

Research statement

Jason Teutsch
teutsch@math.uni-heidelberg.de

November 4, 2011

Summary

My research efforts to date consist of a mixture of pure and applied topics involving mathematics and computer science. I am especially interested in the power of language and the absolute limitations of machines. I have worked in computability theory, the branch of mathematics comprising Gödel's incompleteness theorem, which examines self-reference with respect to formal expression. My work in mathematical logic and theoretical computer science has touched on probability theory and martingales, additive number theory, classical fractal geometry, Gödel numberings (that is, programming languages), Kolmogorov complexity, algorithmic randomness, shortest computer programs, and recursively approximable real numbers. I have also pursued a number of applied topics information theory, including error-correcting codes, efficient data compression, cryptography, network security, computational neuroscience, and (unrelated to information theory) disease modeling.

1. Overview and background

Over the past decade, researchers have precisely pinned down what it means for one object to exhibit more randomness than another [13, 43]. To illustrate this point, consider the following example. One of the following binary sequences was generated using coin flips:

```
01010101010101010101010101010101010101010101...  
00100111010100100011110111010110010110000100...
```

The other was not. Intuition suggests that the second sequence came from coin flips, but how? Both sequences have the same number of 0's and 1's (up to the point shown), and indeed both outcomes result with equal likelihood from a series of coin flips. Hence the scope of the question, "Is sequence X random?" lies outside the realm of classical probability theory.

Had history run a different course, perhaps computability theory would have formalized the notion of chance instead of measure theory. In this exposition, we shall focus our attention on randomness for infinite binary sequences on the following two paradigms¹:

¹A third paradigm is commonly used as well [13, Section 6.2].

- (I) A sequence is random if it is incompressible: its prefixes have no simple pattern recognizable to a computer.
- (II) A sequence is random if it is unpredictable: a computable gambler cannot become rich by betting on it.

Together with the formal definitions of the next two paragraphs, these concepts define the field of *algorithmic information theory* and *algorithmic randomness*. While we cannot expect to obtain a single, definitive notion of random sequence, these formal concepts have proven to be particularly robust [13, 43].

First we treat paradigm (I). A set of finite strings is called *prefix-free* if no element of the set is a proper prefix of another, and a Turing machine is *prefix-free* if its domain is a prefix-free set. For prefix-free Turing machines A and finite binary strings σ and τ , let $K_A(\sigma) = \min\{|\tau| : A(\tau) = \sigma\}$ be the length of the shortest program which computes σ . There exists a *universal* prefix-free machine U which is optimal in the sense that for some constant c and for all σ , $K_U(\sigma) \leq K_A(\sigma) + c$ [35, Theorem 3.1.1]. Fix a universal prefix-free machine U and define $K(\sigma) = K_U(\sigma)$ to be *the prefix-free Kolmogorov complexity of σ* . Now define, according to (I) above, a binary sequence X to be *Martin-Löf random* if there exists a constant c such that $K(X \upharpoonright n) \geq n - c$ for all n , where $X \upharpoonright n$ denotes the length n prefix of X [32, 39]. One might wonder why we have chosen to require “prefix-free” in the above set of definitions, as opposed to discussing *plain* Kolmogorov complexity. The reason is that no Martin-Löf random sequences would exist if we dropped this condition [13, Theorem 3.1.4]. Moreover, prefix-free complexity makes sense as a measure of information content in that one cannot encode “extra” information into the length of input strings [13].

Now let us inspect paradigm (II). Any function M from finite binary strings to nonnegative reals which satisfies the *fairness condition*

$$M(\sigma) = \frac{M(\sigma 0) + M(\sigma 1)}{2}$$

for all strings σ is called a *martingale*. $M(\sigma)$ corresponds to the gambler’s capital after having already bet on the finite string σ . The fairness condition says that the amount of money gained from an outcome of “0” is the same that would be lost from an outcome of “1.” A martingale M *succeeds* on a binary sequence X if M achieves arbitrary sums of money over X , that is, $\limsup_n M(X \upharpoonright n) = \infty$; otherwise X *defeats* M . M *Schnorr-succeeds* on a sequence X if M succeeds on X and there exists a computable, unbounded, nondecreasing function f such that $f(n) < M(X \upharpoonright n)$ for infinitely many n . If we can replace “for infinitely many n ” with “for all n ,” then M *Kurtz-succeeds* on X . We can now define three classical randomness notions in terms of martingales. A sequence X is called *computably random* [51, 52] if X defeats every martingale. If no martingale Schnorr-succeeds on X , then X is *Schnorr random* [52], and if no martingale Kurtz-succeeds on X , then X is *Kurtz random* [12, 29, 65].

In order to explain the interaction between algorithmic information theory and programming languages, we introduce the following concepts. A *numbering* φ is a partial-recursive (p.r.) function $\langle e, x \rangle \mapsto \varphi_e(x)$; φ is an *acceptable* or *Gödel numbering* if for every further numbering ψ there exists a recursive function f

such that $\varphi_{f(e)} = \psi_e$ for all e [44, 53]. If φ is acceptable and in addition for every numbering ψ the corresponding translation function f is linearly bounded, then φ is a *Kolmogorov* numbering [50]. Acceptable numberings capture the notion of reasonable programming language as any algorithm can be effectively coded into an acceptable numbering given a formal description for that algorithm. Kolmogorov numberings are optimal programming languages in the sense that they are universal machines for plain Kolmogorov complexity. Finally a set is called *recursively enumerable*, or *r.e.* if it is the domain of some p.r. function, and *co-r.e.* if it is the complement of an r.e. set [44, 53].

