

Lawrence Berkeley National Laboratory

Lawrence Berkeley National Laboratory

Title

Can An Evolutionary Process Create English Text?

Permalink

<https://escholarship.org/uc/item/2m93p1s3>

Author

Bailey, David H.

Publication Date

2009-09-22

Can An Evolutionary Process Create English Text?

David H. Bailey*

2008-10-29

Abstract

Critics of the conventional theory of biological evolution have asserted that while natural processes might result in some limited diversity, nothing fundamentally new can arise from “random” evolution. In response, biologists such as Richard Dawkins have demonstrated that a computer program can generate a specific short phrase via evolution-like iterations starting with random gibberish. While such demonstrations are intriguing, they are flawed in that they have a fixed, pre-specified future target, whereas in real biological evolution there is no fixed future target, but only a complicated “fitness landscape.” In this study, a significantly more sophisticated evolutionary scheme is employed to produce text segments reminiscent of a Charles Dickens novel. The aggregate size of these segments is larger than the computer program and the input Dickens text, even when comparing compressed data (as a measure of information content).

Keywords: evolution, intelligent design, computational biology, genetic programming

* Lawrence Berkeley National Laboratory, Berkeley, CA 94720. Email: dhbailey@lbl.gov. This work was supported by the Director, Office of Computational and Technology Research, Division of Mathematical, Information and Computational Sciences, U. S. Department of Energy, under contract number DE-AC02-05CH11231.

1. Introduction

A fundamental precept of evolutionary biology is that a combination of random variation and natural selection is the fundamental driving force for evolution. Over the course of many generations, species have diverged and adapted to their local environment, thus producing the remarkable variety of life presently seen on earth [Zimmer2001, pg. xii].

In contrast, skeptics of evolution, including some scholars in the creationist and intelligent design communities, assert that whereas natural biological processes may result in limited diversity among members of a given species, and possibly might result in minor changes in a single species over time, nothing fundamentally new can arise from “random” evolution [Dembski1999, pg. 113]. Thus, according to these writers, we must look elsewhere, possibly to a supernatural Designer, for the true source of novelty in the biological world.

For example, some creationist and intelligent design writers have questioned whether evolution could produce the human alpha-globin molecule, one of the components of hemoglobin. They argue that the probability that the alpha-globin chain in humans (a sequence of 141 amino acids) could form “at random” is something like one in 10^{183} , a number so enormous that such an event is unlikely to occur even once in the 4.5-billion-year history of the planet [Foster1991, pg. 1-20]. Such arguments can be countered by noting that all but about 25 of the 141 positions could have been different (judging from the differences in alpha-globin across the animal kingdom), and yet

still yield the fundamental oxygen transfer process (even though human biology has settled on our particular sequence). This reduces the odds of forming such a molecule in a hypothetical replay of evolution to something on the order of one in 10^{33} . This is still a huge number, but conceivably within the reach of a biological process [Bailey2000]. An even better response to an argument such as this is to note that evolution does *not* posit that biomolecules such as hemoglobin arose in a single shot; instead, they formed as the end product of a long series of structures, each of which was biologically advantageous in its own right. For instance, recent research indicates that hemoglobin arose in primitive bacteria for other purposes, and only later, in animals, did it adopt an oxygen transport function [Hardison1999, pg. 126-137].

Some writers have drawn the analogy to English text. For example, David Foster, in a book skeptical of evolution, discusses and then refutes an argument he attributed to Thomas Huxley, namely that a few monkeys typing randomly for millions of millions of years would type all the books in the British Museum. Foster asserts that even a single line of 50 characters could not be produced in this way, since there are at least 8.5×10^{49} alphabetic strings of length 50 (based on an alphabet of 26 characters and some other assumptions), so that generating a specific given string of length 50 “at random” is unlikely even over the multi-billion-year history of the earth [Foster1991, pg. 57].

