

On Power Set Axiom

Leonid A. Levin* (<https://www.cs.bu.edu/fac/lnd>)

Abstract

Usual math sets have special types: countable, compact, open, occasionally Borel, rarely projective, etc. Generic sets dependent on Power Set Axiom appear mostly in esoteric areas, ST logic, etc. Dropping that Axiom may greatly simplify the foundations of mainstream math. Meanwhile dependence on it of a theorem is worth noting, as dependence on Choice often is.

I always wondered why math foundations as taken by logicians are so distant from the actual math practice. For instance, the cardinality theory - the heart of the Set Theory (ST) - is almost never used beyond figuring out which sets are countable and which are not.

I see the culprit in the Power Set Axiom (somewhat similar to “set of all sets”, albeit not so fatal). Usual math objects are sets of special types: countable, compact, open, occasionally Borel, rarely projective, etc. They by nature are not really sets but classes, defined by ST formulas with countable sets as parameters. Collections of them are really collections of those parameters.

ST power can ever be expanded with more axioms. There is no natural end to this endeavor. A better goal is to find the real match to the actual math needs. Studying “higher level” theories is exciting but can be done without claiming relation to the foundations of mainstream math.

1 A Smaller Purge: Against Pathologies

My first retreat, replacing the Power Set with its opposite “all sets are countable”, is not much unlike others studied so far (e.g., Kripke-Platek ST, Pocket ST [2], etc.).

Yet, some aspects differ, to conform to math practice. Quantifiers will bind all sets (no restriction to Σ_0 for some uses) but not all classes. Classes are meta-objects, not in the domain of the variables. Quantifiers can bind their parameters, in effect, binding classes of any one particular type.

Math would then use classes F_p specified by a set p (parameter) and (often implicit) type: an ST formula $F(q, p)$ giving the formal meaning to colloquial $F_q \in F_p$. To adhere to Foundation axioms, $F(q, p)$ includes a ranking r of parameters with ordinals¹ and a clause for $r(q) < r(p)$. Classes (only informally called “sets”) specified by different types can be the same if the membership relations on their transitive closures have a formula-defined isomorphism.

The restrictions seem to have little effect on actual math practice, besides making some wordings colloquial shorthands for their formal meaning. E.g., “ x is in open $P \subset \mathbb{R}$ ” can be a shorthand for “ $F(x, p)$, where p specifies $P \stackrel{\text{def}}{=} F_p$ as the set of its rational intervals”. (Or P can be Borel, or whatever type clear from the context.) And yet, such version of ST is just a form of second order arithmetic: Hereditarily countable sets reduce to sequences listing \in relation on their transitive closures.

*Boston University, College of Arts and Sciences, Computer Science department, Boston, MA 02215.

¹Foundation axioms preclude infinite descending chains in v.Neumann ordinals (transitive sets totally ordered by \in).

Ordinals can also be defined as sets of rationals isomorphic to v.Neumann ordinals, so forming a class of rank $\omega+1$.

2 A More Radical Purge: Against Tricksome Objects

And making explicit the internal to math (defined by formulas) and external (parameters based) aspects of math objects clarifies their nature. Even for countable collections this allows a better focus on issues brought by these two different (albeit with many similarities) sources.

“Pure” classes defined by parameter-free ST formulas are specific but tricksome. Careless treatment of them easily brings paradoxes, but careful treatment complicates the theory. And they are not to be combined in further definable collections as that would only extend the language of allowed formulas. A formula can take parameters for external data which math handles but does not specify internally. Their values are sets treated as the domain of ST variables.

Sets can be put in collections based on properties and relations with other sets. This forms “mixed” classes, specified by ST formulas with some free variables taken for data parameters. Data may be chaotic, but simpleminded, playing no Goedel tricks. General math objects – mixed classes – carry both tricks and chaos. But their tricks are limited to what a single formula can carry.

In reality, external data do end-up finite. Yet handling their (often ambiguous) termination points is awkward. So they are taken as (potentially) countably infinite.² But distinguishing such data (sets) from pure classes allows a further radical insight: Even having infinite complexity, external objects α can have only finite (actually very small) Kolmogorov Information $\mathbf{I}(\alpha : \mu) < \infty$ (dependence) on pure sequences μ . (Sec. 3 gives some background.) Note, recursive and random transformations can only add $O(1)$ information about a given (e.g., by a math formula) sequence.

