Evolution as Explanation:
The Origins of Neural Codes and their Efficiencies

Han Kim ✉ & Bingni W. Brunton

2021 December 3

Abstract

Neural codes appear efficient. Naturally, neuroscientists contend that an efficient process is responsible for generating efficient codes. They argue that natural selection is the efficient process that generates those codes. Although natural selection is an adaptive process, evolution itself, is not. Evolution consists of not only natural selection, but also neutral stochastic forces that can generate biological inefficiencies. The explanatory power of natural selection cannot be appealed to, without regards for the remaining evolutionary forces. In this paper, we aim to reformulate the explanatory role of evolutionary forces on neural coding, with special attention to neutral forces. We propose a framework that argues for differing contributions of adaptive and stochastic evolutionary forces, for different phenotypic 'levels', including those of neural codes. We assert that this framework is of special interest to neuroscience, because the field has derived much progress from an efficiency-based worldview. We advocate for a pluralistic neuroscience capable of appealing to both adaptive and non-adaptive explanations.

1 Introduction

Neuroscience seeks pluralistic explanations [24, 41, 18, 14]. In doing so, the field borrows from the ethos of multiple disciplines. From engineering-oriented fields, neuroscientists have learned to treat neural systems as reverse-engineering problems. Such approaches hypothesize overarching cost functions or goals, for neural systems, and test those hypotheses against neurobiological observations. Although an optimization-based worldview can be useful, biological features do not possess inherent goals or purposes [13, 25, 9, 19]. The tenets of evolutionary biology are consistent with this agnosticism, and accordingly, are incommensurable with an optimization-based worldview. For this reason, neuroscience's attempts to incorporate both the ethos of evolutionary-biological thinking and those of engineering disciplines have not been straightforward. This paper examines the nature of this pluralism in neuroscience. We argue that the current pluralism is ill-fitting, and reduces the evolutionary process to nothing more than some teleological rendition of natural selection. In response, we attempt to restructure the way in which neuroscientists invoke both optimization-based and historical process-based explanations, in the face of empirical results.

In the current pluralism, neuroscience appeals to only a single evolutionary force—natural selection—to explain neurobiological phenomena. Namely, neuroscientists interpret natural selection as the process that generates proposed neurobiological efficiencies, such as neural codes [5, 3, 37]. In appealing to natural selection alone, however, neuroscientists not only curtail the full explanatory power of evolutionary-biological thinking, but at best, recast evolution as a force limited to natural selection. In actuality, evolution is not just natural selection, but consists also of neutral stochastic forces. We argue that these neutral forces can, in some cases, provide more useful explanations of neurobiological phenomena than optimization frameworks.

On this note, we must mention, at the outset, an important corollary. This paper maintains that adaptive goal-oriented explanations are extremely useful. The hypothesizing of efficient codes yields much progress for neuroscience. Indeed, an automated appeal to adaptive explanations can be a sufficient heuristic, when thinking about the origins of neurobiological features. Like all types of explanations,

1

however, optimization-based explanations have epistemological shortcomings. Namely, the optimization thesis is unfalsifiable, because we can interpret all observable phenomena in optimization terms. Either we have already derived a fitting cost function that the observations appear to match, or if not, we can contend that there still exists some appropriate cost function that has not yet occurred to us. Our position is that when faced with the latter case, the more useful approach—or at least an approach that merits serious consideration—is to interpret the observed phenomenon as the result of a historical process agnostic to any speculative goal or function. We contend that neutral evolutionary forces are one such explanatory process. Importantly, we acknowledge that historical non-adaptive explanations are also unfalsifiable. All empirical observations can be interpreted as the result of some preceding events from its history. In other words, this paper is not about denigrating one type of explanation, in favour of another—both have merits. Rather, we are concerned with deriving appropriate contexts for appealing to each explanatory class.