In response to Foster, biologist Gert Kortof points out that Huxley could not possibly have told this story in 1860, because typewriters were not commercially available until 1874. Furthermore, it was not known at the time that genetic information is contained in a string of symbols (DNA), so it is highly questionable that this argument would have been used at all in the 1800s [Korthof2008]. Furthermore, as both Gert Kortof and Peter Olofsson have noted, this type of argument suffers from failing to define precisely what should truly be counted as “surprising.” To correctly assess the odds of such an occurrence, one should not calculate the probability of some single event (all of which may have the same probability), but instead the probability of all events in a class of similar events [Korthof2008; Olofsson2008].

Probability-based arguments continue to be employed in attempts to demonstrate that the current theory of biological evolution is fundamentally flawed. For example, William Dembski, a leading intelligent design scholar, claims that beyond a certain level of complexity, which he sets as 500 bits of specified information (in other words, a probability of roughly one in 10^{150}), it is completely unreasonable to assert that such highly ordered structures could ever arise in nature [Dembski2002, pg. 159-166].

2. Computational Simulations of Evolution

The author’s principal field of study is the application of state-of-the-art computer systems to questions of scientific research. Highly parallel computer systems are used by researchers in this field to perform large-scale simulations of physical phenomena such as the earth’s climate, supernova explosions, nanostructure physics, jet engine operation, and biological protein interactions. These calculations incorporate numerous known physical laws, utilize sophisticated numerical algorithms and parallel programming techniques, and typically involve enormous amounts of data. The objective of these computations is to simulate nature realistically enough

to permit scientists to draw reliable conclusions from the simulation results. In short, the computer is used as a “laboratory” to test scientific hypotheses.

Thus it is natural to consider using computer simulations to investigate some of the issues that have been raised regarding biological evolution. While a fully detailed simulation of biological evolution, incorporating hundreds of complicated and changing environmental factors, many thousands of competing species and many millions of individual organisms, each with a highly intricate biology, is well beyond the scope of what can be done today even on the most powerful computers, some simplified questions of this type can be addressed. Indeed, numerous studies of this sort have been published in the field of computational biology [Zimmer2001, pg. 94-97].

Along this line, in response to arguments of the type mentioned above, Oxford biologist Richard Dawkins has described a simple computer program he wrote to generate the Shakespearean sentence “Methinks it is like a weasel,” starting from a randomly generated character string [Dawkins1986, pg. 43-50]. The program achieved this in 41 evolution-like iterations, where, at each iteration Dawkins’ population of “sentences” were each scored based on how many letters were in agreement with his target phrase at the appropriate positions. Selective “breeding” slowly improved the score of the best sentence until there were no errors.

While this is an interesting exercise, it has significant flaws, some of which Dawkins himself acknowledged. To begin with, his experiment involved only a single “species,” whereas in the biological kingdom the branching tree of evolution develops in many thousands of directions simultaneously. Secondly, Dawkins’ process was defined by a single pre-specified target, whereas biological evolution is governed instead by a complicated “fitness landscape” involving hundreds of interacting factors such as climate, competing organisms in the same ecological niche, food supply, predators and diseases. Finally, Dawkins’ experiment progressed to a fixed future goal, whereas real biological evolution does not operate with any future goal in mind – each step must bestow some advantage. Nonetheless, Dawkins’ demonstration is intriguing.

3. Genetic Programming and Evolutionary Computing

Computer programs that employ evolutionary strategies have been applied to numerous problems in science and engineering. In this approach, which is variously known as “genetic programming” or “evolutionary computing,” a population of computer programs, electronic designs or molecular structures interchange “genes” and compete according to some “fitness condition” specified by the researcher. After several thousand iterations, the best-scoring “organism” is taken as the result of the process. Genetic programming was first introduced by John Holland in 1975 [Holland1975].

This approach has been quite successful. John Koza of Stanford University, one of the leading researchers in this field, has compiled a list of several dozen “human-competitive” results, judged according to very rigorous criteria, spanning fields as diverse as biotechnology and electronic engineering. In more than 20 cases, a patented 20th-century invention has been reproduced. In six cases, a patented 21st-century invention has been reproduced; in two recent cases, the results out-perform any existing technology [Koza2008].