I outline my research program below in a series of self-contained sections. In Section 2, I describe a single theorem connecting algorithmic information theory to both classical analysis and algebra. In Section 3, I raise an objection to the classical notions of algorithmic randomness based on practical considerations and explore an alternative principle by which a casino might operate. Section 4 and Section 5 examine the entwined relationship between information content and the names used to describe random-like objects. I am interested in applications of algorithmic information theory, both within mathematics and outside, and describe in Section 6 some applied topics where I have made contributions.

2. Fractals by partial randomness

Just as one can use Hausdorff dimension² to identify fractal sets of intermediate, non-integer dimension, one can use *constructive* Hausdorff dimension³ to quantify the amount of “randomness” in a set. Unlike classical Hausdorff dimension, however, with constructive Hausdorff dimension we can meaningfully measure the randomness content of a singleton real. Lutz [37] showed that the constructive Hausdorff dimension of a set E is the supremum over the constructive Hausdorff dimension of singleton sets, or equivalently the *dimension* of the individual points, whose union is E . Moreover any real number $0 \leq X \leq 1$, viewed as a binary sequence expansion, has dimension $\liminf_{n \rightarrow \infty} K(X \upharpoonright n)/n$ so that recursive reals have dimension 0, Martin-Löf random reals have dimension 1, and an interleaving of the digits of a recursive real with a Martin-Löf random real gives dimension $1/2$ [13, Theorem 13.3.4], [31], [40].

The *Cantor set* \mathcal{C} [16, 66], defined as the set of reals in $[0, 1]$ with ternary expansions containing only 0’s and 2’s, is a quintessential fractal due to its sim-

²The formal definition of Hausdorff dimension is well-known. A *cover* \mathcal{G} for a set E is a collection of sets whose union contains E , and \mathcal{G} is a δ -*mesh* cover if the diameter of each member G is at most δ . For a number $\beta \geq 0$, the β -*dimensional Hausdorff measure* of E , written $\mathcal{H}^\beta(E)$, is given by $\lim_{\delta \rightarrow 0} \mathcal{H}_\delta^\beta(E)$ where

$$\mathcal{H}_\delta^\beta(E) = \inf \left\{ \sum_{G \in \mathcal{G}} |G|^\beta : \mathcal{G} \text{ is a countable } \delta\text{-mesh cover of } E \right\}.$$

The *Hausdorff dimension* of a set E is the unique number α where the α -dimensional Hausdorff measure of E transitions from being negligible to being infinitely large; if $\beta < \alpha$, then $\mathcal{H}^\beta(E) = \infty$ and if $\beta > \alpha$, then $\mathcal{H}^\beta(E) = 0$ [16, 66].

³The *constructive β -dimensional Hausdorff measure* is defined exactly in the same way as Hausdorff measure with the restriction that the covers be uniformly r.e. open sets [13, Definition 13.3.3]. This yields the corresponding notion of *constructive Hausdorff dimension*.

ple, recursive construction and self-similarity features. I showed together with Dougherty, Lutz, and Mauldin [11] that some points in the Cantor set “cancel” randomness in the sense that, when added to a Martin-Löf random real, the resulting sum has dimension less than the random itself. More specifically,

Theorem 2.1 (Dougherty, Lutz, Mauldin, Teutsch [11]). *For any α satisfying $1 - \log 2 / \log 3 \leq \alpha \leq 1$, and for any Martin-Löf random $r \in (0, 1)$, the set \mathcal{E}_α of points in $\mathcal{C} + r$ with dimension α has (classical) Hausdorff dimension $\alpha - 1 + \log 2 / \log 3$.*

From this result we obtain a simple relation between the effective and classical Hausdorff dimensions of \mathcal{E}_α ; the difference is exactly 1 minus the dimension of the Cantor set, or $1 - \log 2 / \log 3$. In particular, for $\alpha < 1$, \mathcal{E}_α is nonempty and contains elements with lower dimension than r . We conclude that many points in the Cantor set additively cancel randomness.

While the deepest results used in the proof of Theorem 2.1 belong to the subfield of classical analysis known as fractal geometry, the key algorithmic aspect of our proof hinges critically on an additive number theory result due to G. G. Lorentz:

Lorentz’s Lemma ([36], see also [15]). *There exists a constant c such that for any integer k , if $A \subseteq [0, k]$ is a set of integers with $|A| \geq \ell \geq 2$, then there exists a set of integers $B \subseteq (-k, k)$ such that $A + B \supseteq [0, k]$ with $|B| \leq ck \frac{\log \ell}{\ell}$.*

Using Lorentz’s Lemma, one can effectively obtain points which nearly optimally “cancel randomness” given a description of the translation real r .