We begin the paper by examining the history of the efficient coding hypothesis. We argue that the hypothesis tacitly comprises of two types of conflated explanations: a goal-based explanation, typical of engineering disciplines, and a historical process-based explanation, typical of evolutionary biology. We contend that making this distinction explicit is worthwhile. It reveals how neuroscientists currently view the explanatory capabilities of evolution. We then articulate the explanatory usefulness of non-adaptive evolutionary forces. We show that under certain conditions—namely, when species possess small effective population sizes—neutral stochastic forces dominate a species' evolutionary trajectory. We provide established examples from molecular biology and genomics that convey the explanatory power of non-adaptive forces. We then survey the contrasting efficiencies observed in neural coding. Given the overwhelming evidence that viewing neural codes in terms of efficiency is fruitful, we ask under which circumstances we should incorporate non-adaptive explanations. We articulate one evolutionary biological framework that considers both adaptive and non-adaptive explanations. We argue that the appeal to either adaptive or non-adaptive explanations should depend on both the phenotypic 'level' that one is interested in, as well as the species' effective population size. In doing so, this framework is consistent

with both the efficient codes argued in systems neuroscience, as well as the non-adaptiveness conveyed in genetic codes. We note that this framework can be applied to the thinking about phenotypic evolution across the general biologies. We contend, however, that it is of special interest to neuroscience, because unlike other biological disciplines, an efficiency-based worldview has provided neuroscience with immense progress. We provide neurobiological examples consistent with the proposed framework. We close in support of a pluralistic neuroscience that accommodates for both adaptive and non-adaptive explanations.

2 Two types of explanations and the efficient coding hypothesis

Thinkers of the natural world contend that multiple types of explanations can explain a given observation [1, 2, 24, 41, 14]. Although there exists a rich history on the divvying of explanatory types, in this paper, we will categorize explanations of the biological world in one of two classes. The first is functional. Biological features are the way they are because they help achieve some supposed cost function or goal. This type of explanation is typical of engineering disciplines and can be immensely useful. They are popular with systems neuroscientists. The second type of explanation is historical. Biological features are the way they are because of preceding events in their generative process. This type of explanation is traditional to developmental biologists and evolutionary biologists, and can also be immensely useful. It has also found its way into systems neuroscience. Recognizing this explicit distinction between explanation types is necessary for understanding how neuroscientists use evolution to explain phenomena.

An examination of the efficient coding hypothesis reveals how neuroscience employs both function-based and historical process-based explanations. In Barlow's original articulation of the hypothesis, he asks "what are sensory relays for?" [5] Since then, the hypothesis has successfully explained many empirical observations, and has spread to other aspects of neural coding. It can be found in relation to not only early sensory processing, but also for more central neural processes [30, 34, 6], as well as motor efferents [15, 40, 28]. In all these cases, the spirit of the original hypothesis remains intact: what is the neu-

ral code *for*? In other words, the hypothesis squarely concerns itself with devising a function-based explanation, and in doing so, assumes an implicit goal for the nervous system's properties. We emphasize that there is nothing wrong with this position. It yields much explanatory insight for many neurobiological phenomena.

The logic behind the efficient coding hypothesis, however, is not limited to a function-based explanation. It also tries to explain *how* an organism might end up with an efficient code of sorts. To do so, it first answers its original question, "what are sensory relays for?" It claims they are for representing the animal's natural environment. It hypothesizes that the animal likely does so 'efficiently'—perhaps by using minimal neural impulses—in that the animal best represents its natural environmental statistics, over some non-ecological statistics.

Given that efficient codes exist to optimally represent the animal's natural environment, the efficient coding hypothesis then returns to ask how such a code could arise. It seeks to link the natural environment to the qualities of the organism. To do so, it very intuitively invokes the process of natural selection. The efficient coding hypothesis argues that, over historical time, the process of natural selection alters the structure of a species' neural computations such that it efficiently represents its native environment. In other words, the hypothesis invokes a historical process to explain how an organism ends up with an efficient code. It explains neurobiological observations via a function-based explanation, and it employs a process-based explanation to explain how exactly the code meets its function.

