Randomness and Structure

There is a lot of plain silliness surrounding how people talk about the role of randomness in producing structure in nature. There is no denying that randomness either exists in nature or that we have to invoke random behavior in our models due to our ignorance (which of these you choose is a matter of some philosophical debate). However, there is also no denying that its role is far from understood and that some of the brasher amongst us forget that from time-to-time and end up uttering the most ridiculous pronouncements.

The poster child for this sort of unalloyed over-zealousness is nicely examined and ridiculed in the following homework exercise from the book *Thermal Physics*, by Kittel and Kromer, who broached the question about the role of pure randomness in their question entitled *The meaning of “never”* (page 53 in the 2nd edition). The beginning of the problem reads

<It has been said (footnote: J. Jeans, *Mysterious Universe*, Cambridge University Press, 1930, p. 4. The statement is attributed to Huxley) that “six monkeys, set to strum unintelligently on typewriters for millions of years, would be bound in time to write all the books in the British Museum.”…Could all the monkeys in the world have typed out a single specified book in the age of the universe?>

While the attribution to Huxley by Jeans (presumably Thomas Henry Huxley) seems to be apocryphal, there is no denying that a large segment of the population shares in the expressed sentiment that randomness eventually leads to structure. But the operative question is should they be believing in the power of fluctuations and chance.

To help in answering the posed question, Kittel and Kromer ask the student to make the following assumptions:

* The monkeys have 44 keys on their typewriters. Ignoring the use of the shift key, the breakdown is 26 letters, 10 digits, 8 punctuation marks. Modern laptop keyboards seem to have something more like 10 punctuation mark keys but a smaller number of keys is better for the meandering paws of the monkeys as they have fewer ways of producing gibberish.
* The primate population is $$10^{10}$$ monkeys, which corresponds to 10 billion simian typists, roughly 30% more than the number of people walking the planet at this moment
* Allot $$10^{18}$$ seconds to our monkeys for their unorchestrated typing. This time span totals up at about 31.7 billion years compared to the 13.8 billion years [estimated age of the universe](https://en.wikipedia.org/wiki/Age_of_the_universe).
* Each monkey can type 10 characters per second, which fast even for the most trained typist.
* The specified text is Hamlet, which has $$10^5$$ characters and we ignore case.

Kittel and Kromer calculate that the probability that any 100,000 character string chosen at random matches Hamlet is $$10^{-164,345}$$ and that the probability that the monkeys will produce Hamlet is $$10^{-164,316}$$ to which they state:

<The probability of Hamlet is therefore zero in any operational sense of an event, so the original statement at the beginning of the problem is nonsense: one book, much less a library, will never occur in the total literary production of the monkeys.>

But, before moving onto to a deeper discussion about randomness and structure, it is useful to skim through the probability computations in a simplified setting. To see the nuts and bolts, we will ask what is the probability of producing the inviting string ‘hello’ given a keyboard with only 14 keys corresponding to the first fourteen letters of the alphabet {a,b,c,d,e,f,g,h,i,k,l,m,n,o}. There are five characters in the target string translating to five slots each with 14 possible choices for a grand total of 537,824 possible choices (one of which is shown below).

The only additional pieces used in the Kittel and Kromer problem, which are not needed in this simple example, are the use of logarithms to estimate the number of realizations of Hamlet at $$44^{10^5}$$ and the multiplication of the corresponding probability by $$10^29$$ to account for the number of realizations that the monkey population can produce in the allotted time.

Numbers like these are, or should be, damning evidence to anyone who thinks structure arises solely from random fluctuations and this problem has been a major hurdle that biologists have to deal with when considering how life evolves. Afterall, the idea that many people have about biological evolution is that if one waits long enough structure will appear but, clearly, that can’t happen simply with blind chance.

So, how should one think about the role that randomness plays in physical and biological phenomena? Richard Dawkins claims in *The Blind Watcher* that the key to randomness in biological processes is found in the concept of cumulative selection. He brushes aside the objections raised above in his chapter entitled ‘Accumulating small change’. In that chapter, Dawkins supports this claim by presenting his ‘weasel program’, which is a piece of code that takes a random string (produced, say, by one of the monkeys) and breeds a fitter string. For example, one can take ‘aF!rty.opRSWi’ and breed it towards ‘Hello, World!’.

Dawkins is quite proud of his little tinkering and he uses it as refutation of the idea of intelligent design. However, his program is simply a probabilistic way of traversing a tree of all possible strings of the target length from a random string towards the bottom to a specific string at the top using a fitness function and a rate of randomizing, which he calls mutation. (A pythonic version is included at the end of this post).

The following figure illustrates the 27 random strings that can be created from the simple alphabet of {aer}. Each variation is color-coded based on how many mutations away it is from the target string ‘ear’, with green, yellow, and red corresponding to 1, 2, or 3 mutations, respectively.

For simplicity, assume that when the program breeds a brood from a given parent it can only make one mutation. Then it can move the fitness up a level or down a level or keep it the same (with obvious limitations at the top and bottom). The action of selecting the fittest child is the equivalent of always moving up the ladder provided a sufficient number of children are produced. It is an almost certainty that the program must reach its goal. And so, Dawkins argues, randomness amortized over generations produces the most complex structures we know – living organisms.

But the weasel program merely underscores his sloppy thinking about randomness. Dawkins’s process of cumulative selection is just that, a process that assumes a whole set of deterministic rules, of which the following are a sample:

* The length of the target string is known a priori
* The content of the target string is also known – it is the goal or telos of the algorithm
* A deterministic fitness function exists that clearly finds the best string or strings from a set
* At each generation a fixed number of children are produced
* The best child is always chosen by the fitness function
* The best child always fathers the next generation
* The algorithm exists within a deterministic environment in which strings can be interpreted as commands and those commands do the same thing time and again

With so much structure it is easy to see why a desirable outcome results in a fairly short amount of time. The key point being that a clear set of underlying rules combined with some randomness is really the only way in which a stochastic process leads to structure. In other words, it is the process that matters and blind chance merely adds spice to the recipe. How the underlying process comes about is something which, curiously, Dawkins has little to say, but that hasn’t prevented untold numbers of people being duped by his 20th century version of the ‘Huxley argument’ about randomness. I guess monkey see monkey do.