I: x_i \in F\} \in \mathfrak{U}$, since $\mathfrak{U}$ is a filter, so that $*(E \cap F) = *E \cap *F$. Let ^{c} denote complementation in V or $*V$. Since $\mathfrak{U}$ is an ultrafilter, $\{i \in I: x_i \in E^c\}$ is in $\mathfrak{U}$ if and only if its complement $\{i \in I: x_i \in E\}$ is not in $\mathfrak{U}$, so that $*(E^c) = (*E)^c$. Q.E.D.

If n is a natural number we may identify $(V^n)^I$ with $(V^I)^n$, and if $i \mapsto x_i^1, \ldots, i \mapsto x_i^n$ are equivalent modulo $\mathfrak{U}$ to $i \mapsto y_i^1, \ldots, i \mapsto y_i^n$, respectively, then $i \mapsto \langle x_i^1, \ldots, x_i^n \rangle$ is equivalent modulo $\mathfrak{U}$ to $i \mapsto \langle y_i^1, \ldots, y_i^n \rangle$, and conversely. Therefore we may identify $*(V^n)$ with $(*V)^n$. We define π_j: $P(V^n) \to P(V^{n-1})$ by

$$\pi_j(E) = \{\langle x^1, \ldots, x^{j-1}, x^{j+1}, \ldots, x^n \rangle \in V^{n-1}: \text{for some}$$

$$x^j \text{ in } V \text{ we have } \langle x^1, \ldots, x^{j-1}, x^j, x^{j+1}, \ldots, x^n \rangle \in E\}$$

and we define π_j: $P(*V^n) \to P(*V^{n-1})$ similarly.

THEOREM 8.2. *Let $*V$ be an ultrapower of V. If $E \subset V^n$ then $*\pi_j(E) = \pi_j(*E)$.*

PROOF. We write x for x^j, y for $\langle x^1, \ldots, x^{j-1}, x^{j+1}, \ldots, x^n \rangle$, and (x, y) for $\langle x^1, \ldots, x^{j-1}, x^j, x^{j+1}, \ldots, x^n \rangle$. Let $E \subset V^n$. Then $*\pi_j(E)$ is the set of all equivalence classes modulo $\mathfrak{U}$ of functions $i \mapsto y_i$ such that $\{i \in I: \text{for some } x_i \text{ in } V \text{ we have } (x_i, y_i) \in E\}$ is an element of $\mathfrak{U}$. But this is equal to the set of all equivalence classes modulo $\mathfrak{U}$ of functions $i \mapsto y_i$ such that there is a function $i \mapsto x_i$ such that $\{i: (x_i, y_i) \in E\}$ is an element of $\mathfrak{U}$, and this is $\pi_j(*E)$. Q.E.D.

If $E \subset V^n$ and $x \in V$ we define $E_{x,j}$ by

$$E_{x,j} = \{\langle x^1, \ldots, x^{j-1}, x^{j+1}, \ldots, x^n \rangle$$

$$\in V^{n-1}: \langle x^1, \ldots, x^{j-1}, x, x^{j+1}, \ldots, x^n \rangle \in E\},$$

and similarly for a subset of $*V^n$ and a point in $*V$.

THEOREM 8.3. *Let $*V$ be an ultrapower of V, and identify V as a subset of $*V$ by means of the natural injection. If $E \subset V^n$ and $x \in V$ then $*(E_{x,j}) = (*E)_{x,j}$.*

PROOF. This is obvious. Q.E.D.

An n-ary relation on V is a subset of V^n. Logical connectives among relations, such as negation, implication, etc., are expressed by means of the Boolean operations. Existential quantifiers are expressed by means of the projection operators π_j, and universal quantifiers are expressed by a combina-

tion of them and complementation. Therefore Theorems 1 and 2 have the consequence that any elementary statement about V (any statement involving a fixed finite set of finitary relations and involving quantification only over the elements of V) is true if and only if the corresponding statement about $^*V$, involving the extensions of the given relations, is true. By Theorem 3 the same holds for formulas involving free variables, for all values of the free variables in V. This is the transfer principle for ultrapowers.