This logic, however, risks a nuanced but meaningful misconception. Natural selection, although an adaptive process, is explicitly not goal-oriented. In the words of evolutionary biologist, Ernst Mayr, natural selection and other "[h]istorical processes ... can *not* act purposefully" [24]. The efficient coding hypothesis' invoking of a non-goal-oriented process, in order to explain how something achieves its supposed goal, risks recasting the natural selection process in teleological terms. Indeed, we think that as a result of the coding hypothesis' articulation, many neuroscientists tacitly interpret natural selection as a goal-oriented process. Although we reject the notion that natural selection could be at all purpose-driven, we do not disagree with the idea that natural selection contributed to the rise of purported efficient codes in the peripheral nervous systems. Natural selection, as it

was originally conceived—as an undirected and iterative process that imposes ecological constraints on a species—perhaps provides the best explanation for many efficient neural codes. We simply wish to articulate the necessary evolutionary biological caveat that has been amiss from the incarnations of the efficient coding hypothesis: although natural selection might explain efficient codes, natural selection is not an efficient process with respect to some goal. Goals and adaptive processes are two different things.

We contend that the overlooking of this subtlety persists in the many articulations and applications of the efficient coding hypothesis, and is the culprit behind neuroscience's limited means of appealing to evolution as an explanation. In the worst of such cases, neuroscientists invoke natural selection as an epistemological panacea. They use it to explain how a neural system wound up meeting whatever cost function objective they are interested in, even when the code itself does not appear to be particularly efficient. For example, several decades after Barlow published the efficient coding hypothesis, he revisits his formulation, and argues that he was wrong in overemphasizing the role of 'compression' in an efficient code, and contends that redundancy in neural coding is a crucial function for efficient nervous systems [3]. Although we agree with this reasonable amendment, we disagree with Barlow's continued, and if anything, more explicit appeal to natural selection as something that explains a redundant neural code. The continued allegiance to natural selection in Barlow's revisit may be a vestige from the fact that his original efficient coding hypothesis invoked an adaptive evolutionary process, albeit under more sensible circumstances. Alternatively, or perhaps additionally, Barlow and other neuroscientists are unaware of alternative evolutionary forces that may better explain their phenomena. In the next section, we introduce such alternative forces. We articulate how evolution is more than just natural selection, and we describe how non-adaptive neutral forces can sometimes explain phenomena more usefully than an optimization-based framework.

3 Evolutionary biology usefully explains inefficient codes

Even in the face of biological features that we might intuitively interpret as inefficient—such as redundancy in neural representations—we can still appeal

to goal-oriented explanations. For example, we can argue that the observed redundancy aids the system in achieving the goal of robustness, perhaps from noise or from external perturbations. More generally speaking, we can always formulate a goal-oriented explanation about an observation because we can claim that either the observed inefficiency is exactly the point of the system, or because we can argue that we do not yet know the goal of the system, but the system most certainly has a goal, and with further empirical evidence, such a goal will become clear. We contend that adopting these epistemological positions is limited. As an alternative, we introduce non-adaptive evolutionary forces, which we contend are more useful than efficiency-based frameworks for explaining inefficiencies. We begin by introducing the explanatory power of non-adaptive forces in its native context. The goal of this section is, in part, pedagogical. We aim to explain key evolutionary biology concepts, without being mired by technical derivations.

3.1 The effective population size defines drift

Evolution is not natural selection. Rather, it consists of four forces, of which natural selection is just one. The other three—mutation[1], recombination[2], and genetic drift[3]—are non-adaptive and neutral because they are random and do not depend on an animal's fitness properties. Under what conditions, then, will neutral forces provide the most explanatory power regarding a biological feature's origins? We begin with drift.

Sewall Wright, a forefather of population genetics, first sketched the concept of the effective population size (N_e) to provide a way of calculating the rate of evolutionary change caused by the random sampling of allele[4] frequencies—that is, genetic drift[42]. An intuition for N_e, and how it differs from the 'regular' population size of breeding individuals (N), is essential for understanding neutral forces. N_e is best understood in a highly idealized but important type of random sampling called the Wright-Fisher

