It is easily demonstrated that the rational reals are denumerable. Diagonalization can nevertheless be used to prove that the rationals are nondenumerable ... unless restrictions on the reordering of the diagonalization basis ... are imposed.

Dale Jacquette[1]

[Diagonalization] has no charm for me. I loathe it.

Ludwig Wittgenstein[2]

DIAGONALIZATION

Scott Lehmann[3]

Diagonalization is a simple procedure that has been employed to establish significant results in mathematics and logic. So I was astonished to find Dale Jacquette recently questioning its efficacy in a respectable publication. I dimly recalled that Wittgenstein also had a problem with diagonalization. Maybe I should look into this, I thought.

What I will do here is say what diagonalization is and outline some significant applications. Then I will consider the objections of Jacquette and Wittgenstein. I believe I can identify the problem with Jacquette's argument. I am less clear why Wittgenstein finds diagonalization distasteful, so I am not sure just how to respond. The best I can do is to consider critically what he might have meant by his cryptic remarks.

1 Diagonal Lemma

Basically, diagonalization extracts from a relation R on a set X a subset D_R of X that differs from each subset $R_y = \{x \in X : xRy\}$:

Diagonal Lemma (DL): Let X be a non-empty set and $R \subset X \times X$ be a binary relation on X. Define $R_y =_{df} \{x \in X : <x, y> \in R\}$. We may think of each R_y as a property of members of X. Then $D_R =_{df} \{x \in X : <x, x> \notin R\}$ defines a property of members of X that differs from each of of the R_y.

Proof: $x \in R_x$ iff $<x, x> \in R$ iff $x \notin D_R$. ■

We can picture what is going on here in a way that explains the term "diagonal". We may think of R as given by a "square" array of '0's and '1's, '0' indicating that $<x, y> \notin R$ and '1' indicating that $<x, y> \in R$:

[1] "Diagonalization in Logic and Mathematics" in Gabbay & Guenthner, eds., *Handbook of Philosophical Logic*, 2^{nd} edition, Vol. 11 (Kluwer, 2004), 55–147 at 74.

[2] *Lectures & Conversations* (California, 1966) at 28.

[3] Department of Philosophy, University of Connecticut, Storrs, CT 06269-2054

$$
\begin{array}{cc|cccccc}
 & & \multicolumn{5}{c}{\overbrace{\cdots \quad x \quad x' \quad x'' \quad \cdots}^{X}} & \\
\hline
 & \vdots & \ddots & \vdots & \vdots & \vdots & & \\
 & x & \cdots & \mathbf{1} & 1 & 0 & \cdots & \leftarrow \quad R_x \\
X & x' & \cdots & 0 & \mathbf{1} & 1 & \cdots & \leftarrow \quad R_{x'} \\
 & x'' & \cdots & 0 & 0 & \mathbf{0} & \cdots & \leftarrow \quad R_{x''} \\
 & \vdots & & \vdots & \vdots & \vdots & \ddots &
\end{array}
$$

This picture represents R as relating x to x but not to x' or to x'', as relating x' to x and to x' but not to x'', and as relating x'' to x' but not to x or to x''. Each row of the array corresponds to an element e of X and the '1's in it tell us which members of X belong to R_e; for example, x and x' belong to R_x, while x'' does not. The diagonal of this array is indicated in bold. Like each row, the diagonal determines a subset of X, consisting of those elements for which the diagonal entry is '1'. D_R is the complement (in X) of this set—the subset of X determined by interchanging '1's and '0's on the diagonal. Here x'' belongs to D_R, while neither x nor x' does. Obviously, D_R differs from each of the subsets determined by a row—for any $x \in X$, D_R differs from R_x at x.

It is the applications of DL, rather than the lemma itself, that are interesting. In ordinary life it is sometimes asserted that you can't prove a negative, but this happens all the time in mathematics: we prove that something doesn't exist by deriving a contradiction from the assumption that it does. DL is typically used in indirect proofs of this kind to obtain the desired contradiction. In the next sections, I will sketch a few of these applications.

2 Cantor's Theorem

Cantor's Theorem: If $\mathcal{P}(Z)$ is the set of subsets of Z, there is no 1-1 function $f : \mathcal{P}(Z) \to Z$.

Proof: If Z is empty, this is obvious. So suppose Z is non-empty and that there is such a function f. Let X be the image of $\mathcal{P}(Z)$ under f, i.e., $X = \{f(Y) : Y \subset Z\}$. Then f is onto X, and $f^{-1} : X \to \mathcal{P}(Z)$ is well-defined and onto: any subset of Z is $f^{-1}(y)$ for some $y \in X$. Define $R \subset X \times X$ by $<x, y> \in R$ iff $x \in f^{-1}(y) \cap X$. By DL, D_R is a subset of X distinct from each of the $R_y = \{x \in X : x \in f^{-1}(y)\} = f^{-1}(y) \cap X$. But this is impossible, since $D_R \subset X \subset Z$ and hence $D_R = f^{-1}(y) = f^{-1}(y) \cap X = R_y$ for some $y \in X$. So there is no such function $f : \mathcal{P}(Z) \to Z$, contrary to assumption. ■

From the theorem we may obtain some useful corollaries. First, some notation and terminology. Let $\mathbb{N}$ be the set of natural numbers, which we take to start with 0: $\mathbb{N} = \{0, 1, 2, \ldots\}$. A *segment of* $\mathbb{N}$ is either all of $\mathbb{N}$ or a subset $\{n \in \mathbb{N} : n \leq k\}$, where $k \in \mathbb{N}$. A set X is *denumerable* iff there is a 1-1 mapping from X onto a segment of $\mathbb{N}$, i.e., there is an exhaustive numbering or *enumeration* of the elements of X. A *sequence* is a mapping $s : S \to X$, where S is a segment of $\mathbb{N}$; we usually write s_i for $s(i)$ and $\{s_i\}$ for the sequence s. The *characteristic function* f_X of a set X is the function $X \to \{0,1\}$ such that $f_X(x) = 1$ iff $x \in X$.

COROLLARY 1: $\mathcal{P}(\mathbb{N})$ is not denumerable.
PROOF: Almost immediate from CANTOR'S THEOREM. ■

COROLLARY 2: The set of sequences $s : \mathbb{N} \to \{0,1\}$ is not denumerable.
PROOF: Immediate from COROLLARY 1, once we note that these sequences correspond to the characteristic functions of sets of natural numbers. ■

COROLLARY 3: The real numbers in [0,1] (i.e., the real numbers r such that $0 \leq r \leq 1$) are not denumerable.
PROOF: Each infinite sequence $s : \mathbb{N} \to \{0,1\}$ determines a real number $r_s \in [0,1]$ by

$$r_s = \sum_{i=0}^{\infty} s_i 2^{-(i+1)}$$

where $s_i = 0$ or $s_i = 1$, which we may write as a binary "decimal":

$$r_s = 0.s_0 s_1 s_2 \ldots$$

This mapping is onto the reals in [0,1], but it is not 1-1 since some rational reals have two binary expansions. For example,

$$1/2 = 0.1000\ldots = 0.0111\ldots$$

Despite this, we can argue that if the reals in [0,1] were denumerable, so would be the sequences $s : \mathbb{N} \to \{0,1\}$, contrary to COROLLARY 2. For if there were an enumeration of the reals in [0,1], we could obtain an enumeration of these sequences by replacing each real with (1) the sequence of its binary coefficients, if it has just one binary expansion, or with (2) one such sequence followed by the other, if it has two binary expansions. ■

Note that because $s \to r_s$ isn't 1-1, a direct argument for COROLLARY 3 from DL can't proceed by assuming the reals in $[0,1]$ are enumerated by $\{x_i\}$, setting $X = \mathbb{N}$, and letting $<k, i> \in R$ iff $s_k = 1$, where $s : \mathbb{N} \to \{0,1\}$ gives the coefficients in one of the binary expansions for x_i. In this case R_i will give the sequence of coefficients for x_i, and DL will tell us that the sequence given by D_R is different from any of them. But we cannot be sure that the real number defined by D_R differs from all those defined by the R_i. For all we know, D_R is just giving the other sequence of coefficients for some rational real in $[0,1]$.

I mention this because Jacquette makes a similar error in sketching a diagonalal argument for the non-denumerability of the irrationals in [0,1]. It can be shown that rationals have binary expansions that are *periodic*: after some point in the expansion, a finite sequence of coefficients repeats forever. For example, in

$$2/3 = 0.\underline{10}1010\ldots$$

the underlined sequence 10 repeats from the very beginning. The irrationals have non-periodic binary expansions. If we assume that the irrationals in [0,1] are denumerable and run diagonalization, we can be sure that the sequence of coefficients given by D_R is not the sequence of coefficients for any of the irrationals, but we cannot be sure that it is non-periodic. So we do not reach a contradiction.[4]

3 Tarski's Theorem

The next application is a version of Tarski's Theorem, establishing that truth is not expressible in certain formal languages. Let us take truth to be a property of sentences and identify properties with 1-ary relations. A relation is *expressible* in a language L provided a name for its extension is definable in L, given an interpretation I that fixes the meaning of L's vocabulary.

