

The autopsy of a dissection of a vivisection

The Evolutionary Informatics Lab

May 6, 2011

Thomas D. Schneider has taken the time to respond [9] to our paper “A Vivisection of the *ev* Computer Organism: Identifying Sources of Active Information” [6], which itself is a critique of his previous work on the *ev* system [8]. His response provides some important criticisms of our work and therefore merits a response of our own.

Schneider takes issue with our reference to an “blind, unguided search.” However, we mean that the search is unguided in the same way that Dawkins’ watchmaker is blind. The intent is not to downplay selection, but to distinguish the evolutionary process from one in which an underlying plan is unfolding. Jerry Coyne has described evolution as being unguided in contrast with a guided theistic evolution position [1]. Our intent was to describe the process as non-forward looking and non-teleological. We are not casting Darwinian search as a random search, which is clear from the body of the manuscript.

Schneider’s original paper measures information as a decrease in entropy following the technique of Shannon [10]. In our response, we did not measure information in this way, instead using the same measurement of information as in our previous work [5, 2, 3, 4]. In this context Shannon’s method is not the best way of measuring information. Regardless, while we may have different preferred techniques for measuring information, we agree that the *ev* genome does in fact gain information.

We demonstrate that the information gained by the *ev* genome is not the result primarily of replication, mutation, and selection. Schneider agrees, as he describes the information as coming from the environment. The correct binding sites are located in the environment and are extracted by evolution and stored in the genome. We actually agree with Schneider. However, we are taking a closer look at the process of this information extraction.

There is a difference between information being present in the environment and an information source being present. As an example, consider a scenario depicted in a Calvin and Hobbes comic strip [11]. Calvin asks Hobbes to guess a number between one and seven hundred billion. Hobbes makes two guesses to which Calvin responds, “Nope. Guess again.” At this point, Hobbes gives up, seeing that attempting to guess the number is going to be a fruitless endeavor.

The number was in Hobbes’ environment, located inside Calvin’s head. But Hobbes has no access to that information. The only source of information available is for Hobbes to make guesses at the correct number. However, since Calvin only indicates whether or not the given guess is correct, attempting to extract the correct number is going to take a very long time. Had Calvin indicated whether the given number was too low or too high, he would have been providing much more information and Hobbes would have been able to rapidly determine the number.

In *ev*, the correct binding sites are located in the environment. According to Schneider, *ev* extracts that information. However, *ev* does not have access to that information, it can only access it through secondary information sources. The simple existence of that information in the environment does not guarantee it can be efficiently extracted. The success of *ev* depends on the availability of information sources that make

efficient information extraction possible. This is why are we concerned with the number of queries, as they allow us to measure the efficiency of information extraction.

We demonstrate that the information derived by *ev* results from two primary sources: a perceptron structure, which makes finding the target more likely by orders of magnitude even absent of selection, and the hamming oracle, which is a large source of target-specific information. Schneider reports having eliminated the threshold, with *ev* still managing to converge to its target. However, the bias of the perceptron structure is not due solely to this threshold. In our text we state that:

The random variables in S , though, are not independent. Each, for example, contains the same bias, θ , as one of the seven numbers in its sum.

Our discussion of the threshold does imply that it was primarily responsible for the bias of the perceptron structure. Additionally, we used to term bias to refer to the threshold as well as describing the structure as biased towards producing certain types of strings. We did not analyze the structure to determine the source of the bias; rather, we only calculated the bias of the structure as a whole. Unfortunately, we implied that the threshold was responsible for this bias; this was not our intention.

The threshold in the perceptron does contribute some bias. Choosing the weight matrix for the perceptron is like choosing a coin from a collection of coins of varying bias. The sites which end up being bound are then selected by flipping that coin. There is a relatively high probability of selecting a biased coin and thus producing an output string with relatively low entropy. Since *ev*'s target can be seen as a string with low entropy, the bias of the perceptron helps in locating the target, as we demonstrate in the paper.

Furthermore, we did not claim that *ev* would fail to converge without the perceptron structure. The lack of a perceptron structure would merely cause *ev* to take longer to converge. Without a perceptron structure *ev* would be essentially be a simple version of the "METHINKS IT IS LIKE A WEASEL" presented by Richard Dawkins [7]. The hamming distance oracle is a powerful source of information, and it alone would allow *ev* to find its target in a reasonable amount of time.