population[42, 10, 11]. The Wright-Fisher population assumes a randomly mating population that consists of a number of diploid[5] hermpahroditic individuals. In the Wright-Fisher population, these individuals are the total number of breeding individuals, N. They reproduce with discrete generations, and each generation is counted at the time of breeding. The individuals of each new generation are the result of random sampling, with replacement, from the gametes of the parents, and the parents die immediately after reproduction. If we accept all the idealized assumptions of the Wright-Fisher population, then the rate at which drift, or random sampling, results in a differing genetic composition in the offspring is $\frac{1}{2N}$. The coefficient '2' comes from the fact that the Wright-Fisher population assumes that the parents are diploid, so each individual actually has two inheritable genetic sets. If the members of the Wright-Fisher population were haploid or triploid, then the rate of change from drift across the discrete generations would be $\frac{1}{N}$ or $\frac{1}{3N}$, respectively. In other words, the goal of the Fisher-Wright population is to identify a feasible set of biological assumptions that result in the most simplified expression for genetic drift. This theoretical concept provides us with an intuition for drift, and as we will see, for N_e.

Most biological populations, however, do not reproduce in a fashion similar to the sampling from a bag of marbles. For example, many species have two sexes, select mates non-randomly, and will not produce the progeny of the next generation all at once. For these more complex situations, we invoke the concept of the *effective* population size, N_e. The effective population size has, within it, all the relevant biological complications that distinguish this population from a Wright-Fisher population, such that we can simply substitute N for N_e, for most problems we wish. So, for example, in a diploid non-Wright-Fisher population, we might say that the rate of change in genetic composition due to drift is $\frac{1}{2N_e}$ instead of $\frac{1}{2N}$. In other words, we defer. The consideration for any biological complications when working with a given population genetic expression originally meant for a Wright-Fisher population is offloaded from that particular expression and its use of N, to its substitution for N_e. Many theoretical and computational methods, such as coalescent theory, exist for inferring N_e for a population[7], but they are outside the scope

---

[1] A novel change in genetic sequence. In nature, is the result of a random process.

[2] The rearrangement of genetic material. In nature, is the result of a random process.

[3] The random sampling of genetic material from the parental generation, to produce the offspring generation. Results in differing genetic compositions between the parental and offspring generations.

[4] One of two or more versions of a gene. They arise by mutation.

[5] A cell or organism with two complete sets of genetic information. If the cell or organism has only one complete set of genetic information, it is haploid.

of this paper. The reader needs only know that the effective population size, and more specifically, $\frac{1}{N_e}$, approximates the neutral force that is drift[6].

3.2 The selection coefficient expresses the strength of natural selection

In order to understand when neutral forces outweigh adaptive forces in evolution, we need to introduce the selection coefficient, s. Doing so requires an understanding of fitness. In the broadest sense, fitness involves the ability of organisms or populations to survive and reproduce in their environment[31]. The concept of fitness is important for natural selection and the selection coefficient, because natural selection requires differing level of fitness across members of a population, and that some of those fitness differences be inheritable—that is, that they have a genetic basis. For this reason, we will express our understanding of fitness in more specific genetic terms.

Evolutionary biologists distinguish between two types of fitness metrics[31]. One is the absolute fitness, W. It refers to a genotype's expected total fitness, and encapsulates the complexities that come with integrating all imaginable biological properties, such as viability, mating success, and fecundity. W must be greater than or equal to zero. The other type of fitness, which is more commonly used in evolutionary biology, is relative fitness, or w. The relative fitness of a genotype is simply its absolute fitness, but normalized in some way. Its most common normalization is dividing the absolute fitness of a genotype by the absolute fitness of the fittest genotype. In doing so, w is always bound between 0 and 1, where 1 is the fitness of the fittest genotype. From the definition of w, we can easily understand the selection coefficient, s. It is simply a genotype's w, relative to the fittest genotype's w. That is, $s = 1 - w$, where w is the relative fitness of some genotype. So if some genotype is incredibly fit and has $w = 1$, then for that genotype, $s = 0$, and selection against that genotype is non-existent. Conversely, if some genotype is extremely deleterious and has $w = 0$, then for that genotype, $s = 1$, and selection against that genotype is total. The genotype will contribute nothing to the subsequent generation. From these examples, we can see how the value of s expresses the strength of natural selection. Like many terms in evolutionary biological theory—such as N_e—several methods exist for empirically estimating s[31], but they are outside the scope of this paper.

