

# Infinity in computable probability

Logical proof that William Shakespeare probably was not a dactylographic monkey

Maarten McKubre-Jordens\* & Phillip L. Wilson†

June 12, 2019

## 1 Introduction

Since at least the time of Aristotle [1], the concept of combining a finite number of objects infinitely many times has been taken to imply certainty of construction of a particular object. In a frequently-encountered modern example of this argument, at least one of infinitely many monkeys, producing a character string equal in length to the collected works of Shakespeare by striking typewriter keys in a uniformly random manner, will with probability one reproduce the collected works. In the following, the term “monkey” can (naturally) refer to some (abstract) device capable of producing sequences of letters of arbitrary (fixed) length at a reasonable speed.

Recursive function theory is one possible model for computation; Russian recursive mathematics is a reasonable formalization of this theory [4].<sup>1</sup> Here we show that, surprisingly, within recursive mathematics it is possible to assign to an infinite number of monkeys probabilities of reproducing Shakespeare’s collected works in such a way that while it is impossible that no monkey reproduces the collected works, the probability of *any* finite number of monkeys reproducing the works of Shakespeare is *arbitrarily* small. The method of assigning probabilities depends only on the desired probability of success and not on the size of any finite subset of monkeys.

Moreover, the result extends to reproducing all possible texts of any finite given length. However, in the context of implementing an experiment or simulation computationally (such as the small-scale example in [10]; see also [7]), the fraction among all possible probability distributions of such *pathological* distributions is vanishingly small provided sufficiently large samples are taken.

---

\*Adjunct Fellow, School of Mathematics & Statistics, University of Canterbury, New Zealand.

†Corresponding Author: phillip.wilson@canterbury.ac.nz. School of Mathematics & Statistics, University of Canterbury, New Zealand.

<sup>1</sup>The history of the relationship between classical logic and computation is long and complex, and beyond the scope of this paper. These results do indicate that mathematics using constructive logics, such as the Russian recursive mathematics used here, seems to be more suited to the simulation of the work of a computer than classical logic.

## 2 The classical experiment

The classical infinite monkey theorem [2, 6] can be stated as follows: given an infinite amount of time, a monkey hitting keys on a typewriter with uniformly random probability will almost certainly type the collected works of William Shakespeare [11]. We use a slightly altered (but equivalent) theorem involving an infinite collection of monkeys, and give an intuitive direct proof.

Let a string of characters of length  $w \in \mathbb{N}^+$  over a given alphabet  $A$  (of size  $|A|$ , including punctuation) be called a  $w$ -string. For example, “banana” is a 6-string over the alphabet  $\{a, b, n\}$ . Suppose each monkey is given a computer keyboard with  $|A|$  keys, each corresponding to a different character. Suppose also that the experiment is so contrived that each monkey will type its  $w$ -string in finite time.

**Theorem 1.** *At least one of infinitely many monkeys typing  $w$ -strings, as described in the previous paragraph, will almost certainly produce a perfect copy of a target  $w$ -string in finite time.*

*Proof.* Recall that for this theorem, the probability that any given monkey strikes any particular key is uniformly distributed. Let the target  $w$ -string be  $T_w$ . The chance of a given monkey producing  $T_w$  is simply the probability of him typing each character of the target text in the correct position, or

$$\underbrace{\frac{1}{|A|} \times \frac{1}{|A|} \times \cdots \times \frac{1}{|A|}}_{w \text{ terms}} = \left(\frac{1}{|A|}\right)^w .$$

Therefore, the probability that a given monkey *fails* to produce  $T_w$  is

$$1 - \left(\frac{1}{|A|}\right)^w .$$

Now, if we examine the output of  $m$  monkeys, then the probability that *none* of these monkeys produces  $T_w$  is

$$T_w(m) = \left(1 - \left(\frac{1}{|A|}\right)^w\right)^m .$$

Therefore, the probability that at least one monkey of  $m$  produces  $T_w$  is

$$P(m) = 1 - T_w(m) .$$

Now

$$\lim_{m \rightarrow \infty} P(m) = 1 .$$

□

In other words, as the number of monkeys tends to infinity, at least one will almost certainly produce the required string.