More precisely, let L be a first-order language, let I be a classical interpretation of L, and let U be the universe of I—the range of L's variables. A k-ary relation R in U is expressible in L under I provided there is a formula $\mathbf{A}(\mathbf{x_1}, \ldots, \mathbf{x_k})$ with k free variables $\mathbf{x_1}, \ldots, \mathbf{x_k}$ which is true under I of precisely the k-tuples in R. In this case, we may name R by adding to L a k-place predicate $\mathbf{R}$, interpreted as R, and regarding constructions $\mathbf{R}(\mathbf{t_1}, \ldots, \mathbf{t_k})$ as abbreviations of $\mathbf{A}(\mathbf{t_1}, \ldots, \mathbf{t_k})$, where $\mathbf{A}(\mathbf{t_1}, \ldots, \mathbf{t_k})$ results from $\mathbf{A}(\mathbf{x_1}, \ldots, \mathbf{x_k})$ by (1) changing bound variables so that $\mathbf{x_i}$ is free for $\mathbf{t_i}$ and then (2) substituting $\mathbf{t_i}$ for $\mathbf{x_i}$. More precisely, add $\mathbf{R}$ to L to obtain L^*, extend I to L^* by interpreting $\mathbf{R}$ as R, and provide a truth-preserving translation from L^* to L by $\mathbf{tr}(\mathbf{R}(\mathbf{t_1}, \ldots, \mathbf{t_k})) = \mathbf{A}(\mathbf{t_1}, \ldots, \mathbf{t_k})$.

We may extend this account of expressibility to individuals and functions by assimilating them to relations.

Individuals may be assimilated to properties that are uniquely theirs. An individual i of U may be named in L provided there is a formula $\mathbf{A}(\mathbf{x})$ that is true (under I) only of i. We may then introduce a name $\mathbf{i}$ for i, understanding atomic constructions $\mathbf{B}(\mathbf{i})$ as abbreviations for $\exists\mathbf{x}(\mathbf{B}(\mathbf{x})\&\mathbf{A}(\mathbf{x}))$.

A k-ary function $f : U \to U$ may be identified with its graph, a k+1-ary relation in U. If the graph is expressible by $\mathbf{A}(\mathbf{x_1}, \ldots, \mathbf{x_k}, \mathbf{y})$, we may introduce a name $\mathbf{f}$ for f and understand formulae in which $\mathbf{f}$ appears as abbreviations for what results from successively replacing parts $\mathbf{B}(\mathbf{f}(\mathbf{t_1}, \ldots, \mathbf{t_k}))$ by $\exists\mathbf{y}(\mathbf{B}(\mathbf{y})\&\mathbf{A}(\mathbf{t_1}, \ldots, \mathbf{t_k}, \mathbf{y}))$, where $\mathbf{B}$ is atomic and $\mathbf{f}$ does not occur in $\mathbf{t_i}$, until $\mathbf{f}$ no longer occurs.[5]

[4] Jacquette writes (op. cit. at 56) "The diagonal construction presumably guarantees that the diagonally defined number will also be irrational because it is constructed as an infinite nonrepeating decimal expansion of digits."

[5] Where f is a partial function (one not defined at each k-tuple in U^k), this procedure will produce weird results. For example, if f is 1-ary and not defined at the referent of

TARSKI'S THEOREM: Let L be a first-order language and I a classical interpretation of L in terms of which we understand reference and truth. Assume that:

1. Terms and formulae of L are nameable in L: for each L-term $\mathbf{t}$ there is an L-term $\ulcorner\mathbf{t}\urcorner$ that refers to $\mathbf{t}$ and for each L-formula $\mathbf{A}$ there is an L-term $\ulcorner\mathbf{A}\urcorner$ that refers to $\mathbf{A}$.

2. The naming operation $\mathbf{t} \longrightarrow \ulcorner\mathbf{t}\urcorner$ is expressible in L by $\mathbf{name}$, so that $\mathbf{name}(\mathbf{t})$ and $\ulcorner\mathbf{t}\urcorner$ are co-referring.

3. The substitution operation $<\mathbf{A}(\mathbf{x}), \mathbf{t}> \longrightarrow \mathbf{A}(\mathbf{t})$, where $\mathbf{x}$ is some designated variable, is expressible in L by $\mathbf{sub_x}$, so that $\mathbf{sub_x}(\ulcorner\mathbf{A}(\mathbf{x})\urcorner, \ulcorner\mathbf{t}\urcorner)$ and $\ulcorner\mathbf{A}(\mathbf{t})\urcorner$ are co-referring.

Then the property of being a true sentence of L is not expressible in L.

PROOF: We may apply DL in this situation, setting

$$\begin{aligned} X &= \text{the set of formulae with one free variable, viz. } \mathbf{x} \\ R &= \{<\mathbf{A}(\mathbf{x}), \mathbf{B}(\mathbf{x})> : \mathbf{B}(\ulcorner\mathbf{A}(\mathbf{x})\urcorner) \text{ is true}\} \end{aligned}$$

to obtain $D_R = \{\mathbf{A}(\mathbf{x}) : \mathbf{A}(\ulcorner\mathbf{A}(\mathbf{x})\urcorner) \text{ is not true}\} \neq \{\mathbf{A}(\mathbf{x}) : \mathbf{B}(\ulcorner\mathbf{A}(\mathbf{x})\urcorner) \text{ is true}\} = R_{\mathbf{B}(\mathbf{x})}$ for each $\mathbf{B}(\mathbf{x}) \in X$.

If truth were expressible in L, there would be a formula $\mathbf{T}(\mathbf{x})$ with free variable $\mathbf{x}$ true of precisely the true sentences of L, so that $\mathbf{T}(\ulcorner\mathbf{B}\urcorner)$ is true iff $\mathbf{B}$ is true. Consider $\neg\mathbf{T}(\mathbf{sub_x}(\mathbf{x}, \mathbf{name}(\mathbf{x})))$, obviously a member of X. Note that for $\mathbf{A} \in X$:

$$\begin{gathered} \neg\mathbf{T}(\mathbf{sub_x}(\ulcorner\mathbf{A}\urcorner, \mathbf{name}(\ulcorner\mathbf{A}\urcorner))) \text{ is true} \\ \text{iff (by co-referentiality of } \mathbf{name}(\ulcorner\mathbf{A}\urcorner) \text{ and } \ulcorner\mathbf{A}(\ulcorner\mathbf{A}\urcorner)\urcorner \\ \neg\mathbf{T}(\mathbf{sub_x}(\ulcorner\mathbf{A}\urcorner, \ulcorner\mathbf{A}(\ulcorner\mathbf{A}\urcorner)\urcorner)) \text{ is true} \\ \text{iff (by co-referentiality of } \mathbf{sub_x}(\ulcorner\mathbf{A}(\mathbf{x})\urcorner, \ulcorner\mathbf{t}\urcorner) \text{ and } \ulcorner\mathbf{A}(\mathbf{t})\urcorner \\ \neg\mathbf{T}(\ulcorner\mathbf{A}(\ulcorner\mathbf{A}\urcorner)\urcorner) \text{ is true} \\ \text{iff (by classical } \neg) \\ \mathbf{T}(\ulcorner\mathbf{A}(\ulcorner\mathbf{A}\urcorner)\urcorner) \text{ is not true} \\ \text{iff (by defintion of } \mathbf{T}(\mathbf{x})) \\ \mathbf{A}(\ulcorner\mathbf{A}\urcorner) \text{ is not true} \end{gathered}$$

So if $\mathbf{B}(\mathbf{x}) = \neg\mathbf{T}(\mathbf{sub_x}(\mathbf{x}, \mathbf{name}(\mathbf{x})))$, we have

$$\begin{aligned} R_{\mathbf{B}(\mathbf{x})} &= \{\ulcorner\mathbf{A}\urcorner : \neg\mathbf{T}(\mathbf{sub_x}(\ulcorner\mathbf{A}\urcorner, \mathbf{name}(\ulcorner\mathbf{A}\urcorner))) \text{ is true}\} \\ &= \{\ulcorner\mathbf{A}\urcorner : \mathbf{A}(\ulcorner\mathbf{A}\urcorner) \text{ is not true}\} \\ &= D_R \end{aligned}$$

$\mathbf{t}$, $(\mathbf{f}(\mathbf{t}) = \mathbf{f}(\mathbf{t}))$ is false, since it abbreviates $\exists\mathbf{y}(\mathbf{A}(\mathbf{t},\mathbf{y})\&(\mathbf{y} = \mathbf{y}))$ and $\mathbf{A}(\mathbf{x},\mathbf{y})$ expresses the graph of f. To avoid this, we may use $((\exists!\mathbf{y}\mathbf{A}(\mathbf{x},\mathbf{y})\&\mathbf{A}(\mathbf{x},\mathbf{y})) \vee (\neg\exists!\mathbf{y}\mathbf{A}(\mathbf{x},\mathbf{y})\&(\mathbf{y} = \mathbf{i})))$, where $\mathbf{i}$ names some individual of U_I, to express the graph of a total function on U_I that coincides with f where f is defined.