3.3 Random forces can dominate evolution

From our above definitions of drift and natural selection, we can formulate an expression for evolution that considers the contributions of both non-adaptive and adaptive forces. Given that we can articulate the strength of drift as $\frac{1}{N_e}$, and the strength of natural selection as s, we can define their relative contributions as $\frac{\text{selection}}{\text{drift}} = \frac{s}{\frac{1}{N_e}} = N_e s$.

To appreciate the $N_e s$ ratio, we must consider a classic case of a diploid population species, such as *Drosophila melanogaster* or *Mus musculus*. In such a species, a given gene might have two possible alleles, **A** and **a**, where allele **A** might have a slight selective advantage, s, over **a**. Given that one allele is advantageous over the other, a worthwhile question that directly relates to our original intents, is to ask under which scenario the advantageous version of the gene, **A**, will become predominant, or fixate, in the population. In other words, given the selective advantage that **A** has over **a**, will natural selection always ensure the fixation of **A**? The answer to this question is no. Let's say that the mutation rate of **a** $\rightarrow$ **A** is m times greater than the mutation rate of **A** $\rightarrow$ **a**. If so, then standard population genetics theory[16, 19] argues that the probability of fixing a mutation to **A** is $e^{2N_e s}$ times more likely than fixing a mutation to **a**. That is, the ratio of probabilities of being **A** versus **a** is $m e^{2N_e s}$. This relationship is particularly interesting, because when $\frac{1}{N_e} \gg |s|$, the probability of fixing **A** over **a** in the population approaches exactly the ratio of mutation rates, m. That is, when $\frac{1}{N_e} \gg |s|$, or said another way, when the effective population size, N_e, is extremely small, then the likelihood of fixation for some advantageous allele is a function of mutation pressures alone. Under small N_e, random forces dominate the course of a population's evolution. We can plot this relationship, and observe the effects of altering the $N_e s$ ratio [TODO: CITE FIG]. Small $N_e s$ ratios result in evolutionary trajectories that are nothing more than the product of mutation rates, a neutral stochastic force.

---

[6]Some authors distinguish between various types of effective population sizes, such as N_e, N_l, and N_g, which all take into account different factors such as timescales, genome linkage effects, and the ploidy of a species [19, 20]. For our purposes, we will not be making such distinctions, because they do not matter for our arguments with the reader. We will use exclusively N_e when referring to the effective population size.

3.4 Empirical support for predictions on biological coding inefficiency

From the above pedagogy, one might ask what sorts of populations or species possess small effective population sizes? Such species are likely more susceptible to neutral forces than species with larger N_e values. To answer this question, we must introduce another parameter from evolutionary biology: $N_e\mu$. This composite parameter is the effective population size, N_e, scaled by the mutation rate, μ. It determines the standing equilibrium level of approximately neutral genetic variability in a population[16, 21, 7]. Biologists can estimate this value empirically, by quantifying the extent of synonymous mutations[7] in a species, because synonymous mutations are neutral [21]. Given that across species, the mutation rate per base per cell division, μ, ranges from 5×10^{-11} to 5×10^{-10}, that is, by a single order of magnitude, we can more or less factorize and interpret $N_e\mu$ as N_e. Doing so results in estimates of N_e for prokaryotes that is at least 10^8, for unicellular eukaryotes that is 10^7 to 10^8, for invertebrates that is 10^5 to 10^6, and for vertebrates that is 10^4 to 10^5 [21]. A plot of $N_e\mu$ estimates for various species is shown in Figure [TODO: CITE FIG]. In other words, the effective population sizes of nervous system-possessing animals is far smaller than those without. They are more susceptible to random genetic drift than their neuron-less counterparts. Within nervous system-possessing animals, we see that invertebrates are less privy to random drift than vertebrate animals[8]. We also note that, in general, estimates of N_e are much smaller than estimates of the true number of breeding individuals, N [20, 8, 12]. In other words, most biological populations do not merely differ from Wright-Fisher population dynamics, but do so in a way that promotes the effects of drift.