Let V be a set. We denote the set of all ultrafilters on V by $\hat{V}$. Let $^*V$ be an ultrapower of V. For ξ in $^*V$ let $t(\xi)$ be the family of all subsets E of V such that $\xi \in {}^*E$. By Theorem 8.1, $t(\xi)$ is an ultrafilter on V. Thus we have a natural mapping $t: {}^*V \to \hat{V}$. We call $^*V$ an *adequate* ultrapower of V in case the mapping t is surjective.

THEOREM 8.4. *Let V be a set. Then there exists an adequate ultrapower of V.*

PROOF. We use P_{fin} to denote the set of all finite subsets of a given set. Let $I = P_{\text{fin}}(P(V))$. For $E \subset V$ let $\tilde{E} = \{i \in I: E \in i\}$. The intersection $\tilde{E}_1 \cap \cdots \cap \tilde{E}_n$ is always nonempty, since $i = \{E_1, \ldots, E_n\}$ is an element of it, so that there is a filter on I containing all sets of the form $\tilde{E}$.

We claim that if $\mathcal{F}$ is any filter on a set I then there is an ultrafilter $\mathfrak{U}$ containing $\mathcal{F}$. If $\mathcal{F}$ is not an ultrafilter, let G be a subset of I such that neither G nor $I \setminus G$ is in $\mathcal{F}$. It is easy to verify that the family $\mathcal{G}$ of all supersets of sets of the form $F \cap G$ with F in $\mathcal{F}$ is a filter properly containing $\mathcal{F}$. Also, the union of an increasing chain of filters is again a filter. By Zorn's lemma, therefore, $\mathcal{F}$ is contained in an ultrafilter, which establishes the claim.

Let $\mathfrak{U}$ be an ultrafilter on I containing all sets of the form $\tilde{E}$, and let $^*V = V^I/\mathfrak{U}$. Let $\mathcal{V}$ be an ultrafilter on V. For each i in I let A_i be the intersection of all elements of i which are in $\mathcal{V}$. Then $A_i \neq \varnothing$. For each i in I choose an element x_i of A_i, and let ξ be the equivalence class modulo $\mathfrak{U}$ of the function $i \mapsto x_i$. We will show that $t(\xi) = \mathcal{V}$. Let $E \in \mathcal{V}$. If i is such that $E \in i$ then $A_i \subset E$ so that $x_i \in E$. That is, $\tilde{E} \subset \{i \in I: x_n \in E\}$ and since $\tilde{E}$ is in $\mathfrak{U}$ its superset $\{i \in I: x_i \in E\}$ is in $\mathfrak{U}$. That is, $\xi \in {}^*E$ and by the definition of t this means that $E \in t(\xi)$. Since E is an arbitrary element of $\mathcal{V}$ we have $\mathcal{V} \subset t(\xi)$, since $\mathcal{V}$ and $t(\xi)$ are both ultrafilters we have $\mathcal{V} = t(\xi)$, and since $\mathcal{V}$ is an arbitrary ultrafilter on V we have that t is surjective.   Q.E.D.

THEOREM 8.5. *Let $^*V$ be an adequate ultrapower of V and let $R \subset V^2$ be such that for all finite subsets F of V there is an x in $^*V$ with $\langle x,y \rangle \in {}^*R$ for all y in F. Then there is a ξ in $^*V$ with $\langle \xi,y \rangle \in {}^*R$ for all y in V.*

PROOF. Let $F = \{y_1, \ldots, y_n\}$ be a finite subset of V. By Theorems 8.1 and 8.3,

$$\{x \in {}^*V: \langle x,y_1 \rangle \in {}^*R \wedge \cdots \wedge \langle x,y_n \rangle \in {}^*R\}$$
$$= {}^*\{x \in V: \langle x,y_1 \rangle \in R \wedge \cdots \wedge \langle x,y_n \rangle \in R\}.$$