4. Experimental Design

The objective of the present study is to explore whether an evolutionary computing approach can generate reasonably realistic English text – more than a single, short, targeted phrase as Dawkins produced, but instead a significant volume of text segments that are typical, say, of some genre of English literature.

The computer program that the author has written for this purpose begins by constructing a set of 1024 segments of text, each 64 characters long. The individual characters are chosen at random according to the natural distribution of individual characters in a sample of English text, which for the purposes of this study is Charles Dickens' novel *Great Expectations*. This text, 994,587 characters in length, includes normal punctuation, although all alphabetic characters have been changed to lower case for simplicity. Here is a sampling of some of the initially generated gibberish segments:

```
o ao ,fludoy aocueu feidh,iaemehaiheyh daneny shpesaems y nhte  
nrtnnbaa.nn hymeo t fiilunnw nt t,ntehg eu y' t h l dieosea ii  
mbdsoee lueleciro ,ynaenetg itln h srw l,pn uf svee,ee a'l sl  
snd etke snoymnra lhs gdnunmrs e trlhueafpraa.c.ys f yjser g
```

The program then scores each of these 1024 initial segments as to their “fitness” as Dickens text. The scoring function is as follows: the program finds the longest consecutive match of the given segment in character position 1 up through position 16 to any 16-long segment in the text of *Great Expectations*. This can be done very rapidly since the program sorts, in an initialization step, all 16-long shifted segments of the book's text, thus facilitating rapid lookup. This check is then repeated for positions 2 through 17 of the segment, then for positions 3 through 18, and on until the end of the segment is reached. The sum of the match lengths for these checks is the score for the given 64-long segment. Note that this scoring function has no specific future target, but only measures how typical the given segment is of text in *Great Expectations*. In other words, *Great Expectations* plays the role of “fitness landscape.”

Evolutionary iterations are then initiated, each of which consists of the following: First, the 50 top-scoring segments are permitted to “mate” (i.e., randomly exchange 4-long character strings, beginning at positions 1, 5, 9, etc.) with 4-long strings in the corresponding positions of another segment chosen at random from the 200 top-scoring segments. In addition, the 200 top-segments are permitted to “mate” with a segment randomly chosen from all 1024 segments in the same way. Then four types of “mutations” are performed: (1) for 400 randomly chosen segments, the character at one randomly chosen position is altered, with the new character chosen at random according to natural frequencies; (2) for 200 randomly chosen segments, two consecutive characters are altered in the same way; (3) for 100 randomly chosen segments, a “frame shift insertion” mutation is performed, or in other words, a new randomly chosen character is inserted at a randomly chosen position, and the segment is shifted to the right to accommodate it; and (4) for 100 randomly chosen segments, a “frame shift deletion” mutation is performed in a similar way. After these “mutations” have been performed, each resulting segment is scored, and the segments are sorted according to their new scores. This cycle repeats until all iterations have been performed. This scheme was deliberately chosen to be reminiscent of real biological evolution.

For these tests, 10,000 evolutionary iterations were performed (in other words, the process in the preceding paragraph was repeated 10,000 times). This number was chosen quite arbitrarily – it is roughly same as the number of human generations, assumed to be 20 years apart, over the past 200,000 years of evolutionary history. At the end of these iterations, the highest-scoring segment is taken to be the result of the trial, and the other 1023 segments are discarded.

In the execution of this program, it was observed that most of the improvements occurred fairly early, with subsequent iterations mostly “polishing” the result, and with long periods of “stasis” where nothing much happened. This behavior is not only typical of genetic programming applications, but it is also similar to the phenomenon of punctuated equilibria that has been observed in natural evolution [Zimmer2001, pg. xiii].