3. Martingales and randomness in the real-world

According to the classical notions of algorithmic randomness due to Martin-Löf, Schnorr, and Kurtz, a martingale is permitted to wager any amount of money within his means [13, Section 6.3]. On the other hand, a real gambler who wagers \$0.001, \$0.0001, and then \$0.00001 would be a nuisance to any casino and most likely does not exist. For this reason, Bienvenu, Stephan, and I investigated the consequences of imposing a minimum wager value, as would occur in any modern-day casino, on the definition of “random sequence.” To this end, we put forth the following definition:

Definition 3.1 (Bienvenu, Stephan, Teutsch [2]). For a martingale M and a binary string σ , $|M(\sigma 0) - M(\sigma)|$ is called M ’s *wager at σ* . A martingale whose wagers are integers is called *integer-valued*, and a sequence which defeats every integer-valued martingale is *integer-valued random*.

We discovered that, curiously, integer-valued randomness is incomparable with Schnorr randomness and Kurtz randomness [2]. That is, there exists an integer-valued random sequence which is not Kurtz random and a Schnorr random sequence which is not integer-valued random⁴.

In addition to lower bounds, most casinos also enforce upper limits on betting. Together with Chalcraft, Dougherty, and Freiling [7], I investigated randomness

⁴Incomparability follows by noting that every Schnorr random sequence is Kurtz random.

notions for gamblers whose wagers are restricted to a finite set. Given a set V of nonnegative reals, we say that a martingale is V -valued if for all σ the wager of M at σ belongs to V , unless M does not have enough capital in which case the wager at σ is 0. Furthermore, a sequence is V -valued random if no V -valued martingale succeeds on it.

Theorem 3.2 (Chalcraft, Dougherty, Freiling, Teutsch [7]). *Let A and B be non-empty finite sets of computable real numbers. Then every A -valued random is B -valued random if and only if there exists a $k \geq 0$ such that $B \subseteq A \cdot k$.*

Using Theorem 3.2, one can easily find sets A and B so that A -valued randomness is incomparable with B -valued randomness. Let X be an A -valued random which is not B -valued random. One can imagine a scenario in which a casino puts forth a sequence X and a B -valued “idiot” wins by betting on X . The A -valued casino customers observe B ’s success and are thus lured into gambling, but try as they may X defeats every A -valued customer. In the case of time-bounded customers, the sequence X can even be computable.

Leonid Levin [33] pointed out a potential shortcoming of our restricted-wager gambling model. In the classical paradigms, where a gambler can bet any fraction of her capital, there is a well-known “savings trick” [13, p. 237] which allows a martingale to know when it is safe to take each of her dollars out of the casino and go shopping with them. As Levin feared and I proved [33, 58], there exists a sequence on which some integer-valued martingale succeeds but any integer-valued martingale who attempts to take his winnings outside the casino necessarily goes bankrupt. I continue to investigate extensions of this paradoxical result. Some other open questions remain in this area, for example:

1. Does Theorem 3.2 generalize to infinite sets? In particular, is integer-valued randomness the same as the randomness notion obtained when wagers consist of the set of reals bounded away from 0?
2. Are there meaningful characterizations of integer-valued randomness in terms other algorithmic models, for example the incompressibility paradigm?

4. Recursively approximable reals and sets

Some Martin-Löf random sequences, when viewed as binary expansions of real numbers, admit recursive approximations from below. Let A be a collection of binary strings. We define the *weight* of A to be the real number

$$\sum_{\sigma \in A} 2^{-|\sigma|}.$$

A real number is called *left-r.e.* if it is the weight of some r.e. set of strings. Chaitin [6], after work of Zvonkin and Levin [67], motivated the study of left-r.e. sets by showing that the weight of any universal prefix-free machine U is a Martin-Löf random real. Conversely, Calude, Hertling, Khoussainov, Wang [4] and Kučera, Slaman [26] showed that *every* left-r.e. Martin-Löf random real in $[0, 1]$ is the weight of some universal prefix-free machine’s domain.

Wolfgang Merkle and I [41] recently unearthed an alternate characterization of the left-r.e. Martin-Löf random reals. Given a prefix-free machine M , we say that a binary string σ is M -compressible if $K_M(\sigma) \leq |\sigma|$. Building on preliminary results by Tadaki [57], we proved that a real in $[0, 1]$ is a left-r.e. Martin-Löf random real if and only if it equals the weight of the U -compressible strings for some universal prefix-free machine U . We can extend this characterization to weights of strings which are *compressible by at least a and less than b bits*, that is, those σ satisfying

$$K_M(\sigma) + a \leq |\sigma| < K_M(\sigma) + b.$$

For sufficiently large finite intervals $[a, b)$, the strings which are compressible by a -bits but not by b -bits form a set which is neither r.e. nor co-r.e., hence we can obtain the left-r.e. Martin-Löf reals by methods other than the usual r.e. weight-based constructions.