By hypothesis, sets of this form are nonempty, so there is a filter on V containing all sets of the form $\{x \in V: \langle x,y \rangle \in R\}$, where $y \in V$, and therefore, as shown in the proof of Theorem 8.4, there is an ultrafilter $\mathcal{V}$ on V containing all sets of this form. Since $^*V$ is adequate there is a ξ in $^*V$ such

that $t(\xi) = \mathcal{V}$. This means that $\xi \in *\{X \in V: \langle x, y \rangle \in R\}$ or in other words that $\langle \xi, y \rangle \in *R$, for all y in V.   Q.E.D.

Theorem 8.5 is the restricted idealization principle for adequate ultrapowers, restricted because there are no free variables ranging over $*V$. To obtain the unrestricted idealization principle we turn to the notion of adequate ultralimits.

Let $V = {}^0V$ be a set, let ${}^1V$ be an ultrapower of ${}^0V$, let ${}^2V$ be an ultrapower of ${}^1V$, and so on by induction for every natural number n. Let $j_n: {}^nV \to {}^{n+1}V$ be the natural injection and let $*V$ be the direct limit of the sequence

$$ {}^0V \xrightarrow{j_0} {}^1V \xrightarrow{j_1} {}^2V \xrightarrow{j_2} \cdots . $$

If we identify each ${}^nV$ as a subset of ${}^{n+1}V$ via the injection j_n then $*V$ is simply the union of the ${}^nV$. We call $*V$ an *ultralimit* of V, and if for each n the ultrapower ${}^{n+1}V$ is an adequate ultrapower of ${}^nV$ then we say that $*V$ is an *adequate ultralimit* of V. By Theorem 8.4 every set V has an adequate ultralimit. For $E \subset V$, let ${}^1E$ be the extension of E in ${}^1V$, let ${}^2E$ be the extension of ${}^1E$ in ${}^2V$, and so on, and let $*E$ be the union of the ${}^nE$. We call $*E$ the *extension* of E. We may identify the ultralimit of a Cartesian product with the Cartesian product of the ultralimits, and Theorems 8.1, 8.2, and 8.3 carry over immediately to ultralimits: the mapping $E \mapsto *E$ is an isomorphism of the Boolean algebra $P(V)$ onto a Boolean algebra of $P(*V)$ which commutes with projections and sections in Cartesian products (the transfer principle for ultralimits).

THEOREM 8.6. *Let $*V$ be an adequate ultralimit of V and let $R \subset V^{2+k}$. For all $t_1, \ldots, t_k$ in $*V$, if for all finite subsets F of V there is an x in $*V$ with $\langle x, y, t_1, \ldots, t_k \rangle \in *R$ for all y in F then there is a ξ in $*V$ with $\langle \xi, y, t_1, \ldots, t_k \rangle \in *R$ for all y in V.*

PROOF. Since $t_1, \ldots, t_k$ are in $*V$ they are in ${}^nV$ for some n. Let $S \subset {}^nV \times {}^nV$ be defined by

$$ S = \{ \langle x, y \rangle \in {}^nV \times {}^nV: y \in V \Rightarrow \langle x, y, t_1, \ldots, t_k \rangle \in {}^nR \}. $$

Let $G = \{ y_1, \ldots, y_m \}$ be a finite subset of ${}^nV$, and regard $*V$ as an ultralimit of ${}^nV$. By Theorems 8.1 and 8.3 for ultralimits,

$$ \{ x \in *V: \langle x, y_1 \rangle \in *S \wedge \cdots \wedge \langle x, y_m \rangle \in *S \} $$
$$ =* \{ x \in {}^nV: \langle x, y_1 \rangle \in S \wedge \cdots \wedge \langle x, y_m \rangle \in S \}. $$