We have provided both theoretical and empirical evidence for thinking that non-adaptive evolutionary forces are at least worth acknowledging, when thinking about nervous system evolution. In addition, however, observed biological features can also make more sense under a view of evolution that incorporates stochastic forces. There are several established examples in relation to genetic coding. Prokaryotes possess the largest N_e values, and accordingly, their genome architectures are remarkably efficient. Nearly all of their genomes are dedicated to coding proteins and they possess virtually no non-coding sequences, such as introns and non-coding RNA [20, 27]. In addition, genes that function in common biological processes are found adjacent to each other in the genome, such that the organism can readily achieve co-expression of those related genes [33]. Prokaryotic genes can also be polycistronic, meaning that a single mRNA transcript can encode multiple proteins at a time [17].

In contrast, genetic coding in multicellular eukaryotes appear to be anything but efficient. Eukaryotes are monocistronic, and so more energy and molecular machinery must be expended to transcribe and translate a given coding unit [17]. Moreover, the majority of multi-cellular eukaryotic genomes are non-coding elements. Less than 2% of human genes are protein-coding, whereas 90% to 100% of prokaryotic genomes encode for functional units [27]. The dynamics of coding in multi-cellular organisms is also remarkably inefficient. Most of the eukaryotic genome is transcribed, and yet only a small fraction of those transcripts undergo maturation and subsequent translation into proteins [26]. Transcripts not processed for maturation and translation are digested and decayed [29]. The large investment of energy in producing intermediary products from nearly all of the genome, only to have most of those intermediates degraded, illustrates a highly inefficient biological coding process.

Evolutionary biology can explain these observations without necessarily invoking efficiency and natural selection [19]. Seemingly inefficient biological qualities are nothing more than the outcomes contributed by random sampling processes. Populations with smaller N_e are simply more prone to such processes. In contrast, when faced with putative inefficiencies, optimization-based frameworks such as those seen in engineering disciplines are resigned to one of the two positions mentioned at the very beginning of this section: either the observed inefficiencies are, for some reason, exactly the point of the system, or that the analyst has wrongly identified the goal of the system. We maintain that these two positions are epistemologically limited. They do not explain the observed phenomena by appealing to some scientific process, but rather, defer the explanation to some function that is not yet understood or known.

---

[7]Mutations in the genetic sequence that do not result in the encoding of different amino acids. For example, both GCA and GCG genetic triplets, or codons, encode for the alanine amino acid, even though the sequences differ by a base pair. The GCA $\rightarrow$ GCG mutation is synonymous or silent.

[8]An interesting observation for the invertebrate neuroscientist. Its consequences are slightly outside the scope of this paper.

We reassert, however, that optimization-based frameworks still have an essential place in biology and neuroscience. To demonstrate this point, we provide a brief review of the overwhelming evidence in favour of an optimization framework in neuroscience.

4 Efficient codes and nervous systems

Given the relatively small effective population sizes seen in nervous system-possessing animals, we might expect abundant inefficiencies across neural codes. This hypothesis, however, appears to be exceptionally false. Optimization frameworks can clearly provide much explanatory power, especially in neuroscience. In this section, we provide the reader with diverse examples from across the nervous system that support the epistemological benefits of assuming purposes in natural systems.

4.1 Sensory systems

In feline, human, and grasshopper auditory systems[22, 39], sensory systems appear to use a sparse spike code to represent the acoustic structure of a given stimulus. Kernel functions for an optimal sparse representation learned from auditory stimuli closely approximate the physiological reverse-correlation filters [39]. Importantly, these efficient codes only succeed in using a sparse representation when the auditory stimulus is derived from natural scene statistics. In other words, efficient neural codes appear to at least be correlated with the germane particulars of the animal's natural environment. One reasonable interpretation to the exquisite match between an animal's native environment and the neural code that processes that environment is that the nervous system is somehow tuned or optimized for the environment. A simple process-based and folk-biological explanation of these findings might be that selective forces encompass ethological cues, and animals that fail to efficiently respond to those cues exhibit survival-inappropriate behaviours, and are selected against. Indeed, efficient codes tend to match not only the ethological sensory space, but are especially tuned for those stimuli that are behaviourally relevant[22, 23].