Kjos-Hanssen, Stephan, and I [25] have explored the complex relationship between definability in formal languages and left-r.e. random sequences. In one direction, one might like to know which classes of left-r.e. random reals admit an effective naming scheme, or *left-r.e. numbering*, formally a p.r. function from natural numbers to \mathcal{C} mapping each e to the weight of the e^{th} r.e. set⁵. The answer to this question depends on the choice of randomness notion as the left-r.e. Martin-Löf random reals admit a left-r.e. numbering [3, 25] whereas the left-r.e. Kurtz randoms do not [25], and even the Martin-Löf random reals do not admit an acceptable left-r.e. numbering⁶ [25].

In the other direction, one might ask what left-r.e. reals have to say about the complexity of formal language. We characterize the third and fourth levels of the arithmetic hierarchy⁷ purely in terms of classical randomness notions. We prove a few dichotomy theorems, including the following:

Theorem 4.1 (Kjos-Hanssen, Stephan, Teutsch [25]). *For every acceptable left-r.e. numbering α and every $A \in \Pi_4^0$, there exists a computable function f such that for all e ,*

$$\begin{aligned} e \in A &\implies \alpha_{f(e)} \text{ is computably random;} \\ e \notin A &\implies \alpha_{f(e)} \text{ is not Schnorr random.} \end{aligned}$$

Combining our dichotomy theorems with Theorem 4.2 below, one may obtain a complete picture describing which classical classes of left-r.e. random reals and left-r.e. nonrandom reals admit left-r.e. numberings. We would like to expand the vocabulary to include other recursively approximable sets which are not necessarily left-r.e., a goal which leads us to the following numberings-free question:

⁵For technical reasons which we elaborate in our paper [25, Section 2], it is essential, in any detailed discussion of left-r.e. numberings to view numberings in terms of sets rather than reals. In order to obtain a quick, intuitive picture, however, we appeal to a concise definition for reals.

⁶We define “acceptable left-r.e. numbering” analogous to “acceptable numbering” from Section 1.

⁷For Theorem 4.1, recall that a Π_4^0 set A is one which can be written in the form

$$x \in A \iff (\forall u) (\exists v) (\forall y) (\exists z) : P(u, v, x, y, z)$$

for some recursive predicate P [53].

if $W = \{w_0 < w_1 < w_2, \dots\}$ is an r.e. set and Ω is a left-r.e. Martin-Löf random, is

$$\Omega(w_0)\Omega(w_1)\Omega(w_2)\dots$$

necessarily Martin-Löf random?

The versatility of left-r.e. numberings over the usual (r.e.) numberings makes them interesting to study in their own right, even without ties to randomness. In the world of left-r.e. numberings, existence of numberings is equivalent to arithmetic complexity in the sense of the following theorem:

Theorem 4.2 (Kjos-Hanssen, Stephan, Teutsch [25]). *Let \mathcal{C} be a class of infinite left-r.e. reals which contains a shift-persistent element⁸. Then for any left-r.e. numbering α of the left-r.e. reals, the following are equivalent:*

- (I) $\{e : \alpha_e \in \mathcal{C}\}$ is a Σ_3^0 -set⁹.
- (II) There exists a left-r.e. numbering of \mathcal{C} .
- (III) There exists an effective enumeration of \mathcal{C} without repetitions.

In another direction, Frank Stephan and I have studied the self-referential properties of left-r.e. numberings [55]. We say that a left-r.e. set A (defined appropriately) *can be made into itself* if there exists a left-r.e. numbering α of the left-r.e. reals such that $\{e : \alpha_e = A\} = A$, and similarly a class of sets *can be made into itself* if there exists a left-r.e. numbering β of the left-r.e. sets such that $\{e : \beta_e \in \mathcal{C}\} \in \mathcal{C}$. Examples of our results include that the Martin-Löf random reals can be made into themselves, but the non-recursive left-r.e. sets cannot. In another project, Stephan, Jain, and I have examined sets which reduce only to left-r.e. sets under strong recursion-theoretic reductions [22]. Our work suggests a possible connection between the strength of such reductions and initial segment complexity for such downward closed sets.

5. The set of shortest programs

The study of minimal indices is motivated philosophically by Occam's razor, a principle which says that the simplest solution is the correct one. The *set of shortest programs* is

$$\text{MIN}_\varphi = \{e : (\forall j < e) [\varphi_j \neq \varphi_e]\},$$

where φ is an acceptable numbering. Strictly speaking, MIN_φ consists of the minimal indices for each p.r. function. The set of shortest programs has applications to computational learning theory [1, 8, 20, 21], and also serves as a convenient counterexample in recursion theory [19, 23, 24, 38, 49, 56, 54, 60, 59]. It is perhaps not too surprising that properties minimal indices, which are themselves names for programming languages, interact closely with numberings. For

⁸*Shift-persistent* means that finite prefixes can be added without leaving the class, which would actually make sense if we were discussing sequences rather than reals.

⁹A set is Σ_3^0 , in the usual sense from computability theory, if it is definable using three alternating quantifiers followed by a recursive predicate with the first quantifier being existential.

example, Stephan and I showed [56] that whether or not the set of shortest descriptions contains a co-r.e. set depends on the underlying acceptable numbering. We shall focus on acceptable numberings because in the general case every index could be minimal which is not too interesting [49].