See [4] for discussion in its section “The Independence Postulate” and for references.

Due to this Postulate (IP), only computable sequences can double as both pure classes and sets.

IP conflicts with Specification Axioms for sets. (For classes these axioms are merely definitions.) So they must be restricted to just one: “Being a set is preserved by Turing reductions”. Instead, the Foundation axiom becomes a family, like Induction Axioms: one for each class of ordinals.

And “Primal Chaos” Axiom is added: “Each $\alpha \in \{0, 1\}^{\mathbb{N}}$ is computable from even-indexed digits of some Martin-Lof random β : $\mathbf{T}(\beta) < \infty$ ”. (For classes, i.e. in ZFC, it is a theorem: see [5].)³

Then comes a great simplification, reducing any statement to one with only integer quantifiers. IP excludes α satisfying $P(\alpha, \beta)$ unless computing them from randomly chosen $\gamma \in \{0, 1\}^{\mathbb{N}}$ has a positive chance conditioned on β computed from that γ , too. The converse comes from “Primal Chaos” Axiom. (That axiom may need a somewhat stronger form: I will clarify in the final version.)

Of course, treating math objects as combining pure math classes with sets independent of them (by Independence Postulate) would require thought about terminology to minimize its effects on regular math practice. But I do not see why well designed conventions would require a very different care level than proper use of ZFC now does.

²An example of more technical conveniences: the ring (class) of p-adic integers, unlike \mathbb{Z} , is compact.

³“Computable from” here can be equivalently restricted to “weak truth table reducible to” (see sec. 3).

And to build a model (say, in ZFC) take a $\gamma \in \{0, 1\}^{\mathbb{N}}$ that is not in any arithmetic set of measure 0. The model includes all α for which $\exists k \in \mathbb{N}$ such that α is computable from digits of γ with indexes not divisible by k . ([3] used a different version of “Primal Chaos” including – postulating existence of – such “universal” γ as a set.)

3 Kolmogorov Information

Computably enumerable (c.e.) functions to $\overline{\mathbb{R}^+}$ are supremums of c.e. sets of basic continuous ones.

Dominant in a convex class C of functions is its c.e. $f \in C$ if all c.e. g in C are $O(f)$.

Such is $\sum_i g_i/(i^2+i)$ if (g_i) is a c.e. family of all c.e. functions in weakly compact C .

Uniform measure on $\Omega \stackrel{\text{df}}{=} \{0,1\}^{\mathbb{N}}$ is $\lambda(x\Omega) \stackrel{\text{df}}{=} 2^{-n}$ for $x \in \{0,1\}^n$. (Same for $\Omega^2 \simeq \Omega$).

λ -test is $\|\mathbf{T}(\alpha)\|$ for a dominant \mathbf{T} on Ω with expectation $\lambda(\mathbf{T}) \leq 1$. $\|t\| \stackrel{\text{df}}{=} \lceil \log_2 t \rceil - 1$.

$R_\lambda \stackrel{\text{df}}{=} \{\alpha : \mathbf{T}(\alpha) < \infty\}$ is the class of Martin-Lof λ -random α . Let $x^f \stackrel{\text{df}}{=} f^{-1}(x\Omega)$ for algorithms $f : \Omega \rightarrow \Omega$, and $U(p\alpha) = f_p(\alpha)$ be a universal algorithm. R_λ consists of all α with $\sup_x \frac{\lambda(x^U)}{\lambda(x\Omega)} < \infty$. (A similar property holds for μ -random α with other measures μ .) Note that any arithmetic property invariant under changing single digits holds either for all $\beta \in R_\lambda$ adhering to IP or for none.

Mutual Information $\mathbf{I}(\alpha_1 : \alpha_2)$ is $\min_{\beta_1, \beta_2} \{\|\mathbf{T}(\beta_1, \beta_2)\| : U(\beta_i) = \alpha_i\}$.