However, any real-world experiment that attempts to show this will, unless the target  $w$ -string and  $|A|$  are quite small, be very likely to fail, since the probabilities involved are tiny. For example, taking the English alphabet (together with punctuation and capitalization) to have 64 characters, a simple computation shows that, if the monkeys are typing 6-strings, the chances of a monkey typing “banana” correctly are

$$\left(\frac{1}{64}\right)^6 = \frac{1}{68719476736} \approx 1.5 \times 10^{-11}. \quad (1)$$

If it takes one second to check a single monkey’s output, then the number of seconds that will elapse before we have a 50% chance of finding a monkey that has typed “banana” correctly is outside the precision of typical computing software. Of course, if some monkeys have a preference for typing a certain letter more often than others — say ‘a’ — then this probability can be much larger. Indeed, it is non-uniformity among monkeys that we exploit to derive our main result in §4.

Results such as (1) have been interpreted [8, p.53] as saying that “The probability of [reproducing the collected works of Shakespeare] is therefore zero in any operational sense...”. In §4 we show that this probability can be made arbitrarily small in any sense, operational or otherwise.

### 3 A simple, classical non-uniform version

What if the monkeys do not necessarily strike their keys in a uniformly distributed manner? In this case, we might prescribe a certain probability for a particular monkey to type a particular  $w$ -string (and this probability need not be the same from one monkey to the next). Before we reach our main result, we outline a non-uniform classical probability distribution such that for any  $\varepsilon > 0$  the probability of success by monkey  $m$  is arbitrarily small, but with the probability of failure still zero. If we allow our probability distribution to be a function of  $m$  as well as  $\varepsilon$  then the following distribution will suffice:

$$1 - p_k(m, \varepsilon) = \delta(m - k)(\varepsilon - \sigma) + \delta(m + 1 - k)(1 - \varepsilon + \sigma),$$

where  $p_k$  is the probability of failing at monkey  $k$ , the Dirac delta function  $\delta(s) = 1$  for  $s = 0$  and is zero otherwise, and  $0 < \sigma < \varepsilon$ . Here, the probability of finding the target  $w$ -string at or before the  $m^{\text{th}}$  monkey,  $P(m)$ , is less than the prescribed  $\varepsilon$ , but success is still certain — we need merely look at  $m + 1$  monkeys.

In the following section we show that, surprisingly, this can be achieved with a probability distribution dependent only on  $\varepsilon$ , and not on  $m$ . That is, it is possible to produce a computable distribution so that, while each monkey produces Shakespeare’s works with nonzero probability, actually finding the culprit among *any* finite subcollection is very unlikely. To do so, we invoke a result from recursive mathematics.

### 4 The successful monkey is arbitrarily elusive

Within recursive mathematics, there is a theorem sometimes referred to as the *singular covering theorem*, originally proved by Tseitin and Zaslavsky (1956), and independently

by Kreisel and Lacombe (1957) (see [9]): given a compact set  $K$ , for every positive  $\varepsilon$ , one can construct a computable open rational  $\varepsilon$ -bounded covering of  $K$ .<sup>2</sup> It can be restricted to the interval  $[0, 1]$  as follows:

**Theorem 2.** *For each  $\varepsilon > 0$  there exists a sequence  $(I_k)_{k=1}^\infty$  of bounded open rational intervals in  $\mathbb{R}$  such that*

- (i)  $[0, 1] \subset \bigcup_{k=1}^\infty I_k$ , and
- (ii)  $\sum_{k=1}^n |I_k| < \varepsilon$  for each  $n \in \mathbb{N}^+$ .

Our principal result, Theorem 3, follows from this theorem, and highlights the tension between classical probability theory and its constructive counterpart as outlined in [5].

To set up our principal theorem, we first define  $M$  to be an infinite, enumerable set of monkeys (the *monkeyverse*), and for any natural number  $m$  the  $m$ -troop of monkeys to be the first  $m$  monkeys in  $M$ . Note that for any given monkey it is decidable whether that monkey has produced a given finite target string.

**Theorem 3.** *Given a finite target  $w$ -string  $T_w$  and a positive real number  $\varepsilon$ , there exists a computable probability distribution on  $M$  of producing  $w$ -strings such that:*

- (i) *the classical probability that no monkey in  $M$  produces  $T_w$  is 0; and*
- (ii) *the probability of a monkey in any  $m$ -troop producing  $T_w$  is less than  $\varepsilon$ .*

*Proof.* Suppose that the hypotheses of the theorem are satisfied. As above, let  $P(m)$  be the probability that a monkey in the  $m$ -troop has produced  $T_w$ , and let  $p_k$  be the probability that the  $k^{\text{th}}$  monkey has *not* produced  $T_w$ . Then

$$P(m) = 1 - \prod_{k=1}^m p_k.$$

Given  $0 < \varepsilon < 1$ , compute  $\varepsilon_0 = -\log(1 - \varepsilon)$ . For this  $\varepsilon_0$ , construct the singular cover  $(I_k)_{k=1}^\infty$  as per Theorem 2. Then set

$$p_k = \exp(-|I_k|).$$

To prove (i), observe that  $0 < p_k < 1$  for each  $k$ . The monotone convergence theorem now ensures that the product  $\prod_{k=1}^m p_k$  *classically* tends to 0, hence it is (classically) impossible that no monkey produces  $T_w$ .