I describe the two classical complexity problems regarding minimal indices and the progress I have made towards their solution. We appeal to some notions from recursion theory, where reductions can informally be interpreted as comparing information content. The symbol $'$ denotes the jump operator¹⁰ and \leq_T denotes Turing reduction¹¹. Stronger yet, A is *truth-table* reducible¹² to B , written $A \leq_{tt} B$, if membership in A can be decided by effectively constructing a Boolean formula and evaluating it using membership values from B . If the truth-table f can be chosen so as to depend on only a constant number of membership queries from B , then a stronger reduction called *bounded truth-table*, is satisfied. A set is *complete* if it has the same degree as the halting problem, and *incomplete* otherwise. See [44] and [53] for further background on recursion theory.

In 1972, Albert Meyer posed the following question:

Meyer's Problem ([42]). *Is $\text{MIN}_\varphi \equiv_{tt} \emptyset''$ for every acceptable numbering φ ?*

Meyer immediately proved that $\text{MIN}_\varphi \equiv_T \emptyset''$ for every acceptable numbering φ . Five years later, Kinber [24] showed that the bounded truth-table degree of MIN_φ depends on the underlying acceptable numbering φ , and Schaefer [49] gave an acceptable numbering ψ which makes MIN_ψ truth-table complete, yet 41 years since its introduction, Meyer's Problem remains unresolved.

In 1998, Marcus Schaefer [49] introduced the *set of shortest descriptions*:

$$\text{SD}_\varphi = \{e : (\forall j < e) [\varphi_j(0) \neq \varphi_e(0)]\},$$

a sort of “one-dimensional” version of the set of shortest programs. A couple of years earlier, Martin Kummer [28] had shown that the similar-looking set of Kolmogorov random strings was truth-table complete under every Kolmogorov numbering (though not so for acceptable numberings), and Schaefer wondered whether a similar construction might yield the truth-table degree of SD_φ . Frank Stephan and I [54] recently showed that the truth-table degree of the set of shortest descriptions depends on the underlying acceptable numbering, although our techniques do not carry over to Meyer's Problem. Our construction makes a particularly concrete connection between shortest descriptions, which are themselves incompressible objects, and algorithmic randomness: our truth-table incomplete set of shortest descriptions has the truth-table degree of a Martin-Löf random sequence. We also showed that the truth-table of the *set of domain-random strings*, the complement of

$$\text{NRW}_\varphi = \{x : (\exists j < x) [\max(W_j^\varphi \cup \{0\}) = x]\},$$

¹⁰This makes \emptyset' is the halting set, \emptyset'' is the halting set for the halting set oracle, and so on.

¹¹For sets A and B , $A \leq_T B$ if A can be computed using B as an oracle.

¹²More formally, $A \leq_{tt} B$, if there exist recursive functions f and g such that for all x ,

$$x \in A \iff f[x, B(0), B(1), \dots, B(g(x))] = 1.$$

and closer cousin of the Kolmogorov random strings, depends on its underlying acceptable numbering. We extend our result on the set of domain-random strings to obtain chains and antichains of bounded truth-table degrees inside of various r.e. truth-table degrees. Some degree related problems remain. For example, the truth-table degree of SD_φ and NRW_φ for Kolmogorov numberings φ and the Turing degree of NRW_ψ for acceptable numberings ψ are unknown.

John Case introduced [5] the following variant of the set of shortest programs:

$$\text{MIN}^* = \{e : (\forall j < e) [\varphi_j =^* \varphi_e]\}$$

where $=^*$ denotes equality except for a finite set. Schaefer [49] showed that MIN_φ^* joined with the halting set has the Turing degree of \emptyset''' for every acceptable φ . This begs the question of whether, in fact, MIN_φ^* has the Turing degree of \emptyset''' , or equivalently,

Schaefer's Problem ([49]). *Is $\text{MIN}_\varphi^* \geq_T \emptyset'$ for every acceptable numbering φ ?*

Using the Owings Cardinality Theorem [27, 47], Jain, Stephan, and I showed that $\text{MIN}_\varphi^* \equiv_T \emptyset'''$ for every Kolmogorov numbering φ , thus solving Schaefer's Problem for the case of standard universal machines. We also show that Schaefer's result, $\text{MIN}_\varphi^* \oplus \emptyset' \equiv_T \emptyset'''$, carries over to $=^*$ -minimal indices of r.e. sets for any *arbitrary* numbering φ of the p.r. functions. Our result is optimal. Our paper [23] also continues the investigation begun in my earlier papers, [56, 59, 60], which examines minimal indices for r.e. sets in place of functions; we identify \emptyset''' (and higher [59]) with sets of minimal indices [23]. Schaefer's original problem remains open.

Having noted that arbitrary numberings preserve some properties of minimal indices, we wondered whether the structure of the familiar index sets¹³ from recursion theory, a classical topic in the case of acceptable numberings, carry over to arbitrary numberings. We discovered a disturbing numbering which makes the *index set for equality*, the pairs $\langle i, j \rangle$ such that the i^{th} r.e. set equals the j^{th} one, to have 1-generic¹⁴ Turing degree [23]. In other cases, such as the index set for pairs with disjoint domains, the Turing degree can remain fixed across all numberings. We wonder whether the *inclusion problem*, the pairs $\langle i, j \rangle$ such that the i^{th} r.e. set contains the j^{th} r.e. set, has a fixed Turing degree of \emptyset'' in every numbering.