Remark 1. If $\lambda(A) = 0$ then a sequence s of clopen $s_n \subset \Omega$, $\lambda(s_n) < 2^{-n}$ exists s.t. each $\alpha \in A$ is in infinitely many of them. So, if $\Omega \setminus A$ is Turing-reduction-closed (or $\mathbf{T}(\alpha) < \infty$) then $\mathbf{I}(\alpha : s) = \infty$.

W.t.t. reductions. Let u be the restriction of U to have i -th bits of output depend only on the first, say, $3i$ input bits. The outputs distribution is not affected: $\lambda(x^u) = \Theta(\lambda(x^U))$, see e.g., [5].

Let $\lambda_\beta(Q) = \lim_{x: \beta \in x\Omega} \frac{\lambda(Q \cap x^u)}{\lambda(x^u)}$ be the distribution λ on γ conditioned on $u(\gamma) = \beta$.

IP and $\exists \alpha \in R_\lambda U(\alpha) = \beta$ imply $\exists \alpha \in R_\lambda u(\alpha) = \beta$: Let $U(\gamma)$ take $T_i(\gamma)$ time to output 2^i bits. $U_*(\gamma)$ avoids divergence by diluting β with t_i blanks $\#$ between 2^i bits segments. Here t_i is the least run time $T(p) > T_i$ of $3\|i\|$ bits programs. $t = \infty$ if no short p take $> T_i(\gamma)$ time, but such γ violate IP. Then $U'(\gamma) \stackrel{\text{df}}{=} \lambda(U_*^{-1}(\{\#\mathbb{N}, U_*(\gamma)\}))$, and $u(\alpha) \stackrel{\text{df}}{=} U(\gamma)$ for $\alpha = U'(\gamma)$; u diverges if $\alpha \notin U'(\{\#\mathbb{N}, 0, 1\}^{\mathbb{N}})$.

3.1 Elimination of 2nd Order Quantifiers with Independence Postulate

Some simplifications. To *reduce* all ST predicates, i.e., prove them equivalent to those with only integer quantifiers, we need to eliminate $\exists \alpha$ from $\exists \alpha P(\alpha, \beta)$, $\beta = \beta_1, \dots, \beta_k$ for reduced P .

We can restrict P to the form $\alpha, \beta \in R_\lambda \& \beta = u(\alpha) \& \alpha \in P_*$: First, note that $\lambda_\beta(Q) > t$ is reduced if Q is. Also $\forall \beta \exists \beta' \in R_\lambda \beta = u(\beta')$, so let us replace β with such β' by $P'(\alpha, \beta') \stackrel{\text{df}}{=} \beta' \in R_\lambda \& P(\alpha, u(\beta'))$. If $\exists \alpha P'(\alpha, \beta')$ reduces to $\beta' \in Q$ then $\exists \alpha P(\alpha, \beta)$ being $\Leftrightarrow \lambda_\beta(Q) > 0$ reduces, too.

Next, any pair (α, β) is $u_2(\gamma) \stackrel{\text{df}}{=} (u(\gamma, 0), u(\gamma))$ for a $\gamma \in R_\lambda$, so take $P_* \stackrel{\text{df}}{=} \{\gamma \in R_\lambda : P'(u_2(\gamma))\}$ for the above form of P . Let $P_0 \stackrel{\text{df}}{=} \{\gamma \in P_* : \lambda_{u(\gamma)}(P_*) = 0\}$, and $P_n \stackrel{\text{df}}{=} \{\gamma : \gamma \in P_* \vee \lambda_{u(\gamma)}(P_*) \leq 1/n\}$.

Then $\exists \alpha P(\alpha, \beta)$ reduces to $(\exists n \lambda_\beta(P_*) > 1/n \& \exists \gamma \in P_n \cap u^{-1}(\beta)) \vee \exists \gamma \in P_0 \cap u^{-1}(\beta)$. Note: $\lambda(P_0) = 0$, so $\beta \notin u(P_0)$ by IP and we can replace P_* with P_n for some n , to have $\forall x \lambda(P_* \cap x^u) > \lambda(x^u)/n$.

For that new P_* (which is reduced and can be routinely made compact) proving $\forall \beta \exists \alpha P(\alpha, \beta)$ is easiest with another version of Primal Chaos axiom: “For each β exists a $(\lambda_\beta)^\mathbb{N}$ -random $(\gamma_i)_i$ ”. Generating such γ with $\forall i \gamma_i \notin P_*$ has 0 chance, so, IP assures $\gamma_i \in P_*$ for some i .