We observe theoretical and experimental evidence for efficient neural codes in modalities apart from audition. For example, a survey of ommatidia diameter and eye height from 27 *Hymenoptera* species suggest that their facets have evolved to maximize visual sampling, while minimizing the blur from diffraction [4]. This conclusion derives from a fundamental physics principle showing that the optimal resolving power of a lens with diameter d is equivalent to $\sqrt{a\lambda}$, where a is the angle subtended by two point sources that can still be detected as double, and λ is the wavelength of incoming light. The surveyed insects indeed possess a linear and proportional relationship between $\sqrt{a}$ and d [4]. In terms of neural coding, the retinal ganglion cells of both salamanders and macaques decorrelate spatial features of visual inputs, as a means of achieving sparse spiking in the retina and optimizing visual coding efficiency [34].

4.2 Motor systems

We also find efficiency at the opposite end of the periphery, although not necessarily with respect to neural coding. Instead, most examples of efficient motor control are seen in terms of the amount of energy expended in muscles, to achieve a particular task. Given that the amount of muscle energy required for a behaviour vastly exceeds those needed for neural spiking [32, 36], one can argue that the relevant cost function at the motor periphery should relate to the joules of muscle work expended, rather than some sparse spike code. Classic work on horse gaits show that freely moving horses self-optimize their movement speeds. Under naturalistic conditions, they largely operate in locomotor regimes that require minimal amounts of consumed lxygen for moving some unit distance. These results hold across multiple gait types, such as walks, trots, and gallops [15]. Similar results of self-optimization with respect to gross energy expenditure have also been reported in human ergometer studies [40].

Despite the foremost relevance of muscle expenditure over neural spiking for motoric efficiency, some recent evidence suggests that the encoding of naturalistic behaviours may also have some efficient basis. Simultaneous microdrive recordings and 3D posture tracking of the head, neck, and back of freely moving rats reveals that proportionately fewer cells in the posterior parietal and motor cortices fire when the animal partakes in common 'default' postures [28]. The encoding of naturalistic poses appears to use a minimal number of cells to represent the repertoire of ethological motor ensembles.

4.3 Central systems

Further examples of efficiency have been argued in deeper processing regions, albeit often with different, and arguably more ambiguous, objectives from the sparse coding seen at the periphery [38]. Learning algorithms that maximize sparseness, for example, succeed in recapitulating the spatially localized, oriented, and bandpass features of mammalian visual cortical cells [30]. In general, sparse codes in the visual cortex appear to be useful for learning and processing incoming spike patterns [30], or for parsing large amounts of signal from background noise [35], in the face of overcomplete representations. Even though these kinds of objective functions differ from those efficient representations seen in the periphery, where the goal is to minimize the number of spikes needed for expressing some maximal description of the environment, both scenarios still offer examples of efficient coding.

Operating on the assumption that nervous systems have innate goals or functions is undeniably useful for neuroscience. An appeal to natural selection as the historical process that explains these findings is also not problematic. In fact, it is likely the most appropriate and useful type of process-based explanation for these observations. In light of the previous section's pedagogy, however, the reader may ask themselves two questions: first, why do small N_e species, such as those with nervous systems, possess efficient neural codes, and second, whether there is a general place or context under which an appeal to neutral evolutionary processes is most appropriate. We directly address this question in the next section.

5 An account of phenotypic evolution

Nervous system-possessing animals have, relatively speaking, the smallest N_e values. These values do not, however, suggest that those species are impervious to natural selection. The previous section's survey of efficient neural codes argues, if anything, that natural selection is well at work for nervous system-possessing animals. Indeed, we remind the reader that a given N_e value should only be considered small relative to some selection coefficient, s, for that species, and that quantifying s for the heritable basis of a species' neural coding scheme is, like for many complex traits, incredibly challenging and unknown. Rather, the purpose of highlighting the small

N_e values of nervous system-possessing animals is to argue that some species are less, but not totally, immune to stochastic evolutionary processes. As we saw in Section 3.4, unicellular eukaryotes, despite having one of the highest N_e values, still exhibit inefficiencies in their genetic coding schemes not seen in their higher N_e counterparts, such as the retention of genomic non-coding elements. For this reason, we think it is not unreasonable to expect the occasional inefficiency from species with the lowest N_e values, like those with nervous systems. In this section, we outline a framework that will provide us with intuitions for when we might expect a biological trait to be the function of a stochastic historical process. We realize that this framework may be broadly applicable for biological observations outside of the neural sciences, but we think its impact might be particularly useful for neuroscience, because of the immense success that an efficiency-based worldview has had for the field.