6. Applied topics in information theory

The majority of my applied research pertains to information theory. I have worked on practical projects in cryptography, network security, data correction, error-correcting codes, and health sciences. I am particularly interested in applications of information theory to computational neuroscience. Below I describe some details of my work in these areas.

¹³Examples of index sets include the indices for the total functions, the indices for functions with recursive domain, etc. [53]

¹⁴A set X is *1-generic* if for every r.e. set of strings A , either X has a prefix in A or there exists a prefix of X which has no extension in A .

Security. I worked on classified cryptography and network security projects as a research staff member at Center for Communications Research–La Jolla (2008–2010). Specific projects have involved data mining, signal processing, analysis of large data sets, and algorithm analysis.

Signal processing. When Lempel and Ziv [30] introduced their famous complexity measure (LZ), which has evolved into one of the most widely used data compression algorithms today, they cited the de Bruijn sequences as examples of strings with high LZ-complexity. *T-complexity* [61] is an alternative complexity measure which also gives rise to efficient compression, however finding strings which are maximal for T-complexity is less straightforward, and this difficulty may have led to the propagation of certain unlikely propositions:

It has . . . been conjectured that T-complexity represents a good approximation to Kolmogorov complexity. [14, p. 2]

Recently Greg Clark and I [10] found an efficient way to generate strings which have maximum T-complexity for their lengths and applied our methods to disprove the above conjecture. We also investigated the T-complexity of de Bruijn sequences, and found that for some lengths de Bruijn sequences can achieve maximal T-complexity and for other lengths not. In general, the T-complexity of a de Bruijn sequence does not lie within a logarithmic factor of maximal for its length.

I have also worked with error-correcting codes. Feldman, Wainwright, and Karger [17, 18] used linear programming to decode messages sent over noisy channels. Their constraint relaxation method results in the creation of pseudo-codewords which must be avoided for successful decoding. Smarandache, Vontobel, Kiyavash, Vukobratovic, and I [64] investigated the pseudo-codewords arising from finite geometry codes and introduced methods to analytically quantify their behavior.

Neuroscience. Recently, Onton and Makeig [45] succeeded in distinguishing among emotional states in human subjects by measuring electroencephalographic activity (EEG). Information theory provides some tools, such as normalized information distance [9, 34, 63], which may facilitate improvement of the independent component analysis (ICA) [46, 48, 62] for these signals. I have a pending grant to study ICA and epilepsy with Scott Makeig (UCSD) and Mina Teicher (Bar-Ilan).

Health. As a postdoctoral fellow at the RAND Corporation, I developed a Markov model with Vargas, Keeler, Weden, Wu, and Zhuo which simulated the progression of chronic kidney disease. The states of our model corresponded to five stages of disease. I estimated transition probabilities for a cohort moving from one stage to the next, given an underlying condition such as hypertension or diabetes, based on available studies from the medical literature.

References

- [1] Andris Ambainis, John Case, Sanjay Jain, and Mandayam Suraj. Parsimony hierarchies for inductive inference. *Journal of Symbolic Logic*, 69(1):287–327, March 2004.
- [2] Laurent Bienvenu, Frank Stephan, and Jason Teutsch. How powerful are integer-valued martingales? (to appear in *Theory of Computing Systems*).
- [3] Paul Brodhead and Bjørn Kjos-Hanssen. Numberings and randomness. In *Mathematical Theory and Computational Practice*, volume 5635 of *Lecture Notes in Computer Science*, pages 49–58, Berlin, Heidelberg, 2009. Springer-Verlag.
- [4] Cristian S. Calude, Peter H. Hertling, Bakhadyr Khoussainov, and Yongge Wang. Recursively enumerable reals and Chaitin Ω numbers. *Theoretical Computer Science*, 255(1-2):125–149, 2001.
- [5] John Case. Homework assignment for students. Computer and Information Sciences Department, University of Delaware, 1990.
- [6] Gregory J. Chaitin. Incompleteness theorems for random reals. *Advances in Applied Mathematics*, 8(2):119–146, 1987.
- [7] Adam Chalcraft, Randall Dougherty, Chris Freiling, and Jason Teutsch. How to build a warped online casino. (submitted).
- [8] Keh-Jiann Chen. Tradeoffs in the inductive inference of nearly minimal size programs. *Information and Control*, 52(1):68–86, 1982.
- [9] Rudi Cilibrasi and Paul Vitányi. Clustering by compression. *IEEE Transactions on Information Theory*, 41(4):1523–1545, April 2005.
- [10] Greg Clark and Jason Teutsch. Maximizing T-complexity. (submitted).
- [11] Randall Dougherty, Jack Lutz, Daniel Mauldin, and Jason Teutsch. Translating the Cantor set by a random. (submitted).
- [12] Rodney G. Downey, Evan J. Griffiths, and Stephanie Reid. On Kurtz randomness. *Theoretical Computer Science*, 321(2-3):249–270, 2004.
- [13] Rodney G. Downey and Denis R. Hirschfeldt. *Algorithmic randomness and complexity*. Theory and Applications of Computability. Springer, New York, 2010.
- [14] Raimund Eimann, Ulrich Speidel, Nevil Brownlee, and Jia Yang. Network event detection with T-entropy. *CDMTCS Research Report Series*, 266, 2005.
- [15] P. Erdős, K. Kunen, and R. Daniel Mauldin. Some additive properties of sets of real numbers. *Polska Akademia Nauk. Fundamenta Mathematicae*, 113(3):187–199, 1981.

