

## ANALYSIS OF MUTATION AND FIXATION FOR LANGUAGE

BART DE BOER<sup>\*1</sup>, BILL THOMPSON<sup>2</sup>, ANDREA RAVIGNANI<sup>1,3</sup>,  
and CEDRIC BOECKX<sup>4</sup>

<sup>\*</sup>Corresponding Author: bart@ai.vub.ac.be

<sup>1</sup>AI-lab, Vrije Universiteit Brussel, Brussels, Belgium

<sup>2</sup>Social Science Matrix, UC Berkeley, Berkeley, USA

<sup>3</sup>Seal Centre, Pieterburen, the Netherlands

<sup>4</sup>ICREA and Universitat de Barcelona, Barcelona, Spain

Did the biological basis for language emerge instantaneously from a single mutation (Berwick & Chomsky, 2016) or gradually through many (Hurford, 2014)? Advances in what we know about how humans have evolved make it possible to formally analyze these alternative evolutionary scenarios. Evolution operates through transmission, variation and selection. Theoretical biology has developed a set of (mathematical) tools with which these processes can be studied (see Blythe & McKane, 2007 for a review relevant to language evolution). These can be used to assign likelihoods to different evolutionary scenarios, and also to calculate estimates of how much time would be needed for each scenario.

The two relevant techniques are: 1) a way to calculate probabilities of mutations and 2) a way to calculate the probability that a mutation spreads in the population (in biological parlance: goes to fixation), as well as the time it would take to spread. Multiplying these probabilities gives the probability of a mutation occurring and subsequently going to fixation. The first is provided by extreme value theory applied to the probability of beneficial mutations, developed by Orr (2003). From this it follows that the probability of a beneficial mutation is approximately  $p_{mut}(\varphi) = \alpha \cdot e^{-\alpha \varphi}$ , where  $\varphi$  is the fitness effect of the beneficial mutation (i.e. how much more offspring an organism with the mutation has on average than an organism without it) and  $\alpha$  a constant, with a value that is at least 30 (de Boer, Thompson, Ravignani, & Boeckx, to appear).

The probability of fixation and the time to fixation have been worked out by among others Kimura (Kimura, 1957; Kimura & Ohta, 1969). The fixation

probability is given by:  $p_{fix}(\varphi) = (1 - e^{\beta}) / (1 - e^{\beta \cdot N})$ , where  $\beta = -4\varphi / (2 + \varphi)$  and  $N$  is the effective population size. The expression for the fixation time can be determined as well, but is too complicated to include here.

Combining these it is found that in order to achieve a large increase of functionality (approximated by the fitness) the most likely scenario consists of a medium number of mutations with a medium fitness effect (i.e around 0.1). Larger mutations are exceedingly rare, while smaller mutations tend not to reach fixation in the population. This is in line with what has been proposed by biologists (Eyre-Walker & Keightley, 2007). The precise relation between the distribution of beneficial mutations,  $\alpha$ , the fitness improvement to reach,  $I$  and the needed number of mutations  $m$ , is given in figure 1. Estimating the duration of these scenarios is still a challenge, but even small mutations reach fixation relatively quickly – if it does not disappear from the population – and the time it takes for a mutation to reach fixation depends more on population size than on the strength of the fitness effect. Multiple mutations can evolve in parallel, so the time it takes is about equal to a small multiple of the time it takes for the slowest mutation to reach fixation.

Among other things, these results show that the minimalist account of language is not supported by evolutionary theory. Minimalism was proposed because it was assumed language emerged suddenly and recently. However, our analyses show that even a multi-mutation scenario does not need to take a long time. In other words, complex traits can evolve rapidly if population sizes are not too large, and if the selection pressures remain constant.

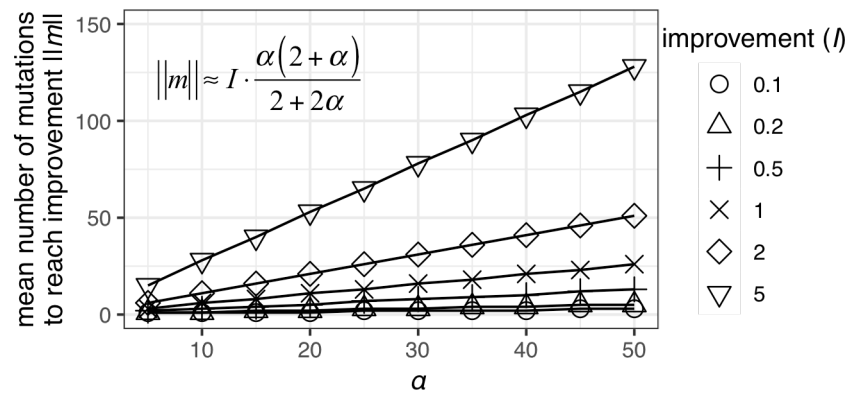


Figure 1. Most likely number of mutations  $\|m\|$  needed to reach a fitness improvement  $I$  for different values of the mutation parameter  $\alpha$ . Realistic values of  $\alpha$  are higher than 30.

## References

- Berwick, R. C., & Chomsky, N. (2016). *Why only us: Language and evolution*. MIT press.
- Blythe, R. A., & McKane, A. J. (2007). Stochastic models of evolution in genetics, ecology and linguistics. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(07), P07018.
- de Boer, B., Thompson, B., Ravignani, A., & Boeckx, C. (to appear). Evolutionary Dynamics Do Not Motivate a Single-Mutant Theory of Human Language. *Scientific Reports*.
- Eyre-Walker, A., & Keightley, P. D. (2007). The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8(8), 610.
- Hurford, J. R. (2014). *The Origins of Language: A Slim Guide*. Oxford: Oxford University Press.
- Kimura, M. (1957). Some problems of stochastic processes in genetics. *The Annals of Mathematical Statistics*, 28(4), 882–901.
- Kimura, M., & Ohta, T. (1969). The Average Number of Generations until Fixation of a Mutant Gene in a Finite Population. *Genetics*, 61(3), 763–771. Retrieved from PMC. (PMC1212239)
- Orr, H. A. (2003). The distribution of fitness effects among beneficial mutations. *Genetics*, 163(4), 1519–1526.