By hypothesis, sets of this form are nonempty (the condition $\langle x, y_j \rangle \in *S$ is trivially satisfied if $y_j \notin V$) so there is an x in ${}^nV$, and consequently in ${}^{n+1}V$, such that $\langle s, y_1 \rangle \in S \wedge \cdots \wedge \langle x, y_m \rangle \in S$. By Theorem 8.5 for the adequate ultrapower ${}^{n+1}V$ of ${}^nV$, there is a ξ in ${}^{n+1}V$, and consequently in $*V$, such that $\langle \xi, y \rangle \in *S$, and consequently $\langle \xi, y, t_1, \ldots, t_k \rangle \in *R$, for all y in V.   Q.E.D.

Theorem 8.6 is the idealization principle for adequate ultralimits.

If only we could take an adequate ultralimit of the universe we would be through. Instead, we let $R(0) = \varnothing$ and by transfinite induction for all ordinals α we let

$$R(\alpha) = \bigcup_{\beta < \alpha} P(R(\beta)).$$

Every set x is in $R(\alpha)$ for some ordinal α. Let A be a statement of ZFC. The relativization of A to V is the formula A^V with free variable V obtained by replacing $\forall x$ by $\forall x(x \in V) \Rightarrow$ and $\exists x$ by $\exists x(x \in V) \wedge$, for every variable x occurring in A. In the following theorem, $A_1, \ldots, A_n$ are a finite number of statements of ZFC.

THEOREM 8.7. *There is an ordinal α such that*

$$(8.1) \qquad\qquad (A_1 \Leftrightarrow A_1^{R(\alpha)}) \wedge \cdots \wedge (A_n \Leftrightarrow A_n^{R(\alpha)}).$$

PROOF. Since $(\sim A)^V \Leftrightarrow \sim (A^V)$ it is immaterial, for each $i = 1, \ldots, n$, whether we have A_i or $\sim A_i$. Since one of the terms in the expansion of $(A_1 \vee \sim A_1) \wedge \cdots \wedge (A_n \vee \sim A_n)$ holds, we may, without loss of generality, assume $A_1 \wedge \cdots \wedge A_n$ in the proof. We may write A_i as

$$(8.2) \qquad\qquad \forall x_i^1 \exists y_i^1 \cdots \forall x_i^k \exists y_i^k \, B_i(x_i^1, y_i^1, \ldots, x_i^k, y_i^k)$$

for $i = 1, \ldots, n$, where B_i has no quantifiers. For all V there are functions $\tilde{y}_i^j$ on V^j such that

$$(8.3) \quad \forall x_i^1 \in V \cdots \forall x_i^k \in V \, B_i(x_i^1, \tilde{y}_i^1(x^1), \ldots, x_i^k, \tilde{y}_i^k(x_i^1, \ldots, x_i^k)).$$

Let $r(V)$ be the least ordinal β such that V and the ranges of the $\tilde{y}_i^j$ are all elements of $R(\beta)$. Let $\alpha_1 = r(\varnothing)$, let $\alpha_2 = r(R(\alpha_1))$, and by induction $\alpha_{m+1} = r(R(\alpha_m))$. Then $m \mapsto \alpha_m$ is a strictly increasing sequence of ordinals; let α be its limit. Then (8.1) holds, because if the x_i^j are all in $R(\alpha)$ they are in $R(\alpha_m)$ for some m, and so there are solutions to (8.3), and hence to (8.2), with the y_i^j in $R(\alpha_{m+1})$ and, consequently, in $R(\alpha)$.   Q.E.D.

THEOREM 8.8. *Every internal theorem of IST is a theorem of* ZFC.