The pseudorandom number generator used for these experiments is based on a recent paper by Richard Crandall of Reed College and the present author [Bailey2002; Bailey2004]. As it turns out, this scheme can generate more than 3×10^{15} pseudorandom 64-bit floating-point numbers without repeating (a 64-bit floating-point number is a binary computer word with approximately 15-digit numeric precision). Note, for instance, that a “random” integer between 0 and 1023 can be produced using this generator simply by multiplying one of the pseudorandom values (a fraction between 0 and 1) by 1024 and then taking the greatest integer of the result. A sequence of pseudorandom floating-point numbers can be constructed beginning with any of a large number of “seeds.” If desired, by making some changes to this computer code (such as by employing higher precision arithmetic), the number of possible distinct pseudorandom “seeds” and sequences that can be produced by this scheme can be increased virtually without limit.

5. Computational Results and Analysis

The computer program ran for 24,576 repetitions of the process described above, using parallel programming facilities, thus generating 24,576 segments of length 64 characters each. These runs were done in batches of 8,192, and were run on 1024 processing cores of a parallel computer system at the author’s institution.

Many segments generated by the program, such as these four examples, contain syntax errors and nonsensical or misspelled words:

```
had i learn a lesson - looked at the stars, and held the gate.  
i felt as if he were a surgeon or a dentistrate in the table.  
did, in a comfortable about it and hear a triale beside her.  
he is sure to be executed on mond another in the mire of time.
```

Many other segments, such as these four, are syntactically acceptable but don’t make much sense:

```
and gloves, and as there no one and between his countenance.  
asked me why i wanted it and at her, said i, almost in a french  
for three in the station that he was in it rather resented.  
at remained, all these reasons for my part, he were a file.
```

But other segments are entirely reasonable, and could easily pass as fragments of literary text, clearly refuting the claims by some that a “random letter gun” plus evolution cannot possibly generate English text. A sampling of some of the high-scoring segments is shown in Table 2 on the next page. The number preceding each segment is the score that this segment achieved.

Along this line, a quiz was constructed as shown in Table 1, which was then presented to some college students at a large university in the western U.S. They were told only that some of these twenty segments of English text are extracted from the writings of Charles Dickens, and some are computer generated. Neither the professor who administered the quiz nor any of the students in these classes had previously seen the quiz or knew any of the answers.

1. up at it for an instant. but he was down on the rank wet grass,
2. or do any such job, i was favoured with the employment. in order,
3. at the fire as she took up her work again, and said she would be
4. the monster was even careless as to the word that i had him so.
5. as to go with him to his father's house on a visit, that i might
6. fitted it to nothing and get the ashes between me to the last.
7. as no relation into another that it is the same room - a little
8. a separation to be made for the desolater, like the man he was.
9. we said that as you put it in your pocket very glad to get it, you
10. that he had treated him to a little bee, he was to call the
11. if he had for a time such an interest here and contented me.
12. great iron coat-tails, as he had done, and then ran to that.
13. he saw me going to ask him anything, he looked at me with his glass
14. on my objecting to this retreat, he took us into another room with
15. been born on there, or that i had the greatest indignature.
16. the chimney as though it could not bear to go out into such a night
17. later to settle to anything i had hesitated as to the sound.
18. the greatest slight and injury that could be done to the many far
19. of it on the hearth close to the fear that she had done rather
20. out of my thoughts for a few moments together since the hiding had

Table 1: Which of these 20 segments are authentic Dickens text,
and which are computer-generated?

The reader is invited to try to identify which of these are authentic snippets of Dickens’ writings and which are computer-generated segments produced by the scheme described above, without consulting any references. The answers are given in the Appendix below.