On the other hand, we have (computably)

$$-\log(p_k) = |I_k|,$$

---

<sup>2</sup>Related theorems with detailed proofs and discussion were published by Tseitin and Zaslavsky in [12].

We hasten to add that while this may seem esoteric, the results really provide commentary on much more mainstream ideas such as computer simulations, since constructive logics are much more suited to theorizing about these.

whence, by the singular covering theorem,

$$\sum_{k=1}^m -\log(p_k) = \sum_{k=1}^m |I_k| < \varepsilon_0 = -\log(1 - \varepsilon)$$

for all  $m \in \mathbb{N}^+$ . Some rearranging shows that

$$\log\left(\prod_{k=1}^m p_k\right) = \sum_{k=1}^m \log(p_k) > \log(1 - \varepsilon)$$

and hence

$$\prod_{k=1}^m p_k > 1 - \varepsilon.$$

Then the probability of any member of the  $m$ -troop producing  $T_w$  is

$$P(m) = 1 - \prod_{k=1}^m p_k < \varepsilon$$

for *any* positive natural number  $m$ . This proves (ii).  $\square$

Thus, the chances of us actually *finding* the monkey that produces the collected works of Shakespeare can be made arbitrarily small, and the classical intuition that, since we have an infinite number of monkeys, Shakespeare's works must be typed by *some* monkey is of no help in *locating* the successful monkey.

We emphasize that, in contrast to the case in §3, the pathological distribution in Theorem 3 does *not* depend on  $m$ , the size of the troop we search.<sup>3</sup>

One might argue that it is easy to assign probabilities in such a way that any finite search will almost certainly not yield the monkey that produced it — by letting each monkey produce the target  $w$ -string with probability zero. However, in this case, *no* monkey will produce it. Our theorem shows that, even in the case where it is (classically) impossible that no monkey produces the target, it is still possible to make the probability of finding the monkey that accomplishes the necessary task arbitrarily small.

## 5 Target-free writing

One criticism of the above line of reasoning is that the experimenter requires knowledge of the target. There, the output of each monkey was tested against the collected works of

---

<sup>3</sup>Contrasting the classical with the computational view in the same proof may prove counterintuitive. We are hoping to shed light on why the *intuitive* result—that it is (in the classical abstract world) impossible that *no* monkey produces Shakespeare's works—clashes with the fact that it may be incredibly difficult (in the concrete computational world) to nail the cheeky monkey that did it.

What sense to make of the product  $\prod p_k$  of monkeys failing to produce Shakespeare classically tending to 0? The problem here is the *rate* at which it does so—this rate is computationally untractable.

Shakespeare: only if every character matched would it pass the test. However, suppose now that we wish to recreate Shakespeare’s work armed only with knowledge of the total character length in some alphabet. That is, we know that we require one of the  $|A|^w$  possible  $w$ -strings. Can we guarantee to complete the list (without repetition) and therefore recreate the collected works of Shakespeare (somewhere)? We note that the list can be shortened by checking for grammar etc.<sup>4</sup>; here we consider the worst case of the complete list, without repetition, of  $w$ -strings.

**Corollary 4.** *Any list of finite strings is completed in finite time with arbitrarily small probability.*

The proof relies on applying Theorem 3 multiple times using standard calculations.

## 6 Pathological distributions are arbitrarily rare

At first sight, Theorem 3 might appear to destroy any hope of finding the successful monkey. However, we have the following:

**Theorem 5.** *The probability that the probability distribution on the monkeyverse is constructed in such a way as to make the constructive probability of finding the desired monkey arbitrarily small, is arbitrarily small.*

*Proof.* Given  $0 < \varepsilon < 1$ , in order for the probability distribution to be pathological, the probability of *any* monkey in the  $m$ -troop outputting  $T_w$  cannot exceed  $\varepsilon$ . Therefore the fraction of pathological distributions over an  $m$ -troop is at most  $\varepsilon^m$ , and

$$\lim_{m \rightarrow \infty} \varepsilon^m = 0.$$

□

In short, we can make the fraction of pathological distributions arbitrarily small if we search sufficiently large  $m$ -troops. Here, then, is an a priori justification for large sample sizes in the case of computational simulations.