PROOF. Let A be an internal theorem of IST. The axioms of IST are the axioms of ZFC and the axiom schemes (I), (S), and (T). Only finitely many of the axioms of ZFC, say $A_1, \ldots, A_n$, can be used in the proof of A within IST. By Theorem 8.7 there is an ordinal α such that

$$(8.4) \qquad\qquad (A \Leftrightarrow A^{R(\alpha)}) \wedge A_1^{R(\alpha)} \wedge \cdots \wedge A_n^{R(\alpha)}.$$

Let $*R(\alpha)$ be an adequate ultralimit of $R(\alpha)$, let $j: R(\alpha) \to *R(\alpha)$ be the natural injection, and let $E(\alpha) = \{\langle x, y \rangle \in R(\alpha) \times R(\alpha): x \in y\}$. We claim that $*R(\alpha)$ with the interpretations

$$(8.5) \qquad\qquad \langle x, y \rangle \in *E(\alpha) \quad \text{for } x \in y,$$

$$(8.6) \qquad\qquad x \in jR(\alpha) \quad \text{for } x \text{ standard}$$

is a model for the axiom system $A_1, \ldots, A_n$, (I), (S), (T). If B is a formula of IST let $*B$ be the formula of ZFC obtained by replacing each occurrence of $x \in y$ by its interpretation (8.5) and each occurrence of x standard by its interpretation (8.6), for all variables x and y. By the transfer principle and

principle of idealization for adequate ultralimits, we have*(T) and*(I). Using the fact that any subset of an element of $R(\alpha)$ is an element of $R(\alpha)$, we have*(S). From $A_1^{R(\alpha)}, \ldots, A_n^{R(\alpha)}$ we have $*A_1, \ldots, *A_n$. This establishes the claim.

Now the proof of A from $A_1, \ldots, A_n$, (I), (S), (T) gives a proof of $*A$ from $*A_1, \ldots, *A_n, *(I), *(S), *(T)$. Since A is internal, $A^{R(\alpha)}$ follows from $*A$ and so by (8.4) we have A. This gives a proof of A within ZFC. Q.E.D.

9. Notes. The axioms of internal set theory are simply the basic properties of the internal sets in the usual approach to nonstandard analysis (provided sufficient saturation has been assumed). In our approach we refer to the internal sets simply as sets. Once one recovers from the shock of being told that infinitesimals and other idealized elements were there all along in the sets with which we are familiar, and learns to avoid illegal set formation, one will find our approach very easy to use. All familiar results are available without change; they do not have to be extended to a new system. External sets are available for convenience of language. It is surprising how seldom one needs to quantify external sets: perhaps this is what distinguishes model theory from nonstandard analysis.

In Robinson's book [10] only the restricted idealization principle (no internal free variables) is available. This is because he uses enlargements, which are given by adequate ultrapowers rather then adequate ultralimits. The use of the unrestricted idealization principle produces a natural closure, and allows the reduction algorithm to work. Luxemburg [7] gives examples of results which may be false for enlargements.

The distinction between $\in$ and $\mathbf{\in}$ in §3 is necessary. We cannot treat external sets on the same footing as internal sets when we are discussing all of set theory. As customarily defined in set theory, a natural number is equal to the set of all its predecessors. Let n be an unlimited natural number in a model of IST. Then $n, n-1, n-2, n-3, \ldots$ is an external infinite sequence with the property that each term is $\in$ its predecessor. The axiom of regularity [1] precludes the existence of such a sequence where each term is $\mathbf{\in}$ its predecessor, and in fact each term is $\mathbf{\in}$ the * of its predecessor. Within IST, of course, the sequence is finite and terminates after n steps. For a fragment of set theory it is possible to identify $\in$ and $\mathbf{\in}$, but only at the cost of distinguishing types of sets and considering stratified formulas [10, pp. 19–30].

The theorem on coloring graphs in Example 4 of §4 is due to de Bruijn [2, p. 137], and the nonstandard proof we give is essentially Beth's model-theoretic proof [8, p. 96]. The theorem on internal monads of Example 13 of §4 is due to Luxemburg [7] who gives a lengthy proof which is also valid for enlargements. The theorem of Example 14, and its consequence in Example 15, are also due to Luxemburg [7].