- [16] Kenneth Falconer. *Fractal geometry*. John Wiley & Sons Inc., Hoboken, NJ, second edition, 2003. Mathematical foundations and applications.
- [17] Jon Feldman. *Decoding error-correcting codes via linear programming*. PhD thesis, Massachusetts Institute of Technology, 2003.
- [18] Jon Feldman, Martin J. Wainwright, and David R. Karger. Using linear programming to decode binary linear codes. *IEEE Transactions on Information Theory*, 51(3):954–972, March 2005.
- [19] Stephen Fenner and Marcus Schaefer. Bounded immunity and btt-reductions. *MLQ Math. Log. Q.*, 45(1):3–21, 1999.
- [20] Rusins Frievālds. Inductive inference of minimal programs. In *Proceedings of the third annual workshop on Computational learning theory, COLT '90*, pages 3–22, San Francisco, CA, USA, 1990. Morgan Kaufmann Publishers Inc.
- [21] Sanjay Jain, Efim Kinber, and Rolf Wiehagen. On learning and co-learning of minimal programs. In Setsuo Arikawa and Arun Sharma, editors, *Algorithmic Learning Theory*, volume 1160 of *Lecture Notes in Computer Science*, pages 242–255. Springer Berlin / Heidelberg, 1996.
- [22] Sanjay Jain, Frank Stephan, and Jason Teutsch. Closed left-r.e. sets. In *Theory and Applications of Models of Computation (TAMC 2011)*, volume 6648 of *Lecture Notes in Computer Science*, pages 218–229. Springer Berlin / Heidelberg, 2011.
- [23] Sanjay Jain, Frank Stephan, and Jason Teutsch. Index sets and universal numberings. *Journal of Computer and System Sciences*, 77(4):760–773, 2011.
- [24] Jefim Kinber. On btt-degrees of sets of minimal numbers in Gödel numberings. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 23(3):201–212, 1977.
- [25] Bjørn Kjos-Hanssen, Frank Stephan, and Jason Teutsch. Arithmetic complexity via effective names for random sequences. (to appear in *ACM Transactions on Computational Logic*).
- [26] Antonín Kučera and Theodore A. Slaman. Randomness and recursive enumerability. *SIAM Journal on Computing*, 31(1):199–211 (electronic), 2001.
- [27] Martin Kummer. A proof of Beigel’s cardinality conjecture. *The Journal of Symbolic Logic*, 57(2):677–681, June 1992.
- [28] Martin Kummer. On the complexity of random strings (extended abstract). In *STACS 96 (Grenoble, 1996)*, volume 1046 of *Lecture Notes in Comput. Sci.*, pages 25–36. Springer, Berlin, 1996.
- [29] Stuart Kurtz. *Randomness and Genericity in the Degrees of Unsolvability*. PhD thesis, University of Illinois at Urbana, 1981.

- [30] Abraham Lempel and Jacob Ziv. On the complexity of finite sequences. *Institute of Electrical and Electronics Engineers. Transactions on Information Theory*, IT-22(1):75–81, 1976.
- [31] L. A. Levin. The concept of a random sequence. *Doklady Akademii Nauk SSSR*, 212:548–550, 1973.
- [32] L. A. Levin. Laws of information conservation (nongrowth) and aspects of the foundation of probability theory. *Problems of Information Transmission*, 10(3):206–210, 1974.
- [33] Leonid Levin. Personal communication.
- [34] Ming Li, Xin Chen, Xin Li, Bin Ma, and Paul M. B. Vitányi. The similarity metric. *IEEE Transactions on Information Theory*, 50(12):3250–3264, December 2004.
- [35] Ming Li and Paul Vitányi. *An introduction to Kolmogorov complexity and its applications*. Texts in Computer Science. Springer, New York, third edition, 2008.
- [36] G. G. Lorentz. On a problem of additive number theory. *Proceedings of the American Mathematical Society*, 5:838–841, 1954.
- [37] Jack H. Lutz. The dimensions of individual strings and sequences. *Information and Computation*, 187(1):49–79, 2003.
- [38] G. Marandžjan. On the sets of minimal indices of partial recursive functions. In Jir Becvr, editor, *Mathematical Foundations of Computer Science 1979*, volume 74 of *Lecture Notes in Computer Science*, pages 372–374. Springer Berlin / Heidelberg, 1979.
- [39] Per Martin-Löf. The definition of random sequences. *Information and Control*, 9:602–619, 1966.
- [40] Elvira Mayordomo. A Kolmogorov complexity characterization of constructive Hausdorff dimension. *Information Processing Letters*, 84(1):1–3, 2002.
- [41] Wolfgang Merkle and Jason Teutsch. Weights of deeply compressible strings. (submitted).
- [42] Albert R. Meyer. Program size in restricted programming languages. *Information and Control*, 21:382–394, 1972.
- [43] André Nies. *Computability and Randomness*. Oxford University Press, Inc., New York, NY, USA, 2009.
- [44] P. G. Odifreddi. *Classical recursion theory. Vol. II*. Studies in Logic and the Foundations of Mathematics. North-Holland Publishing Co., Amsterdam, 1999.
- [45] Julie Onton and Scott Makeig. High-frequency broadband modulations of electroencephalographic spectra. *Frontiers in Human Neuroscience*, 3(61):1–18, 2009.