But this is impossible by DL. So there is no such formula $\mathbf{T}(\mathbf{x})$ and truth is not definable in L. ■

I should note that the argument can be modified to avoid use of DL, which is not essential to it. Simply observe that $\mathbf{B}(\ulcorner\mathbf{B}\urcorner) = \neg\mathbf{T}(\mathbf{sub_x}(\ulcorner\mathbf{B}\urcorner, \mathbf{name}(\ulcorner\mathbf{B}\urcorner)))$, so that $\mathbf{B}(\ulcorner\mathbf{B}\urcorner)$ is true iff $\neg\mathbf{T}(\mathbf{sub_x}(\ulcorner\mathbf{B}\urcorner, \mathbf{name}(\ulcorner\mathbf{B}\urcorner)))$ is true iff $\mathbf{B}(\ulcorner\mathbf{B}\urcorner)$ is not true. Thus there can be no formula $\mathbf{T}(\mathbf{x})$ such that for any sentence $\mathbf{A}$, $\mathbf{T}(\ulcorner\mathbf{A}\urcorner)$ is true iff $\mathbf{A}$ is true.

The theorem actually establishes that truth is inexpressible in a stronger sense. Normally, a relation that cannot be expressed in a language can be expressed in an extension of it, simply by adding a new predicate for the relation. For example, while the less-than relation is not expressible in the language L_{0S} of zero and successor under its standard interpretation, we can express it in the extension $L_{0S<}$ of L obtained by adding the 2-place predicate '$<$' and specifying that its exension is $\{<n, m> : n < m\}$. But this strategy will not work here. In general, if the conditions of the theorem hold for L, they will also hold for the language L_T that results from L by adding a new predicate T. The proof of TARSKI'S THEOREM then shows that we cannot (consistently) interpret T as truth, since we would have $\mathbf{B}(\ulcorner\mathbf{B}\urcorner)$ is true iff $\mathbf{B}(\ulcorner\mathbf{B}\urcorner)$ is not true.

The proof of the theorem assumes that $\neg$ is classical in the sense that $\neg\mathbf{A}$ is true *if* $\mathbf{A}$ is not true. This can fail in free semantics where sentences may be valueless and the negation of such a sentence is also valueless. Whether the proof fails as well depends upon how we understand truth and expressibility in such a case.

If we think of truth as a property that some sentences have and the others lack and of expressing truth as coming up with a formula $\mathbf{T}(\mathbf{x})$ that is true of the former and *false* of the latter, the proof will go through, for in it $\neg$ is applied to $\mathbf{T}(\mathbf{sub}(\ulcorner\mathbf{A}\urcorner, \mathbf{name}(\ulcorner\mathbf{A}\urcorner)))$, which by assumption will be either true or false. So truth in this sense won't be expressible, just in the classical case.

If we think of truth as a property that some sentences have and others lack, but only in a domain that excludes 'meaningless' sentences, and of expressing truth as coming up with a formula that is true of the true sentences, false of the false sentences, and neither true nor false of the others, then the proof will not go though, for $\neg\mathbf{T}(\mathbf{sub}(\ulcorner\mathbf{A}\urcorner, \mathbf{name}(\ulcorner\mathbf{A}\urcorner)))$ may be valueless. This of course does not show that truth *is* expressible in this sense. But even if it were—say, by $\mathbf{T}'(\mathbf{x})$—there will be other semantic notions that are not. One is clearly truth in the other sense—the property that true sentences, and no others, have. Another is having a truth-value—a property that every sentence has or lacks; for if we could express this by $\mathbf{V}(\mathbf{x})$, we could express the other sense of truth by $(\mathbf{V}(\mathbf{x})\&\mathbf{T}'(\mathbf{x}))$, assuming that a false conjunct renders a conjunction false.

My final illustration of diagonalization is a version of CHURCH'S THEOREM, for which we need the notion of a Gödel numbering. This is simply a 1-1 function g that assigns natural numbers to the terms and formulae of L. If L is interpreted so that its variables range over natural numbers (and perhaps additional entitites), then their Gödel numbers may be treated as proxies for terms and formulae of L, so that a term of L that names $g(\mathbf{t})$ or $g(\mathbf{A})$ may be regarded as naming $\mathbf{t}$ or $\mathbf{A}$. To see how this works, consider a variant of TARSKI'S THEOREM, which is essentially a corollary.

COROLLARY: Let L be a first-order language with a Gödel numbering g of terms and formulae and an interpretation I whose universe U_I includes $\mathbb{N}$. Suppose that:

1. The natural numbers are nameable in L: for each $n \in \mathbb{N}$, there is a term—a *numeral*—$\overline{\mathbf{n}}$ of L that names n.

2. The operation $n \longrightarrow g(\overline{\mathbf{n}})$ is expressible in L by $\mathbf{num}$, so that $\mathbf{num}(\overline{\mathbf{n}})$ and $\overline{g(\overline{\mathbf{n}})}$ are co-referring.

3. The substitution operation $<g(\mathbf{A}(\mathbf{x})), g(\mathbf{t})> \longrightarrow g(\mathbf{A}(\mathbf{t})$, where $\mathbf{x}$ is some designated variable, is expressible in L by $\mathbf{sb_x}$, so that $\mathbf{sb_x}(\overline{g(\mathbf{A}(\mathbf{x}))}, \overline{g(\mathbf{t})})$ and $\overline{g(\mathbf{A}(\mathbf{t}))}$ are co-referring.

Then the property of being the Gödel number of a true sentence of L is not expressible in L.

PROOF: The conditions of TARSKI'S THEOREM hold, provided we regard Gödel numbers as proxies for terms and formulae. Numerals for Gödel numbers then function as names for the terms and formulae they represent: $\ulcorner\mathbf{t}\urcorner = \overline{g(\mathbf{t})}$ and $\ulcorner\mathbf{A}\urcorner = \overline{g(\mathbf{A})}$. The naming operation $n \longrightarrow g(\overline{\mathbf{n}})$ subsumes the naming operation $\mathbf{t} \longrightarrow \ulcorner\mathbf{t}\urcorner$, and the substitution operation $<g(\mathbf{A}(\mathbf{x})), g(\mathbf{t})> \longrightarrow g(\mathbf{A}(\mathbf{t})$ mirrors the substitution operation $<\mathbf{A}(\mathbf{x}), \mathbf{t}> \longrightarrow \mathbf{A}(\mathbf{t})$. ■

4 Church's Theorem

CHURCH'S THEOREM establishes that first-order theories strong enough to represent the recursive sets are undecidable if consistent. The application of diagonalization is very similar to TARSKI'S THEOREM.

A first-order theory T consists of a first-order language L, a set Ax of sentences (of L) taken as axioms, and a proof method. A sentence $\mathbf{A}$ is *provable* ($\vdash \mathbf{A}$) if it is provable by this method from Ax; such sentences are *theorems* of T. A sentence is *refutable* if its negation is provable. T is *consistent* if no sentence $\mathbf{A}$ is both provable and refutable.

Suppose that the universe of I includes the natural numbers and that for every natural number n, there is a term $\overline{\mathbf{n}}$ of L naming n (under I). Roughly speaking, a set of numbers is representable in T provided it can be named in L and correct statements about what numbers belong or do not belong to it are theorems of T.

More precisely, an k-ary numerical relation R is *representable* in T if there is a formula $\mathbf{A}(\mathbf{x_1}, \ldots, \mathbf{x_k})$ of L such that

$$\text{if } <n_1, \ldots, n_k> \in R, \text{ then } \vdash \mathbf{A}(\overline{\mathbf{n_1}}, \ldots, \overline{\mathbf{n_k}})$$

$$\text{if } <n_1, \ldots, n_k> \notin R, \text{ then } \vdash \neg\mathbf{A}(\overline{\mathbf{n_1}}, \ldots, \overline{\mathbf{n_k}})$$

A relation in **N** is *recursive* if its characteristic function is recursive; for the somewhat involved definition of 'recursive function', see the appendix. The only things we need to know about recursive functions for the proof of CHURCH'S THEOREM are that (1) functions formed by composition (substitution) from recursive functions are recursive and (2) the characteristic function of the complement of a recursive numerical relation is also recursive.

Recursiveness is of interest because there is good reason to believe Church's Thesis that it precisely characterizes computability for a numerical functions and decidability for numerical relations. A numerical function is computable if there exists an algorithm for obtaining its value for any given argument. A numerical relation is decidable if there is an algorithm for determining of any given numbers whether they stand in the relation, so a relation is decidable iff its characteristic function is computable. These accounts are obviously only as precise as the notion of an algorithm.

A theory is *decidable* if there is an algorithm for determining, of any given sentence **A**, whether it is a theorem. Given an appropriate Gödel numbering g and Church's Thesis, a theory's decidability reduces to whether $Thm =_{df} \{g(\mathbf{A}) : \vdash \mathbf{A}\}$ is recursive.

CHURCH'S THEOREM: Let T be a first-order theory and I an interpretation of L_T such that:

1. The universe of I includes $\mathbb{N}$, and each natural number n is named by a term $\overline{\mathbf{n}}$ of L_T.

2. Every recursive set is representable in T.

3. g is a Gödel numbering of terms and formulae of L_T such that 0 is not the Gödel number of any formula and both num and $sb_{\mathbf{x}}$ are recursive, where

$$\begin{aligned} num(n) &= g(\overline{\mathbf{n}}) \\ sb_{\mathbf{x}}(n, m) &= \begin{cases} g(\mathbf{A}(\mathbf{t})) & \text{if } n = g(\mathbf{A}(\mathbf{x})) \text{ and } n = g(\mathbf{t}) \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Then the set Thm of Gödel numbers of sentences that are theorems of T is not recursive if T is consistent.