Schneider addresses the hamming oracle issue by assuming that nature provides a correct fitness function (a hamming function) that allows for positive selection in the direction of our target. He argues that this fitness is based on a

biologically sensible criteria: having functional DNA binding sites and not having extra ones.

But this describes a target; this is the desired goal of the simulation. The fitness function actually being used is a distance *to* this target. This distance makes efficient information extraction possible.

Schneider is concerned with the biology and expresses concern that we are "veering" away from it. We demonstrated that *ev*'s success derives from the use of a perceptron structure and a hamming distance oracle. Neither of these concepts comes from biology, rather they are concepts developed within the world of computer science. As such, it is not clear the *ev* models biology closely enough to claim any particular merit in following the biology. Regardless, our particular interest was in the question of information gain in the genome. Perhaps these structures are perfectly good models of what we find in biology. Our intention is to show the existence of these information sources and they are necessary and responsible for the information gain, not any magic of the evolutionary process.

While we appreciate Schneider's correction on some typographical errors, the results of the paper still hold. We agree with Schneider that information is gained by the genome through extraction of it from the environment. However, we show that information sources must exist which allow access to that information. Schneider has assumed two powerful sources of information within *ev*, the perceptron structure and the hamming distance oracle. These information sources are the cause behind the gain in information, not

evolution or the mere existence of information in the environment. The information in the environment takes the form of a point of high fitness in a fitness landscape. However, there mere existence of that point does not allow reaching it. Reaching that point requires a particular shape to the fitness landscape to guide evolution.

1 Errata

We would like to thank Schneider for bringing the following errors and unclear sections to our attention:

1. Page 2: “genome consisting of a 256 base string (excluding five extra bases at the end that are not part of the genome proper)”

We did not intend to convey that the 5 extra bases were not part of the genome. By genome proper we meant the portion of the genome at which it was possible for a binding site to start.

2. Page 2: The coordinates given here differ slightly from those in the original paper. The actual results are not significantly affected. The conclusions, though, stand strong.

3. Page 3: “In the search for the binding sites, the target sequences are fixed at the beginning of the search. The weights, bias, and remaining genome sequence are all allowed to vary”

The target sequence being referred to is the sequence of 1’s and 0’s corresponding the correct locations of the binding sites. The “remaining genome sequences” is referring to the genome aside from the weights and bias.

References

- [1] Jerry Coyne. New survey on science and religion. <http://whyevolutionistrue.wordpress.com/2010/07/12/vcu-survey-on-science-and-religion/> Last accessed: 03/01/2011, July 2010.
- [2] W.A. Dembski and R.J. Marks. Bernoulli’s principle of insufficient reason and conservation of information in computer search. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, pages 2647–2652, October 2009.
- [3] W.A. Dembski and R.J. Marks. Conservation of information in search: Measuring the cost of success. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 39(5):1051–1061, 2009.
- [4] W. Ewert, W.A. Dembski, and R.J. Marks. Evolutionary synthesis of nand logic: Dissecting a digital organism. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, pages 3047–3053, October 2009.
- [5] Winston Ewert, George Montañez, et al. Efficient per query information extraction from a Hamming oracle. In *Proceedings of the 42nd Meeting of the Southeastern Symposium on System Theory, IEEE, University of Texas at Tyler*, pages 290–297, 2010.
- [6] G. Montañez, W. Ewert, W. Dembski, and R. Marks. A Vivisection of the ev Computer Organism: Identifying Sources of Active Information. *BIO-Complexity*, 2010:1–6, 2010.
- [7] Dawkins R. *The blind watchmaker: why the evidence of evolution reveals a universe without design*. Norton, City, 1996.
- [8] T. D. Schneider. Evolution of biological information. *Nucleic Acids Res.*, 28(14):2794–2799, 2000.

- [9] Thomas D. Schneider. Dissection of “a vivisection of the ev computer organism: Identifying sources of active information”. <http://www.ccrnp.ncifcrf.gov/~toms/papers/ev/dembski/dissection.html> Last accessed: 02/28/2011.
- [10] C. E. Shannon. A mathematical theory of communication. *Bell system technical journal*, 27, 1948.
- [11] Bill Watterson. Calvin and hobbes. <http://www.gocomics.com/calvinandhobbes/2009/05/22> Last accessed: 03/01/2011.