The term "strongly differentiable" is used by Nijenhuis [9] for what we have called "derivable" in §5. We have avoided the use of "strongly differentiable" because this term has a different meaning in the context of Banach space valued functions.

We have followed Lamperti [5] closely in §7.

Theorem 8.7 is a reflection principle. Our proof is taken from an argument of A. Lévy [6] who credits the method to A. Montague. Our proof uses implicitly the theorem that for all V, if $\forall x \in V \exists y\, A(x,y)$ then there exists a function $\tilde{y}$ on V such that $A(x, \tilde{y}(x))$ for all x in V. If the y is unique this is an instance of the replacement axiom scheme. Let $B(x, \beta)$ be the formula: $\exists y\, A(x,y)$ and $y \in R(\beta)$ and β is the least ordinal with this property. Then the β is unique, so there is a function $\tilde{\beta}$ on V. Let γ be such that the range of $\tilde{\beta}$ on V is in $R(\gamma)$. Then $\forall x \in V \exists y\, A(x,y)$ implies $\forall x \in V \exists y \in R(\gamma)\, A(x,y)$, and now we may apply the axiom of choice to obtain a function $\tilde{y}$.

Many of the theorems and examples in this paper occur in Robinson's book [10]. The magnitude of Abraham Robinson's achievement is astonishing. In the last paragraph of [10] Robinson says "... 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". This is what we have tried to express by means of internal set theory. We share his hope [10, p. 5] "... that some branches of modern Theoretical Physics, in particular those that are afflicted with divergence problems, might be treated with profit by Nonstandard Analysis".

I thank Carl Herz for a simpler proof of Theorem 7.6, and William C. Powell for permission to include his result, Theorem 8.8. I thank Simon Kochen, the originator of many ideas in this field, for many long conversations which, in addition to being fruitful, were great fun.

REFERENCES

We wish to draw the reader's attention to a Bibliography of Nonstandard Analysis, Lecture Notes in Mathematics, no. 3, compiled by D. Randolph Johnson and complete through June 1975, which is available at one dollar per copy by writing to the Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

1. Paul J. Cohen, *Set theory and the continuum hypothesis*, Benjamin, New York, 1966. MR **38** #999.

2. P. Erdös, *Some remarks on set theory*, Proc. Amer. Math. Soc. **1** (1950), 127–141.

3. Charles N. Friedman, *Perturbations of the Schrödinger equation by potentials with small support, semigroup product formulas, and applications to quantum mechanics*, Ph. D. Dissertation, Princeton Univ., Princeton, N.J., 1971.

4. ———, *Perturbations of the Schrödinger equation by potentials with small support*, J. Functional Analysis **10** (1972), 346–360. MR **49** #5529.

5. John Lamperti, *Probability*, Benjamin, New York, 1966. MR **34** #6812.

6. Azriel Lévy, *Axiom schemata of strong infinity in axiomatic set theory*, Pacific J. Math. **10** (1960), 223–238. MR **23** #A1522.

7. W. A. J. Luxemburg, *A general theory of monads*, Applications of Model Theory to Algebra, Analysis, and Probability (Internat. Sympos., Pasadena, Calif., 1967), Holt, Rinehart and Winston, New York, 1969, pp. 18–86. MR **39** #6244.

8. E. Mendelson, *Introduction to mathematical logic*, Van Nostrand, Princeton, N. J., 1964. MR **29** #2158.

9. A. Nijenhuis, *Strong derivatives and inverse mappings*, Amer. Math. Monthly **81** (1974), 969–980. MR **50** #13405.

10. Abraham Robinson, *Non-standard analysis*, 2nd ed., American Elsevier, New York, 1974. (1st ed., 1966. MR **34** #5680.)

DEPARTMENT OF MATHEMATICS, PRINCETON UNIVERSITY, PRINCETON, NEW JERSEY 08540