PROOF: Apply DL, with $X = \mathbb{N}$ and

$$R = \{<k, n> : n = g(\mathbf{A}(\mathbf{x})) \text{ and } \vdash_T \mathbf{A}(\overline{\mathbf{k}})\}$$

to obtain a set $D = D_R = \{k : <k, k> \notin R\} \neq \{k : <k, n> \in R\} = R_n$, for each $n \in \mathbb{N}$.

Note that if T is consistent, each recursive set S is R_n for some n. (Assume T is consistent. If S is recursive, it is represented in T by a formula $\mathbf{A}(\mathbf{x})$. But then $S = R_{g(\mathbf{A}(\mathbf{x}))}$. For if $k \in S$, then $\vdash_T \mathbf{A}(\overline{\mathbf{k}})$ and $k \in R_{g(\mathbf{A}(\mathbf{x}))}$; and if $k \notin S$, then $\vdash_T \neg \mathbf{A}(\overline{\mathbf{k}})$, so $\not\vdash_T \mathbf{A}(\overline{\mathbf{k}})$ (by consistency) and $k \notin R_{g(\mathbf{A}(\mathbf{x}))}$.) It follows that D cannot be recursive, since it is distinct from all of the R_n.

Next note that

$$k \in D \text{ iff } sb_{\mathbf{x}}(k, num(k)) \notin Thm$$

1. If $k \in D$, then $<k, k> \notin R$, so either (a) k isn't the Gödel number of a formula $\mathbf{A}(\mathbf{x})$ or (b) $k = g(\mathbf{A}(\mathbf{x}))$ but $\not\vdash_T \mathbf{A}(\overline{\mathbf{k}})$. In case (a), $sb_{\mathbf{x}}(k, num(k)) = 0 \notin Thm$; in case (b), $sb_{\mathbf{x}}(k, num(k)) = g(\mathbf{A}(\overline{\mathbf{k}})) \notin Thm$. So if $k \in D$, $sb_{\mathbf{x}}(k, num(k)) \notin Thm$.

2. If $sb_{\mathbf{x}}(k, num(k)) \notin Thm$, it's either because (a) $sb_{\mathbf{x}}(k, num(k)) = 0$ or because (b) $k = g(\mathbf{A}(\mathbf{x})$ but $\not\vdash_T \mathbf{A}(\overline{\mathbf{k}})$. In case (a), k isn't the Gödel number of a formula $\mathbf{A}(\mathbf{x})$, so in either case $<k, k> \notin R$ and $k \in D$. So if $sb_{\mathbf{x}}(k, num(k)) \notin Thm$, then $k \in D$.

Now let $f_{\neg Thm}$ be the characteristic function of $\neg Thm$, the complement of Thm. If f_D is the characteristic function of D, we have

$$f_D(x) = f_{\neg Thm}(sb_{\mathbf{x}}(x, num(x)))$$

so that f_D is obtained by composition from $f_{\neg Thm}$, $sb_{\mathbf{x}}$, and num. If Thm were recursive, $f_{\neg Thm}$ would also be recursive, and therefore f_D and D would be recursive. Since this is impossible, Thm is not recursive. ∎

To make the proof of Church's Theorem easy, I have made some high-level assumptions about T. But it does not take much to make them true:

1. We can get names $\overline{\mathbf{n}}$ in L for all the natural numbers n provided L has a name for zero and a name for the successor function.

2. Though the proof is not trivial, it can be shown that the recursive *functions* are representable in T, provided L contains names for zero, successor, addition and multiplication, and Ax contains some simple truths relating them. A k-ary numerical function f is representable in T provided there is a formula $\mathbf{A}(\mathbf{x_1}, \ldots, \mathbf{x_k}, \mathbf{y})$ such that $\vdash \forall \mathbf{x_1} \ldots \forall \mathbf{x_k} \exists! \mathbf{y} \mathbf{A}(\mathbf{x_1}, \ldots, \mathbf{x_k}, \mathbf{y})$ which represents the graph of f. A numerical relation R will be representable if its characteristic function is representable (provided $\vdash (\mathbf{0} \neq \mathbf{1})$).

3. It is easier, if tedious, to show that the operations num and $sb_{\mathbf{x}}$ are recursive, provided that L's vocabulary is effectively specified and the Gödel numbering g is standard.

5 Jacquette's objection to diagonalization

I now turn to Jacquette's misgivings about diagonalization. Jacquette suggests that the diagonal method, as standardly employed to show that the real numbers in [0,1] are non-denumerable, is "unsound", because the same procedure can be employed to show something that is demonstrably false, viz. that the rational numbers in [0,1] are non-denumerable. He is wrong about this—but wrong in an interesting way. His argument rests on the following claim:

JACQUETTE'S CLAIM (JC): Let $S = \{s_0, s_1, \ldots\}$ be an infinite sequence of infinite sequences s_i of 0s and 1s—the k-th member $s_{i\,k}$ of each $s_i \in S$ is 0 or 1—such that for each $k \in \mathbb{N}$ there are infinitely many members s_i of S for which $s_{i\,k} = 0$ and infinitely many for which $s_{i\,k} = 1$. Let $\{x_i\}$ be *any* infinite sequence of 0s and 1s. Then S may be rearranged in a such a way that the k-th member of its k-th sequence is x_k.

PROOF: If S is not such that $s_{k\,k} = x_k$ for each $k \in \mathbb{N}$, let j be the least k for which $s_{k\,k} \neq x_k$. By assumption, there are infinitely many sequences s_n of S such that $s_{n\,j} = x_j$, so we may let m be the least $n \geq j$ for which this is so. By interchanging s_j and s_m in S, we obtain a rearrangement of S that has the desired property up through its j-th member. Repeat as needed to obtain a reordering of the original enumeration of S with diagonal $\{x_i\}$. ∎

Given JC, Jacquette can use diagonalization to conclude, absurdly, that the rationals in [0,1] are non-denumerable. Here is the argument:

> Suppose the rationals in [0,1] are denumerable. Then so is the set Y of sequences of coefficients of their binary expansions, where these include both expansions for rationals that have two. Because the rationals are dense, any enumeration S of the members of Y satisfies the condition of JC. Let $\{x_i\}$ be a sequence of 0s and 1s such that $\{\overline{x}_i\}$, where
>
> $$\overline{x}_i = \begin{cases} 0 & \text{if } x_i = 1 \\ 1 & \text{if } x_i = 0, \end{cases}$$
>
> gives the coefficients of some rational in [0,1]. By JC, we may assume that $S = \{s_0, s_1, \ldots\}$ is an enumeration of Y such that $s_{k\,k} = x_k$. Apply DL here, with $X = \mathbb{N}$ and $R = \{<i, j> : s_{j\,i} = 1\}$ to obtain $D_R = \{k : s_{k\,k} = 0\} = \{k : x_k = 0\} = \{k : \overline{x}_k = 1\}$ different from any $R_j = \{k : s_{j\,k} = 1\}$. Accordingly, $\{\overline{x}_i\}$ is distinct from all the members of Y, though it defines a rational in [0,1]. So the supposition that the rationals in [0,1] are denumerable is false.

Unfortunately for Jacquette, the same argument reveals that JC is false. If $\{\overline{x}_i\}$ is in S and S could be rearranged so that the diagonal sequence is $\{x_i\}$, then DL would tell us that $D_R = \{\overline{x}_i\}$ is not in S! Where does the proof of JC go wrong?

If the sequence $\{\overline{x}_i\}$ is in S to begin with—suppose it is the j-th member, so that $s_j = \{\overline{x}_i\}$—the sequence of interchanges Jacquette describes will not reorder the original sequence, because s_j will never be assigned a position in it. When the interchange process reaches s_j, it will be bumped along to some more remote position, where the same thing will happen again, *ad infinitum*. What is enumerated by this process is not S but what results from S by removing $s_j = \{\overline{x}_i\}$. The diagonal sequence of the corresponding array will indeed be

$\{x_i\}$, so that diagonalization will generate $s_j = \{\overline{x}_i\}$. But there is no problem here because this sequence is no longer in the enumeration.

6 Orderings and reorderings

Perhaps stepping back a bit will help clarify the situation. While Jacquette sometimes speaks of his reordering as a "permutation" of the original ordering, it would be more accurate to say that it *results* from an infinite sequence of *reorderings by permutation*.

A binary relation $\prec$ on X *orders* X if $\prec$ is transitive: $(x \prec y \,\&\, y \prec z) \rightarrow x \prec z$; anti-reflexive: $\neg x \prec x$; and complete: $x \neq y \rightarrow (x \prec y \vee y \prec x)$.[6] For example, $<$ orders $\mathbb{N}$. Ordered sets X (under $\prec_X$) and Y (under $\prec_Y$) are *similar* iff there is a 1-1 mapping $f : X \longrightarrow Y$ from X onto Y such that $x \prec_X x'$ iff $f(x) \prec_Y f(x')$.[7] For example, $\mathbb{N}$ under $<$ is similar to $\mathbb{E}$ under $<$, where $\mathbb{E}$ is the set of evens, but not to $\mathbb{N}$ under $\prec$, where $n \prec m$ iff (1) $n < m$ and n and m are both even or both odd or (2) n is even and m is odd—i.e., the evens in order precede the odds in order.