## 7 Discussion and further work

Recall that, throughout this paper, we take the term “monkey” to refer to some device capable of producing arbitrary but finite sequences of letters — computers satisfy this criterion. The theorems presented in this paper therefore have implications for computer simulations. In particular, when performing simulations of a probabilistic nature, the experimenter needs to ensure that pathological distributions do not arise, or arise rarely enough to provide a measure of confidence in the conclusion.

---

<sup>4</sup>Truncating the list in this way may be desirable in order to avoid being overwhelmed by “meaningless cacophonies, verbal farragoes, and babblings” [3].

It should also be noted that the classical non-uniform distribution presented above suggests that a pathological situation can never be ruled out with certainty, since if the experimenter tests just one more monkey, the result may be vastly different than observed earlier in the simulation. With practical considerations in mind, there will be some point at which costs (ethical and/or material) outweigh the benefit of testing further monkeys.

The proof of Theorem 3 required a result from constructive mathematics. We conjecture that such a result is classically impossible, since the singular covering theorem is classically not true.

A deeper fact here is that from the classical viewpoint, the computable reals have zero measure, and all finite texts produced by monkeys correspond to the rationals (or some other convenient subset of computable reals). The context of the results, then, would indicate that a careful constructive study of probability distributions provides *a priori* motivation for repetition of simulations for accuracy (to rule out accidental pathological distributions generated by computer programs), and has potentially more to say about issues involving computer simulations.

There is the further issue of what model of constructive mathematics provides a good framework for this sort of work. Philosophically there is tension between the intuitionistic free choice-sequence approach and the computable sequence approach, and within these approaches are further complications by sensitivity of the theory to the validity (or invalidity) of the various versions of König's Lemma. It is not the aim here to go deeply into these issues, which could lead to a lengthy series of papers. In the interest of brevity, we leave such explorations for future research.

It has not escaped our attention that science and mathematics have each been considered to be “games” of recombining a finite set of characters (even if we do not yet know what they all are). Even if we consider only finite strings which are syntactically sound, and not contradicted by empirical evidence, our result shows that completing such a list is not necessarily even *likely* to happen within *any* finite time, such as a human lifespan, the duration of a civilisation, or even the age of the universe.

**Acknowledgements:** The authors would like to acknowledge the contributions of the anonymous referees for substantial improvements to the paper. McKubre-Jordens was partially funded by Marsden Fund Fast-Start Grant UC1205.

## References

- [1] Aristotle (350 BCE) *Metaphysics*. Translation by W.D. Ross. <http://classics.mit.edu/Aristotle/metaphysics.html>. Retrieved 18 June 2010.
- [2] É. Borel (1913) ‘Mécanique Statistique et Irréversibilité’. *J. Phys.*, 5e série 3, 189–196.
- [3] J.L. Borges (1939) In *The Total Library: Non-Fiction 1922-1986*. Translated by E. Weinberger (2000). Penguin, London, 214–216.

- [4] D.S. Bridges & F. Richman (1987) *Varieties of Constructive Mathematics*. LMS Lecture Notes Series, Cambridge University Press.
- [5] Y.K. Chan (1974) Notes on Constructive Probability Theory. *Ann. Prob.*, 2(1), 51–75. Institute of Mathematical Statistics.
- [6] A. Eddington (1928) *The Nature of the Physical World: The Gifford Lectures*. New York: Macmillan.
- [7] Elmo, Gum, Heather, Holly, Mistletoe, & Rowan *Notes Towards The Complete Works of William Shakespeare* (2002) Khave-Society & Liquid Press, UK.
- [8] C. Kittel & H. Kroemer (1980) *Thermal Physics* (2nd ed.). W.H. Freeman Company.
- [9] B.A. Kushner (1999) Markov’s constructive analysis; a participant’s view. *Theoretical Computer Science* 219, 267–285.
- [10] ‘Give six monkeys a computer, and what do you get? Certainly not the Bard’, <https://www.theguardian.com/uk/2003/may/09/science.arts>.
- [11] W. Shakespeare *The Complete Works of William Shakespeare* (2001) Geddes & Grosset, Scotland.
- [12] I.D. Zaslavsky & G.S. Tseitin (1962) Singular coverings and properties of constructive functions connected with them, *Problems of the constructive direction in mathematics. Part 2. Constructive mathematical analysis*, Collection of articles, Trudy Mat. Inst. Steklov., 67, Acad. Sci. USSR, Moscow–Leningrad, 458–502. English translation: A.M.S. Translations (2) 98 (1971), 41–89, MR 27#2408.