862 of it on the hearth close to the fear that she had done rather
813 gate of the gardens, and found the street and they of the rest.
811 waiting for me near the deed as to that, while they rested.
809 be as bad as to the hearth and saw that it is not of my eye.
807 his hint had passed within a boat at the side opening lines,
807 when i saw the room, seemed at once to be in a little distance.
804 to let me he was heard, as in the last we went into the street.
803 want to see the more he disposed of, it was when i ascended it.
803 they had lasted, and held a pirate that the trials were on.
800 a time when i had treated him as it was the business of the
797 it is there for a little time said in this to be tired of me."
797 was to be seen that he showed the point on the marshes on the
793 come out to me, and he started as if she had a time to this!
791 catching his head at me in the old tone, that it is not the
787 a hit at me at last, when we got to the steerer i would not.
787 it was one of the restless and with the rest of those down.
784 lies, and both entreated me with it and the last i was late.
784 it were the wish of your own and desperate and ran with that."
783 we are not the way as i was soon at the battery with a far more
783 at these stores in detail it was administered to me as if there
782 point in at his to be the sharpened by the wind in the rain.
782 the old one, and here there was recompense in it. i had her.
782 if he had for a time such an interest here and contented me.
780 the discussion, that the notion on that basement of the world?
778 on the page came in and say nothing that he has - of the rest.
778 i went and said to my prisoners who attended on them at the
777 for three in the station that he was in it rather resented.
776 of me and the returned one of that man said he did, i shout.
775 for he had a bar, and that the fireside, and he is one of them.
774 was best not to be other in our present life of her with me.
774 that he had treated him to a little bee, he was to call the
774 referred to in the first faint dust, so that one me a start.
773 and fastened at her, as his intention he had done another."
772 a separation to be made for the desolater, like the man he was.
772 himself, years then, he saw me, and motioned that gentleman.
771 but stand that she was not to go astonished to see that the
771 as no relation into another that it is the same room - a little
771 if i had known the wine to begin to be trying my abhorrence.
771 not to be in a winter in the air to get the horrors on him."
766 in his hand, and he saw that her not on the stones of the water
763 fitted it to nothing and get the ashes between me to the last.
762 turned his face that when there was an air of the intent i had.
762 the time to catch my attention to his as he sat in the dead of
762 but this is another name for me to rise and used it to be, and
761 ate, and i nodded at her a state of my attendant on the head.
761 meditating on it into brass and eyes of the set of the word.
760 one who had it in the candles. we all three times said yes.
758 he had a tiresome journey of it in trying to the point to see.
758 no time to startle me from there, as if he had made the sides.
757 to death, and it is needless to add there is on to all the
757 the monster was even careless as to the word that i had him so.
756 as that he had come home to in his guard were ready at all the
754 woman to be advised by at this stage of the lords and trees.

Table 2: A sampling of high-scoring computer-generated text segments.

Looking collectively at the 66 sets of responses that the author received for this quiz, the average number of correct responses among the 20 items is 40 (60.6%), which is statistically significant, although not a great deal higher than the 33 correct responses (50%) that one would expect at random. If we look at “majority vote” statistics, indeed the majority of the 66 responses is correct for most items, but it is wrong for items #8, 9, 11, 13, 20, and in two other cases (#1 and #15) the margin of the “vote” is slim. All of the computer-generated items had at least 18 incorrect responses out of 66. Items #8 and #9 proved especially troublesome to these students, with only 17 and 18 correct responses, respectively (#8 is computer-generated; #9 is from Dickens’ *Great Expectations*).

It is important to note that none of the 24,576 segments produced by the author’s computer program coincides with any 64-character segment of *Great Expectations*. In other words, the computer program is not merely “regurgitating” portions of the input text file. What’s more, none of these 24,576 computer-generated segments coincides with any other of the 24,576 segments in more than 17 consecutive characters, even when shifts are allowed – in other words, all 24,576 segments are substantially distinct.

It is also interesting to note that the computer program constructed, in the 24,576 generated segments, numerous legitimate English words that do not appear anywhere in *Great Expectations*. Some examples include the following:

administer, agitate, allowing, arrangers, assail, assessed, attenuated, attraction, auctioned, baroness, batter, bellow, breather, chastened, coached, conspire, contentions, credited, deceived, descension, despot, detained, detriment, discriminate, dispensable, dispenses, distances, easiness, elected, enhance, formations, foundered, generate, generation, gentile, glisten, gradation, handler, hitches, inconvenient, increase, intentionally, intentioned, intimations, iterate, lacerate, liberate, liberated, likened, mattered, mediated, migration, ministered, mission, necessitated, operated, positioned, possibilities, powered, prostrate, releases, remonstrations, renderings, retirements, retreated, searches, session, silenced, simmer, situations, slinging, soothing, spheres, statements, steamed, steers, straits, stratified, stressed, teased, tendered, termination, thickens, threatenings, threshes, torments, traitors, trench, utters, wandered, wither, weathers

Such considerations are significant with regards to the objection that has been raised by some that this type of exercise does not truly generate anything novel that was not already present in the beginning, buried in the complexity of the computer program and input data.