Similar ordered sets are said to have to same *type* of order (or the same *order-type*). The order-type of $<$ on $\mathbb{N}$ is designated ω; the order-type of $\prec$ on $\mathbb{N}$ is designated $\omega + \omega$.[8] A *permutation* of a set X is just a mapping $p : X \longrightarrow X$ that is 1-1 and onto. If $\prec$ orders X, a permutation p of X induces a reordering $\prec_p$ of X according to $x \prec_p x'$ iff $p^{-1}(x) \prec p^{-1}(x')$. Obviously, X under $\prec_p$ is similar to X under $\prec$, so $\prec_p$ and $\prec$ are of the same order-type. Accordingly, any reordering by permutation of $\mathbb{N}$ under $<$ has order-type ω.

In Jacquette's case, we are reordering an infinite sequence $\{s_i\}$ by permuting its terms. To simplify notation, let us simply represent s_i by i, so that Jacquette's reorderings can be represented as reorderings by permutation of $\mathbb{N}$. Each permutation here is a simple interchange or *2-cycle*: there are numbers n and m such that $p(n) = m$, $p(m) = n$, and $p(k) = k$ otherwise.

Successive reorderings by 2-cycles can move 0 up in the ordering of $\mathbb{N}$, as depicted in the diagram below, where $\prec_i$ is left-to-right, $\prec_0$ is $<$, and the k-th

[6] These conditions define a *linear* ordering; if completeness is omitted, we have a *partial* ordering.

[7] A broader notion of similarity can be obtained by dropping the requirement that the relations $\prec_X$ and $\prec_Y$ be orderings.

[8] In pure set theory, ω, the first limit ordinal, is ordered by $\in$ in the same way that the natural numbers are ordered by $<$. $\omega+\omega$, the second limit ordinal, is ordered by $\in$ in the same way that the natural numbers are ordered by $\prec$. The ordinals are well-ordered by $\in$—every non-empty subset of an ordinal has a $\in$-least element—and names for ordinals are used for the order-types of well-ordered sets of the same order-type. For extended discussion, see Chapter III ("Order and Similarity") of Abraham A. Fraenkel, *Abstract Set Theory* (Amsterdam; North-Holland, 1968).

permutation takes us from $\prec_k$ to $\prec_{k+1}$:

$$\begin{array}{cccccc}
\prec_0 & 0 & 1 & 2 & 3 & 4 & \cdots \\
\prec_1 & 1 & 0 & 2 & 3 & 4 & \cdots \\
\prec_2 & 1 & 2 & 0 & 3 & 4 & \cdots \\
\prec_3 & 1 & 2 & 3 & 0 & 4 & \cdots \\
\prec_4 & 1 & 2 & 3 & 4 & 0 & \cdots \\
\vdots & & & \vdots & & &
\end{array}$$

This models the essentials of the case where $\{\overline{x}_i\}$, here represented by 0, is bumped up in the ordering *ad inf.* Each permutation gives us a ordering of type ω. It does not follow, however, that the sequence of reorderings gives us a ordering of type ω; indeed it seems to be heading for an ordering of type $\omega+1$:[9]

$$1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow \cdots \qquad 0$$

where the arrows go to immediate successors. However, we have yet to characterize the ordering of X (if any) that results from an infinite sequence $\prec_i$ of reorderings.

Before doing so, let me give a couple of other examples of reordering $\mathbb{N}$ by permutation, chosen in part as test cases for the characterization I shall propose.

1. Use the 2-cycle that takes 1 to 2 and 2 to 1 to reorder $\mathbb{N}$ *ad inf*:

$$\begin{array}{cccccc}
\prec_0 & 0 & 1 & 2 & 3 & \cdots \\
\prec_1 & 0 & 2 & 1 & 3 & \cdots \\
\prec_2 & 0 & 1 & 2 & 3 & \cdots \\
\prec_3 & 0 & 2 & 1 & 3 & \cdots \\
\prec_4 & 0 & 1 & 2 & 3 & \cdots \\
\vdots & & & \vdots & &
\end{array}$$

2. Use progressively larger k-cycles to reorder $\mathbb{N}$ so that progressively larger sections of the evens precede corresponding sections of the odds. The cycles are indicated by arrows in the diagram:

$$\begin{array}{cccccccccc}
\prec_0 & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & \cdots \\
 & & \downarrow & \downarrow & & & & & & \\
\prec_1 & 0 & 2 & 1 & 3 & 4 & 5 & 6 & 7 & \cdots \\
 & & & \downarrow & \downarrow & \downarrow & & & & \\
\prec_2 & 0 & 2 & 4 & 1 & 3 & 5 & 6 & 7 & \cdots \\
 & & & & \downarrow & \downarrow & \downarrow & \downarrow & & \\
\prec_3 & 0 & 2 & 4 & 6 & 1 & 3 & 5 & 7 & \cdots \\
\vdots & & & & \vdots & & & & &
\end{array}$$

(Think of the natural numbers as an endless zipper, mating the evens with the odds, which is being unzipped by reordering.)

[9]Named for the successor $\omega \cup \{\omega\}$ of ω.

If we ask where these reorderings are heading, it appears that the first sequence results in a partial ordering of $\mathbb{N}$ indicated by

$$\begin{array}{ccc} & \nearrow 1 \searrow & \\ 0 & & 3 \rightarrow 4 \rightarrow 5 \rightarrow \cdots \\ & \searrow 2 \nearrow & \end{array}$$

while the second results in an ordering of $\mathbb{N}$ (of type $\omega + \omega$) in which the evens (in order) precede the odds (in order):

$$0 \rightarrow 2 \rightarrow 4 \rightarrow 6 \rightarrow \cdots \qquad 1 \rightarrow 3 \rightarrow 5 \rightarrow 7 \rightarrow \cdots$$

Let us define the *limit* $\lim_{i\to\infty} \prec_i$ of an infinite sequence $\{\prec_i\}$ of relations on X (which need not order X) to be the relation on X such that

$$<x, y> \in \lim_{i\to\infty} \prec_i \ \text{ iff } \exists k \forall i((i \geq k) \rightarrow x \prec_i y)$$

Note that, in contrast to the usual situation, $\lim_{i\to\infty} \prec_i$ always exists, though it may be the empty relation on X. It is easy to see that in each of the examples above, $\lim_{i\to\infty} \prec_i$ is the relation on $\mathbb{N}$ that intuitively *results from* the sequence $\{\prec_i\}$ of reorderings by permutation.

In particular, in the example that models Jacquette's sequence of reorderings we see that $\lim_{i\to\infty} \prec_i$ on $\mathbb{N}$ has order-type $\omega + 1$. The k-th sequence in the resulting reordering is s_{k+1}, and s_0 appears nowhere in this sequence. In terms of DL, $X = \mathbb{N}$ and $R = \{<k, i> : s_{(k+1)_i} = 1\}$. DL establishes that D_R differs from $R_{s_{k+1}}$ for each k, but not that D_R differs from R_{s_0}.

Jacquette does consider the possibility that his reordering procedure is not "conservative", i.e., that what issues from it does not include all the elements of the original list. But he finds this mysterious:

> Now if that is true, where do the missing numbers go? They must vanish from the diagonalization basis altogether, which seems implausible to say the least. ... [The procedure does] nothing more than leave each expansion where it is in the list if the simple condition of having a certain digit in its nth expansion place is satisfied, and otherwise to shift it from a finite row n in the list to a finitely lower location in the list at row $n + m$... if the condition is not satisfied. This by itself does not appear capable of exporting any real number expansion entirely outside the list....[10]

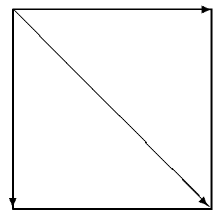
Jacquette's mistake here is to assume that successive reorderings of a list must generate a list of the same items. This is true if the sequence of reorderings is finite, but it need not be true if the sequence is infinite, as it is in this case. Here we end up not with a list, but an ordering of type $\omega + 1$. Its initial ω portion is a list, but it omits one member of the original list—the one bumped to the $\omega + 1$ position in the ordering.

In the end, the situation is like that depicted by the second diagram in one of Wittgenstein's remarks: [11]

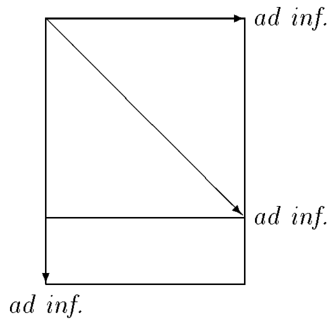
[10] Jacquette, op. cit. at 85–6.

[11] *Remarks on the Foundations of Mathematics*, Rev. Edn. (MIT, 1978), at 128 (Part II).