- [46] Julie Onton, Marissa Westerfield, Jeanne Townsend, and Scott Makeig. Imaging human EEG dynamics using independent component analysis. *Neuroscience & Biobehavioral Reviews*, 30(6):808–822, 2006. Methodological and Conceptual Advances in the Study of Brain-Behavior Dynamics: A Multivariate Lifespan Perspective.
- [47] James C. Owings, Jr. A cardinality version of Beigel’s Nonspeedup Theorem. *The Journal of Symbolic Logic*, 54(3):761–767, Sep. 1989.
- [48] John E. Richards. Recovering dipole sources from scalp-recorded event-related-potentials using component analysis: principal component analysis and independent component analysis. *International Journal of Psychophysiology*, 54(3):201–220, 2004.
- [49] Marcus Schaefer. A guided tour of minimal indices and shortest descriptions. *Archive for Mathematical Logic*, 37(8):521–548, 1998.
- [50] C. P. Schnorr. Optimal enumerations and optimal Gödel numberings. *Theory of Computing Systems*, 8:182–191, 1974.
- [51] Claus-Peter Schnorr. A unified approach to the definition of random sequences. *Mathematical Systems Theory. An International Journal on Mathematical Computing Theory*, 5:246–258, 1971.
- [52] Claus-Peter Schnorr. *Zufälligkeit und Wahrscheinlichkeit. Eine algorithmische Begründung der Wahrscheinlichkeitstheorie*. Lecture Notes in Mathematics, Vol. 218. Springer-Verlag, Berlin, 1971.
- [53] Robert I. Soare. *Recursively enumerable sets and degrees*. Perspectives in Mathematical Logic. Springer-Verlag, Berlin, 1987. A study of computable functions and computably generated sets.
- [54] Frank Stephan and Jason Teutsch. An incomplete set of shortest descriptions. (to appear in *Journal of Symbolic Logic*).
- [55] Frank Stephan and Jason Teutsch. Things that can be made into themselves. (in preparation).
- [56] Frank Stephan and Jason Teutsch. Immunity and hyperimmunity for sets of minimal indices. *Notre Dame Journal of Formal Logic*, 49(2):107–125, 2008.
- [57] Kohtaro Tadaki. A new representation of Chaitin Ω number based on compressible strings. In Cristian Calude, Masami Hagiya, Kenichi Morita, Grzegorz Rozenberg, and Jon Timmis, editors, *Unconventional Computation*, volume 6079 of *Lecture Notes in Computer Science*, pages 127–139. Springer Berlin / Heidelberg, 2010.
- [58] Jason Teutsch. A savings paradox for integer-valued martingales. Manuscript.
- [59] Jason Teutsch. On the Turing degrees of minimal index sets. *Annals of Pure and Applied Logic*, 148:63–80, 2007.

- [60] Jason R. Teutsch. *Noncomputable Spectral Sets*. PhD thesis, Indiana University, 2007.
- [61] Mark R. Titchener. Digital encoding by means of new T-codes to provide improved data synchronization and message integrity, July 1984. Technical note.
- [62] Ricardo Vigário, Jaakko Särelä, Veikko Jousmäki, Matti Hämäläinen, and Erkki Oja. Independent component approach to the analysis of EEG and MEG recordings. *IEEE Transactions on Biomedical Engineering*, 47(5):589–593, May 2000.
- [63] Paul M. B. Vitányi, Frank J. Balbach, Rudi L. Cilibrasi, and Ming Li. Normalized information distance. In Frank Emmert-Streib and Matthias Dehmer, editors, *Information Theory and Statistical Learning*, pages 45–82. Springer US, 2009.
- [64] Pascal O. Vontobel, Roxana Smarandache, Negar Kiyavash, Jason Teutsch, and Dejan Vukobratovic. On the minimal pseudo-codewords of codes from finite geometries. In *Proceedings of IEEE International Symposium on Information Theory*, pages 980–984, Adelaide, Australia, Sept. 2005.
- [65] Yongge Wang. *Randomness and complexity*. PhD thesis, Mathematisch-Naturwissenschaftlichen Gesamtfakultät, Universität Heidelberg, 1996.
- [66] William P. Ziemer. *Modern Real Analysis*. Second edition, 2004. <http://www.indiana.edu/~mathwz/PRbook.pdf>.
- [67] A. K. Zvonkin and L. A. Levin. The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Mathematical Surveys*, 25(6):83–124, 1970.