Such objections can be quantitatively tested here. Recall that 24,576 distinct 64-long segments of text were generated by the computer program, for a total of 1,572,864 bytes. Note that this figure is higher than the length of the computer program (17,622 bytes) plus the length of the *Great Expectations* input file (994,587 bytes), which total 1,012,209 bytes. An even more telling comparison can be made by first compressing these files using the Unix “gzip” utility, a widely used and highly effective data compression program (which thus can be used as an objective measure of information content) [Gzip2008]. The compressed file of output segments is 541,297 bytes long (34.4% of the original length). By comparison, the compressed *Great Expectations*

text file is 363,899 bytes long (36.5% of the original), and the compressed computer program is 5,139 bytes long (29.2% of the original), for a total of 369,038 bytes. In other words, the computer program is generating some 46% more compressed information than contained in the compressed computer program and input data file. What's more, it is clear that if the program were run longer, it could generate far more than the 24,576 segments produced here. The only limiting factor appears to be the maximum period of the pseudorandom number generator employed, which as noted above can be greatly increased by making minor programming changes.

In this regard, it is worth recalling the “birthday paradox” of elementary probability theory. This is the counter-intuitive mathematical fact that if more than 23 persons are assembled in a room, the chances are at least 50-50 that some pair of persons has the same birthday. In general, the probability P that among a group of n persons with d equally probable birthdays ($d = 365$ in this case) at least one pair has the same birthday, is closely approximated by the formula $P = 1 - \exp(-n^2 / (2d))$, where \exp denotes the exponential function [Weisstein2008]. Recall the observation above that none of the 24,576 generated segments coincides with any other of the 24,576. The birthday paradox formula then implies that the total number of distinct 64-long segments of text that the program is capable of producing in its current configuration is, with high probability, at least 400,000,000 – otherwise it is likely that at least one pair among the 24,576 segments would coincide. A file of 400,000,000 segments would be 25.6 billion bytes long, which even when compressed (assuming the 34% ratio above) would still be roughly 8.7 billion bytes long.

6. Conclusion

This modest study reports some computational experiments testing whether a process akin to biological evolution can generate English text. By constructing a computer program based on methodology developed in the genetic programming community, it has been shown that English text segments reminiscent of Dickens literature can indeed be generated. At the least, some of the better resulting text segments are sufficiently good to fool human judges in an informal test – college students were only correct in distinguishing true Dickens from computer-generated segments about 61% of the time (on average).

Additional improvements could be made to this computer program, which very likely would result in higher quality output text. For instance, it was noticed that in numerous cases in this study, spelling errors in the final output segment were not corrected during the 10,000 trial iterations, not because mutations and mating never produced a correct spelling (they did), but instead because the corrected text did not score higher for some reason. Thus, a more sophisticated scoring function should result in higher-quality output text.

Further, it is clear from related research that significantly better quality output could be achieved by employing a much larger sample of Dickens literature, and by operating word-by-word instead of character-by-character. For example, Google's remarkable success in machine translation, according to Franz-Josef Och, who now leads Google's translation effort, is based on utilizing a bilingual text collection of at least one million words and two monolingual collections of roughly one billion words each. Statistical models obtained from this data are then used to translate between the pair of languages [Brants2007; Google2008].

It should be kept in mind that the computer program used by the author is deterministic, in the sense that if it were re-run it would produce exactly the same set of results, and a deterministic program technically cannot “create” information. But this is paradoxical only because the pseudorandom number generator is generating iterates of sufficiently “random” quality that compression utilities cannot distinguish them from those produced by a truly random process.