11. Since my drawing is after all only the *indication* of infinity, why must it be like this



and not like this



As we move down the diagonal in the second diagram, we run through the members of X (shown on the x-axis) before we run through those of Y (the y-axis). In this case, the subset D_R of X constructed from the diagonal differs from R_y for any y along the left edge of the square, but we have no assurance that it differs from R_y for any y below. But that's where $y = s_0$ is in Jacquette's case.

7 Wittgenstein's problems with diagonalization

Having invoked Wittgenstein, let me turn to his objections—or what I can make of them. Like Jacquette, Wittgenstein is concerned with Cantor's use of diagonalization to show that the reals are non-denumerable. Recall that in this case we assume for *reductio* that the reals in [0,1] are denumerable and then apply diagonalization to an enumeration of their binary expansions to obtain a binary expansion not in the enumeration but nevertheless belonging to a real number in [0,1]. Wittgenstein writes:

> 3. If someone says: "Shew me a number different from all these", and is given the rule of the diagonal for answer, why should he not say: "But I didn't mean it like that!"? What you have given me is a rule for the step-by-step construction of numbers that are successively different from each of these.
>
> "But why aren't you willing to call this too a method of calculating a number?"—But what is the method of calculating, and what the

> result, here? You will say that they are *one*, for it makes sense now to say: the number C is bigger than ... and smaller than ...; it can be squared etc. etc.
>
> Is the question not really: What can this number be *used* for? True, that sounds queer.—But what it means is: what are its mathematical surroundings?[12]

and a bit later:

> 5. Let us say—not: "This method gives a result", but rather: "it gives an infinite series of results". ...[13]

Here Wittgenstein appears to be arguing that what the diagonal method produces in this case is not a real number r, but an infinite sequence of rational numbers $\{r_n\}$ that, from the perspective of real number theory, converges to r, namely, the sequence of partial sums

$$r_n = \sum_{i=0}^{n} c_i 2^{-(i+1)}$$

where

$$r = \sum_{i=0}^{\infty} c_i 2^{-(i+1)} =_{df} \lim_{n\to\infty} r_n$$

As a method of calculation, then, the diagonal method at best generates successive approximations to a real number, but not the number itself. I say "at best" because in order to calculate the approximation r_n, we need to be able to determine the k^{th} binary coefficient of the k^{th} real in the enumeration for each $k \leq n$, and so far we've not been given enough information to do this.

But why is any of this relevant? The "mathematical surroundings" here are the background theory of real numbers and the *reductio* assumption that the reals can be enumerated. The theory of real numbers takes sequences like $\{r_n\}$ to converge to a real number—in this case, one that differs from the k^{th} real in the supposed enumeration in its k^{th} binary coefficient. The number produced by the diagonal method seems perfectly useful in this context (though it might not serve other purposes well). So what's the problem?

For Wittgenstein the problem seems to be the theory of real numbers. In his view, this theory makes use of illegitimate notions and we don't need it to do the practical work of mathematics.

It is clear that what the diagonal method produces, in case where we assume that the real numbers in [0,1] are denumerable, is, according to the theory of real numbers, a real number in [0,1] that is distinct from those in the assumed enumeration. Wittgenstein's complaints about this "diagonal number"—that we are not given enough information to calculate it, or even rational approximations thereof—therefore seem to carry over to this theory.

[12] Id. at 126.
[13] Id.

That Wittgenstein questions the authority of the theory of real numbers here is supported by the following remarks:

> 16. ...Asked: "Can the real numbers be ordered in a series?" the conscientious answer might be: "For the time being I can't form any precise idea of that". —"But you can order the roots and the algebraic numbers for example in a series; so surely you understand the expression!"—To put it better, I *have got* certain analogous formations, which I call by the common name 'series'. But so far I haven't any certain bridge from these cases to that of 'all real numbers'. Nor have I any general method of trying whether such-and-such a set 'can be ordered in a series'.
>
> Now I am shewn the diagonal procedure and told: "Now here you have the proof that this ordering can't be done here". But I can reply: "I don't know—to repeat—what it is that *can't be done* here". Though I can see that you want to shew a difference between the use of "root", "algebraic number", etc. on the one hand, and "real number" on the other. Such a difference as e.g. this: roots are called "real numbers", *and so too* is the diagonal number formed from the roots. And similarly for all series of real numbers. For this reason it makes no sense to talk about a "series of all real numbers", just because the diagonal number for each series is also called a "real number". Would this not be as if any row of books were itself ordinarily called a book, and now we said: "It makes no sense to speak of 'the row of all books', since this row would itself be a book."[14]

When Wittgenstein denies being able to form "any precise idea of that", to what is he referring, and why the difficulty? One hypothesis is: "that" refers to a sequence $\{r_i\}$ of all the real numbers. If forming a precise idea of such a thing requires specifying the sequence by identifying r_i for each i, then of course we cannot do this: Cantor's proof shows there is no such sequence. If Cantor's proof required such a precise idea, then there would be a problem, since the proof would presuppose something incompatible with what it establishes. But this is not the case. The proof establishes a negated existential: there is no enumeration $\{r_i\}$ of the real numbers. It does so, as usual, by showing that a contradiction follows from assuming the existential (there exists an enumeration $\{r_i\}$ of the reals). If (like Wittgenstein) you demand constructive proofs of existence claims—require that a proof of $\exists \mathbf{x} \mathbf{A}(\mathbf{x})$ exhibit $\mathbf{t}$ such that $\mathbf{A}(\mathbf{t})$, rather than simply showing that $\forall \mathbf{x} \neg \mathbf{A}(\mathbf{x})$ leads to a contradiction—then you will require that proof of the existential exhibit such a sequence of reals. But you cannot require as much in making an existential assumption.

Elsewhere Wittgenstein raises this sort of problem, only to put it immediately to rest:

[14] Id. at 130.

> 28. ... The difficulty which is felt in connexion with *reductio ad absurdum* in mathematics is this: what goes on in this proof? ... How—one would like to ask—can one so much as assume the mathematically absurd at all? That I can assume what is physically false and reduce it *ad absurdum* gives me no difficulty. but how to think the—so to speak—unthinkable?
>
> What an indirect proof says, however, is: "If you want *this* then you cannot assume *that*: for only the opposite of what you do not want to abandon would be combinable with *that*".[15]

If mathematically false claims are "unthinkable", it is not in a sense that precludes entertaining them for the purposes of indirect proof. Similarly, we should avoid reading Wittgenstein's claims that such and such "makes no sense" as maintaining that they are literally *nonsense*. If it "series of all real numbers" really made no sense (as required by theories that identify meaning with denotation), it is difficult to see how "there is no series of all real numbers" could be meaningful, let alone true. In accord with Wittgenstein's general view that "meaning" refers to use, we should instead take "X makes no sense" or "X means nothing" to indicate something peculiar about certain uses of "X". For example, if we were to go along with Wittgenstein's "sober" suggestion that

> 20. ... If something is called a series of real numbers, then the expansion given by the diagonal procedure is also called a 'real number', and is moreover said to be different from all members of the series.[16]

a claim that *therefore* the reals are non-denumerable might strike us as a little odd. For this conclusion is then just part of the definition of "real number", whereas "therefore" in ordinary contexts outside of logic signals some interesting or unexpected conclusion.[17]

I think that what Wittgenstein can't get his head around is not the notion of a *series* of all real numbers, but simply the standard notion of *real number*, which assures that any infinite sequence of 0s and 1s corresponds to a real number in [0,1], and conversely. From Wittgenstein's perspective, there is no acceptable account of the real numbers that assures that the diagonal number is *real*—unless we make this true by stipulation *via* something like the "sober" sentence of 20. But then, according to him, Cantor's proof reveals nothing exciting about the reals.

(Note that we may use the diagonal procedure to generate the set X of infinite sequences of 0s and 1s (and therefore the reals in [0,1]): X is the smallest set including the constant 0-sequence $\{r_n\}$, where $r_n = 0$ for each n, that is closed under diagonalization. Starting with the constant 0-sequence amounts

[15] Id. at 285 (Part V).

[16] Id. at 131 (Part II).

[17] "10. It means nothing to say: "*Therefore* the X numbers are not denumerable". One might say something like this: I call number-concept X non-denumerable if it has been stipulated that, whatever numbers falling under this concept you arrange in a series, the diagonal number of this series is also to fall under that concept." Id. at 128.

to applying DL with $X = Y = \mathbb{N}$ and $R = \emptyset$; the diagonal sequence that results is the constant 1-sequence. These constant sequences may be arranged in the "diagonal basis" so that any desired sequence of 0s and appears along the diagonal, the arrangement corresponding to some new binary relation R on $\mathbb{N}$. Still, it requires some thought to see that this will work—that we are getting just the desired sequences from the process.)

8 Motivating the real numbers

Now it must be admitted that idea that any sequence of 0s and 1s defines a real number in [0,1] is not something we get for free—it corresponds to a significant 'continuity' assumption about points on the real line. Let us quickly review some steps in the development of the theory of real numbers.

The basic concept of number is described by Newton as "an abstract ratio of a certain quantity to another quantity taken as a unit."[18] Suppose the quantity in which we are interested is the length of line segments. We are going to select some segment as our unit—we set its length = 1—and ask about the length of other segments relative to it. What system of numbers is needed for this?