Acknowledgments. I am grateful to Albert Meyer for valuable criticism.

References

- [1] Paul Halmos. 1974. Naive set theory. Springer-Verlag.
- [2] Randall Holmes. 2006. Alternative Set Theories. *Stanford Encyclopedia of Philosophy*.
https://plato.stanford.edu/entries/settheory-alternative/
- [3] Leonid A. Levin. 1976. On the Principle of Conservation of Information in Intuitionistic Mathematics. *Soviet Math. Dokl.* **17**/2:601-605 = *DAN SSSR* **227**/6:1293-1296.
- [4] Leonid A. Levin. 2013. Forbidden Information. *JACM*, **60**/2. https://arxiv.org/abs/cs/0203029
- [5] Leonid A. Levin. 2022. Gacs-Kucera Theorem. *TCS*, **929**:172-173. https://arxiv.org/abs/2207.12489
- [6] E.Zermelo. 1908. Untersuchungen über die Grundlagen der Mengenlehre. *Math. Ann.* **65**:261–281.

A Appendix: ZFC Axioms

ZFC axioms are sometimes given to undergraduates in an unintuitive, hard to remember list. Setting them in three pairs seems to help intuition.

Sets with a given ST property P (possibly with parameters c) are said to form a **class** $\{x : P_c(x)\}$. They may or may not form a set, but only sets are the domain of ZFC variables.

1. Chains axioms.

1a. Infinity. A non-empty set exists included in the union of its elements,

i.e. having no **sinks** in its membership graph ⁴ :

$$\boxed{\exists S, s \in S \forall x \in S \exists y \in S (x \in y)}$$

1b. Foundation/Induction (anti-dual to 1a) :

$$\boxed{\neg \exists S, s \in S \forall x \in S \exists y \in S (y \in x)}$$

Each non-empty set has membership graph **sources** (members disjoint with it).

2. Formula-defined sets.

2a. Extensionality. Assigning members (as in 2b) defines sets uniquely: sets with the same members are members of the same sets :

$$\boxed{\forall t (t \in x \Leftrightarrow t \in y) \Rightarrow \forall t (x \in t \Rightarrow y \in t)}$$

So, $x=y$ for sets or classes denotes $x \subset y \subset x$. ($x \subset y$, or $y \supset x$ mean $\forall t (t \in x \Rightarrow t \in y)$.)

2b. Replacement family (an axiom for each ST-defined relation $R_c(x, y)$).

The image $R_c(X) \stackrel{\text{df}}{=} \{y : \exists x \in X R_c(x, y)\}$ of any set X is a set if

each point $x \in X$ has a codomain set ⁵ :

$$\boxed{(\forall x \exists Y \supset R_c(\{x\}) \Rightarrow \forall X \exists Y = R_c(X))}$$

3. Inverses of functions. Define functions f , their domains δ_f , and classes f^{-1} of their inverses $g : f(\delta_f) \rightarrow \delta_f$, such that $f(g(x)) = x$ (so, $g \subset f^T \stackrel{\text{df}}{=} \{(f(x), x) : x \in \delta_f\}$).

3a. Choice. All functions have inverses :

$$\boxed{\forall f \exists g \in f^{-1}}$$

(The feasibility of computing inverses is the most dramatic open CS problem.)

3b. Powerset. Assures these huge classes f^{-1} of inverses $g \subset f^T$ still are sets :

$$\boxed{\forall x \exists Y \forall z (z \subset x \Rightarrow z \in Y)}$$

⁴Infinite ordinals are built by unique ordinal-valued function $f(s) = \sup_{x \in s \cap S} (f(x) + 1)$ on S .

⁵ $R_c(x, y)$ needs free variable c to deduce Paring $\{a, b\}$ (using Powerset). It is equivalent to $R(x, y)$ version (or $R_X(x, y)$ if not using Powerset) plus Paring axiom. (Build $X \times \{c\}$. Conversely, $2^a, 2^b, \{a, \{\}\}, \{a, b\}$.)