Along this line, the recently published book *A New Kind of Science*, written by the well-known mathematical physicist and software entrepreneur Stephen Wolfram, presents numerous examples of very simple computational schemes that produce behavior of arbitrarily high complexity. Wolfram then argues that the enormous randomness inherent in many basic natural processes (which is rooted in the fundamental randomness of quantum mechanics) may itself explain much of the evolutionary novelty observed in nature [Wolfram2002]. Ongoing research in this arena may shed more light on these questions.

Acknowledgment

The author wishes to acknowledge computer time for this study on “Franklin,” a Cray XT4 parallel computer system in the National Energy Research Scientific Computing Center at Lawrence Berkeley National Laboratory.

Appendix

In the exercise presented above, these items are authentic Dickens:

1, 2, 3, 5, 9, 13, 14, 16, 18, 20

These items are produced by the computer program:

4, 6, 7, 8, 10, 11, 12, 15, 17, 19

References

- [Bailey2000] David H. Bailey, "Evolution and Probability," *Report of the National Center for Science Education*, vol. 20 (2000), no. 4, available at <http://www.dhbailey.com/papers/dhb-probability.pdf>.
- [Bailey2002] David H. Bailey and Richard E. Crandall, "Random Generators and Normal Numbers," *Experimental Mathematics*, vol. 11, no. 4 (2002), pg. 527-546.
- [Bailey2004] David H. Bailey, "A Pseudorandom Generator Based on Normal Numbers," 2004, available at <http://crd.lbl.gov/~dhbailey/dhbpapers/normal-random.pdf>.
- [Brants2007] Thorsten Brants, Ashok C. Popat, Peng Xu, Franz J. Och and Jeffrey Dean, "Large Language Models in Machine Translation," *Proceedings of the 2007 Conference on Empirical Methods in Natural Language Processing*, available at <http://acl.ldc.upenn.edu/D/D07/D07-1090.pdf>.
- [Dawkins1986] Richard Dawkins, *The Blind Watchmaker*, W. W. Norton, New York, 1986.
- [Dembski1999] William A. Dembski, *Intelligent Design: The Bridge Between Science and Theology*, InterVarsity Press, Downers Grove, IL, 1999.
- [Dembski2002] William A. Dembski, *No Free Lunch: Why Specified Complexity Cannot Be Purchased without Intelligence*, Rowman and Littlefield, Lanham, MD, 2002.
- [Foster1991] David Foster, *The Philosophical Scientists*, Marboro Books, New York, 1991.
- [Hardison1999] Ross Hardison, "The Evolution of Hemoglobin," *American Scientist*, vol. 87, no. 2 (March-April 1999), pg. 126-137.
- [Google2008] "Google Translate," 2008, available at http://en.wikipedia.org/wiki/Google_Translate.
- [Gzip2008] "Gzip," 2008, available at <http://en.wikipedia.org/wiki/Gzip>.
- [Holland1975] John H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*, University of Michigan Press, Ann Arbor, MI, 1975.
- [Hoyle1981] Fred Hoyle and Chandra Wickramasinghe, *Evolution from Space: A Theory of Cosmic Creationism*, J. M. Dent and Sons, London, 1981.
- [Korthof2008] Gert Korthof, "Does Protein Specificity Destroy the Theory of Evolution?" June 2008, available at <http://home.planet.nl/~gkorthof/korthof15.htm>.

[Koza2008] John Koza, “36 Human-Competitive Results Produced by Genetic Programming,” 2008, available at <http://www.genetic-programming.com/humancompetitive.html>.

[Olofsson2008] Peter Olofsson, “Intelligent Design and Mathematical Statistics: A Troubled Alliance,” *Biology and Philosophy*, vol. 23 (2008), no. 4, pg. 545-553.

[Weisstein2008] Eric Weisstein, “Birthday Problem,” available at <http://mathworld.wolfram.com/BirthdayProblem.html>.

[Wolfram2002] Stephen Wolfram, *A New Kind of Science*, Wolfram Media, Champaign, IL, 2002.

[Zimmer2001] Carl Zimmer, *Evolution: Triumph of an Idea*, HarperCollins, New York, 2001.