We need a number for every segment (congruent segments will have the same number). By appeal to standard geometric assumptions about congruence, we may assume that these segments lie on the same line as our unit and indeed that they have the same left endpoint, which we may label "0". Since any such point p on this line to the right of 0 defines a segment or interval $[0, p]$, we need a different number for each distinct point p to the right of 0, this number being the ratio of the length of $[0, p]$ to that of our unit. The number 1 corresponds to the unit's right endpoint, so we may identify the unit with the unit interval [0,1].

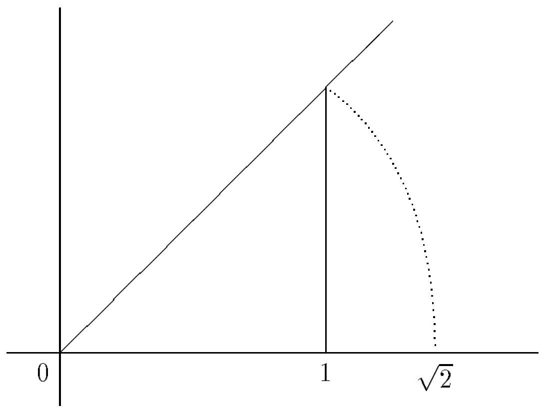
Let us call our line the "real line", anticipating that points to the right of 0 will correspond to the positive reals. Accordingly, I will generally use numerals—names for numbers—to label such points and shall indicate that p is left of q by "$p < q$" so that the left-to-right ordering of points corresponds to the less-than ordering of the corresponding reals.

The issue of what numbers are needed to measure segment or interval length relative to the unit interval thus reduces to the issue of what points p there are such that $0 < p$. Standard assumptions of Euclidean plane geometry (e.g., lines that intersect do so at a point) give us the integral points 2, 3, 4, ... and the (positive) rational points, the latter revealed by constructions involving similar triangles.

As is well-known, however, geometric considerations reveal that there are points P that are "irrational" in the sense that there is no way of dividing [0,1] into equal subintervals so that $[0, p]$ is made up of some integral number of them,

[18] Isaac Newton, *General Arithmetic*, quoted in A. D. Aleksandrov, "A general view of mathematics" in Vol. 1 of Alexandrov, Kolmagorov, and Lavrentev, eds., *Mathematics: its content, methods, and meaning* (Cambridge: MIT Press, 1963), 1–64 at 27

i.e., so that the length of $[0,p]$ is m/n for some integers m and n. The point labelled "$\sqrt{2}$" in the diagram is an example.



For any given point p—say, $\sqrt{2}$—we may construct a sequence $\{I_i\}$ of *nested* intervals (i.e., $I_{i+1} \subset I_i$) whose intersection is $\{p\}$, provided we assume:

ARCHIMEDES' PRINCIPLE (AP): Let $[a,b]$ be a line segment. Then for any line segment $[c,d]$, there is a finite number of points $c_0 < \ldots < c_k$ on the line determined by $[c,d]$ such that (1) $c = c_0$, (2) $d < c_k$, and (3) $[c_i, c_{i+1}]$ is congruent to $[a,b]$.

This looks innocuous, and it is. But it tells us something significant about line segments, namely that they are neither infinitely large nor infinitely small. There is no segment $[c,d]$ so big that it can't be "covered" by a finite number of copies of segment $[a,b]$, laid end to end; nor is there any segment $[a,b]$ so small that a finite number of copies thereof doesn't suffice to "cover" any given segment.

We may now construct a nested sequence $\{I_i\}$ of intervals that narrows down to p. By AP we know that some finite number of copies of [0,1] laid end-to-end suffices to cover $[0,p]$. Let $n+1$ be the least such number, and let $\{I_0\} = [n, n+1]$. Note that $p \neq n$ and $p \in I_0$. Now divide I_0 in half: $\{I_0\} = [n, n+1/2] \cup [n+1/2, n+1]$. If $p \in [n, n+1/2]$, let $I_1 = [n, n+1/2]$; note that $p \neq n$. Otherwise, let $I_1 = [n+1/2, n+1]$; note that $p \neq n+1/2$ in this case. In either case, we have $p \in I_1$. Note that

$$I_1 = \left[n + \frac{m_1}{2}, n + \frac{m_1+1}{2} \right]$$

$$\text{where } m_1 = \begin{cases} 0 & \text{if } p \in [n, n+1/2] \\ 1 & \text{otherwise} \end{cases} \quad \text{and } P \neq n + \frac{m_1}{2}$$

Continuing in this manner, we get

$$I_i = \left[n + \sum_{j<i} \frac{m_j}{2^j} + \frac{m_i}{2^i}, n + \sum_{j<i} \frac{m_j}{2^j} + \frac{m_i+1}{2^i} \right]$$

$$\text{where } m_k = 0 \text{ or } m_k = 1,\ p \in I_i, \text{ and } p \neq n + \sum_{j<i} \frac{m_j}{2^j} + \frac{m_i}{2^i}$$

Since $p \in I_i$ for each i,

$$p \in \bigcap_{i=0}^{\infty} I_i$$

Moreover, p is the only point common to all the intervals I_i. For if there were another point q, the segment $[p, q]$ (assume for convenience that $p < q$) would be infinitesimal, contrary to AP, because it is shorter than $1/2^i$ (and hence shorter than $1/i$) for each i. Finally, notice that

$$p = n + \sum_{i=1}^{\infty} \frac{m_i}{2^i}, \text{ where } m_i = 0 \text{ or } m_i = 1$$

in the sense that the interval being built up by tacking smaller and smaller intervals onto $[0, n]$ is getting as close as you like to including p. The real number corresponding to p is therefore the limit of the sequence of partial sums, where the coefficients m_i are those in "decimal" part of its binary expansion.[19]

The construction establishes that any point p, $0 < p$, determines a sequence of 0s and 1s that we may take as the "decimal" part of a binary expansion of the real number corresponding to p.

What we need for the diagonal argument, however, is the converse of this: every sequence of 0s and 1s determines a point in $[n, n+1]$ via this construction. As above, AP will assure that $\cap_i I_i$ doesn't contain *more* than one point. But to assure that there is *at least* one, we must make an additional assumption about the real line, the easiest in this connection being a version of:

Cantor's Principle (CP): Any nested sequence of (closed) intervals on the real line has a non-empty intersection.[20]

Together, AP and CP assure that the "diagonal number" generated by diagonalization from the assumed sequence of real numbers is a real number.

It is worth noting that the non-denumerability of the reals in [0,1] may be established in a way that appeals explicitly to CP: Given any sequence $\{r_i\}$ of reals in [0,1], we may construct a nested sequence $\{I_i\}$ of (closed) intervals whose intersection does not contain any of the r_i. Let I_0 be some (closed) sub-interval of [0,1] that does not contain r_0, and let I_{i+1} be some (closed) sub-interval of I_i that does not contain r_{i+1} (I_{i+1} could be I_i). Evidently, I_i contains none of the r_j for $0 \leq j \leq i$, so the intersection of the I_i contains none of the reals r_i. By CP, however, this intersection is non-empty. So there is a real in [0,1] that is not among the r_i. Accordingly, $\{r_i\}$ is not an enumeration of the reals in [0,1] and no such enumeration is possible.

[19] This construction will give us the expansion 0.01111... rather than 0.10000... for 1/2, and similarly for other points of the form $n + m/2^i$.

[20] The usual statement is that any sequence $\{I_i\}$ of nested (closed) intervals whose lengths tend, in the limit, to 0 converges to a single point.

Conceived as stepwise construction of a "diagonal number", diagonalization is a special case of this process. In changing 0s to 1s and 1s to 0s on the diagonal as we proceed down the list of reals, we are essentially selecting sub-intervals of [0,1] that exclude the reals listed so far. For example, if the first expansions in the list are those for 1/2 (0.01111...and 0.10000...) the diagonal number begins 0.11, which corresponds to setting $I_0 = [1/2, 1]$ and $I_1 = [3/4, 1]$; the first of these does not exclude 1/2 but the second does.

9 Cuts

Before returning to Wittgenstein, let look at CP and AP from a slightly different angle. Imagine the line divided somewhere into points L to the left and points R to the right. Such a division is called a *cut*. What CP and AP tell us about these cuts is that "somewhere" is always *at a point*: there is a rightmost point in L or a leftmost point in R.[21] Given a cut, we may apply the construction procedure outlined above to obtain a sequence of nested intervals $[a_i, b_i]$ such that $a_i \in L$ and $b_i \in R$ and $b_i - a_i = 1/2^i$. By CP and AP, the intersection of this sequence consists of a single point p, which must belong to L or to R. If p belongs to L, it is the rightmost point of L: if there were a point q to the right of it, p could not be in I_i once $1/2_i < q - p$. Similarly, if p belongs to R it is the leftmost point of R.

CP and AP thus exclude what are termed *open* cuts: cuts in which there is no rightmost point in L *and* no leftmost point in R. If open cuts are allowed on the line, the points that, from the perspective of CP and AP, *close* them are missing: in an important sense the line has gaps. If the real numbers correspond to points on this gappy line, then the real numbers have gaps as well. For example, in this situation we cannot be sure that a continuous real-valued function on [0,1] that is positive at 0 and negative at 1 is zero somewhere in [0,1]!

CP and AP thus assure that the line is a *continuum*, and corresponding principles for the real numbers assure that they also form a continuum. There are other 'continuity' principles to the same effect, most notably:

DEDEKIND'S PRINCIPLE: No cut in the line is open.[22]

Like the assumptions of Euclidean geometry, which got us part way to a view of what numbers are needed, if we wish them to be the ratios of the length of line segments to that of a unit segment, these continuity principles are assumed in the theory of real numbers (or taken to be constraints on acceptable constructions of the real numbers from other mathematical notions, such as the rational

[21] Familiar geometric assumptions assure that we don't have *both*, for any segment has a midpoint.

[22] His own informal statement is: "If all the points of the straight line fall into two classes such that every point of the first class lies to the left of every point of the second class, then there exists one and only one point which produces this division into two classes, this severing of the straight line into two portions." Richard Dedekind, "Irrational numbers", in James R. Newman, ed., *The World of Mathematics* (New York; Simon & Schuster, 1956), 528–36 at 529 (Vol. I)

numbers). Of his own principle, Dedekind remarks:

> ...I am utterly unable to adduce any proof of its correctness, nor has any one the power. The assumption of this property of the line is nothing else than an axiom by which we attribute to the line its continuity, by which we find continuity in the line. If space has at all a real existence it is *not* necessary for it to be continuous; many of its properties would remain the same even were it discontinuous. And if we knew for certain that space was discontinuous there would be nothing to prevent us, in case we so desired, from filling up its gaps, in thought, and thus making it continuous; this filling up would consist in a creation of new point-individuals and would have to be effected in accordance with the above principle.[23]

10 Wittgenstein's response

I have suggested that Wittgenstein's objection to diagonalization in Cantor's proof that the reals are non-denumerable is that there isn't, in his view, an acceptable background theory of real numbers in terms of which we may judge that the "diagonal number" is indeed real. Hence, his odd suggestion that, at best, diagonalization tells us something about how we are to use phrases like "real number" or "series of real numbers". In the preceding two sections, I have presented a bit of the theory of real numbers and its motivation, in terms of which the "diagonal number" is clearly real. So what's not to like about it?

I'm don't believe I can give a good account of Wittgenstein's position here. He seems to think that people like Dedekind and Cantor are committed to the view that the subject of mathematics is some external mathematical reality, rather than mathematical conceptions, which may or may not apply to anything. He claims that:

> 19. The dangerous, deceptive thing about the idea: "The real numbers cannot be arranged in a series", or again "The set ...is not denumerable" is that it makes the determination of a concept—concept formation—look like a fact of nature.[24]

whereas

> 168. The mathematician is an inventor, not a discoverer.[25]

Wittgenstein doesn't explain what's crazy about regarding mathematics as the science of mathematical reality. Be that as it may, it's hard to see that Dedekind, at least, is committed to such a view. For him, the real line and its structure of points are objects of thought, things we dream up. The nature of the real

[23] Id. at 530.
[24] Wittgenstein, op. cit. at 131 (Part II).
[25] Id. at 99 (Part I).

line is a matter of stipulation; it is determined by various assumptions that *we* make.

Where Wittgenstein and Dedekind may part company is whether pictures and diagrams supply any content to these assumptions. Wittgenstein's view seems to be that they do not: assumptions are linguistic expressions, and whatever content they have is *exhausted* by rules for manipulating them. Mathematical practice consists essentially of symbol-mongering, of rule-governed manipulations of expressions in calculations or proofs. Although mathematical sentences may, in virtue of the misleading forms assigned them by logicians, look like ordinary sentences, their truth or falsity is wholly a matter of derivability or provability. There are no mathematical objects, properties and relations, even in thought, to which the singular terms and predicates of mathematics refer and which determine, in the way described by logical semantics, the truth or falsity of mathematical claims.

Here are some passages that suggest this view:

> 38. ... A 'series' in the mathematical sense is a method of construction for series of linguistic expressions.[26]

If so, the sequences mentioned in CANTOR'S PRINCIPLE will have to issue from explicit instructions about how to get from a notation for I_i to one for I_{i+1}—instructions that specify what I_{i+1} is, given I_i. This limits the points generated by CP to those for which we can devise such instructions, certainly far fewer than Cantor imagines are available.

> 10. ... One learns the meaning of "all" by learning that 'fa' follows from '$(x).fx$'.
>
> 11. ... the meaning of '$(x).fx$' is made clear by our insisting on 'fa' 's following from it.[27]

If so, it's hard to make sense of non-denumerability, since the meaning of "all" is going always to be limited by the expressions available for 'naming' individuals.

> 8. ... Just as we ask: " 'provable' in what system?", so we must also ask: " 'true' in what system?" 'True in Russell's system means ... : proved in Russell's system; and 'false in Russell's system' means: the opposite has been proved in Russell's system. ...[28]

If so, truth = proof and there can be no unprovable truths, contrary to the usual interpretation of Gödel's incompleteness theorem.

As indicated, these passages point to constraints on mathematical concepts, if their content is given by the rules of some mathematical 'language game'. People like Cantor and Dedekind would, I believe, respond by rejecting the view that the content of mathematical concepts is wholly given by such rules. And I believe they would be joined by most contemporary logicians. The standard

[26] Id. at 136 (Part II).
[27] Id. at 41-2 (Part I)
[28] Id. at 118 (Part III).

view of the Incompleteness Theorem and other limitive results in logic (the Löweheim–Skolem Theorem, the fact that you can't axiomatize identity, etc.) is not that our mathematical conceptions are limited, but that you can't wholly capture them in a net of language. Truth cannot be reduced to proof, that minimaly rich mathematical concepts have content that eludes axiomatization.

One place—are there others?—we might look for this extra-linguistic content is to pictures and diagrams, which do seem to have the power to make things clear. A picture or diagram in mathematics is indeed frequently worth a thousand words. I don't know how this works, but it seems to me undeniable that pictures can make mathematical concepts clear. It may be true that such pictures can be taken in various ways, so that their content is, from a critical perspective, ambiguous. But this suggests to me not that we don't get important information from pictures but that we don't yet understand how this happens. For in general it seems to me that we do not typically puzzle over the meaning of a picture or diagram in mathematics. On the contrary, we immediately grasp what is going on.

Wittgenstein seems to have no sympathy for this position, as can perhaps be seen from the following remarks (which, alas I do not have time to say more about):

> 32. The picture of the number line is an absolutely natural one up to a certain point; that is to say so long as it is not used for a general theory of real numbers.[29]
>
> 33. ...The proof of Dedekind's Theorem works with a picture which cannot justify *it*; which ought rather to be justified by the theorem.[30]
>
> 37. The misleading thing about Dedekind's conception is the idea that the real numbers are there spread out in the number line. they may be known or not; that does not matter. And in this way all that one needs to do is to cut or divide into classes, and one has dealt with them all.
>
> It is by *combining calculation and construction* that one gets the idea that there must be a point left out on the straight line, namely P, [here appears a diagram, like the one above, but with "P" in place of "$\sqrt{2}$"] if one does not admit $\sqrt{2}$ as a measure of distance from O. 'For, if I were to construct really accurately, then the circle would have to cut the straight line *between* its points.'
>
> This is a frightfully confusing picture.
>
> The irrational numbers are—so to speak—special cases.

[29] Op. cit. at 286 (Part V).

[30] Id. at 287. Wittgenstein is referring here to Dedekind's identification of real numbers with cuts in the rationals, which we may visualize as a division of the rational points on the line into those to the left and those to the right. If we identify such a cut with its left set, we may order the reals by $L < L'$ iff there is a rational in L' that is not in L, and then establish that any cut in the *reals* is made by some real number. This is presumably what is meant by "Dedekind's Theorem".

> What is the *application* of the concept of a straight line in which a point is missing?! The application must be 'common or garden'. The expression "straight line with a point missing" is a fearfully misleading picture. The yawning gulf between illustration and application.[31]

(To be continued.)

Appendix

The recursive functions are numerical functions obtainable from simple *initial* functions by a finite number of applications of simple *operations*. The initial functions are the constant zero function ($Z(x) = 0$), the successor function ($S(x) = x + 1$), and the index functions ($\pi_i^n(x_1, \ldots, x_n) = x_i$). The operations are:

1. Composition: h is obtained from f, g_1,…,g_n by composition if

$$h(x_1, \ldots, x_m) = f(g_1(x_1, \ldots, x_m), \ldots, g_n((x_1, \ldots, x_m))$$

2. Recursion: h is obtained from f and g by recursion if

$$\begin{aligned} h(x_1, \ldots, x_m, 0) &= g(x_1, \ldots, x_m) \\ h(x_1, \ldots, x_m, y+1) &= f(x_1, \ldots, x_m, y, h(x_1, \ldots, x_m)) \end{aligned}$$

3. μ-operation: h is obtained from f by the μ-operation if

$$\begin{aligned} h(x_1, \ldots, x_n) &= \mu y(f(x_1, \ldots, x_n, y) = 0) \\ &= \text{the least } y \text{ such that } f(x_1, \ldots, x_n, y) = 0 \end{aligned}$$

provided $\forall x_1 \ldots \forall x_n \exists y(f(x_1, \ldots, x_n, y) = 0)$.

[31] Id. at 290-9.