References

[1] *Physics.* Dover Thrift Editions. Dover Publications, 2017.

[2] Aristotle and H. Lawson-Tancred. *The Metaphysics.* Classics Series. Penguin Publishing Group, 1998.

[3] H. Barlow. Redundancy reduction revisited. *Network: Computation in Neural Systems*, 12(3):241–253, 2001.

[4] Horace B Barlow. The size of ommatidia in apposition eyes. *The Journal of Experimental Biology*, 29:667–674, 1952.

[5] Horace B Barlow. Sensory communication - possible principles underlying the transformation of sensory messages. page 216234, 1961.

[6] Bingni W. Brunton, Matthew M. Botvinick, and Carlos D. Brody. Rats and Humans Can Optimally Accumulate Evidence for Decision-Making. *Science*, 340(6128), 2013.

[7] Brian Charlesworth. Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3):195–205, 2009.

[8] James F Crow and Newton E Morton. Measurement of Gene Frequency Drift in Small Populations. *Evolution*, 9(2):202, 1955.

[9] Theodosius Dobzhansky. Nothing in Biology Makes Sense Except in the Light of Evolution. *The American Biology Teacher*, 75(2):87–91, 1973.

[10] R. A. Fisher. Xxi.on the dominance ratio. *Proceedings of the Royal Society of Edinburgh*, 42:321341, 1923.

[11] R. A. Fisher. Xvii.the distribution of gene ratios for rare mutations. *Proceedings of the Royal Society of Edinburgh*, 50:204219, 1931.

[12] Richard Frankham. Effective population size/adult population size ratios in wildlife: a review. *Genetical Research*, 66(2):95–107, 1995.

[13] S J Gould and R C Lewontin. The spandrels of san marco and the panglossian paradigm: a critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B, Biological sciences*, 205(1161):58198, 1979.

[14] Vojtch Hladk and Jan Havlek. Was Tinbergen an Aristotelian? Comparison of Tinbergen's Four Whys and Aristotle's Four Causes. *Human Ethology Bulletin*, 28(4):3–11, 2013.

[15] Donald F. Hoyt and C. Richard Taylor. Gait and the energetics of locomotion in horses. *Nature*, 292(5820):239–240, 1981.

[16] Motoo Kimura. *The Neutral Theory of Molecular Evolution.* Cambridge University Press, 1983.

[17] Marilyn Kozak. Initiation of translation in prokaryotes and eukaryotes. *Gene*, 234(2):187–208, 1999.

[18] John W Krakauer, Asif A Ghazanfar, Alex Gomez-Marin, Malcolm A MacIver, and David Poeppel. Neuroscience Needs Behavior: Correcting a Reductionist Bias. *Neuron*, 93(3):480 490, 2017-02.

[19] Michael Lynch. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proceedings of the National Academy of Sciences*, 104:85978604, 2007.

[20] Michael Lynch. *The Origins of Genome Architecture.* Sinauer Associates, 2007.

[21] Michael Lynch and John S Conery. The origins of genome complexity. *Science*, 302(5649):14011404, 2003.

[22] Christian K Machens, Tim Gollisch, Olga Kolesnikova, and Andreas V M Herz. Testing the efficiency of sensory coding with optimal stimulus ensembles. *Neuron*, 47(3):447 456, Aug 2005.

[23] Christian K. Machens, Martin B. Stemmler, Petra Prinz, Rdiger Krahe, Bernhard Ronacher, and Andreas V. M. Herz. Representation of Acoustic Communication Signals by Insect Auditory Receptor Neurons. *Journal of Neuroscience*, 21(9):3215–3227, 2001.

[24] Ernst Mayr. Cause and Effect in Biology. *Science*, 134(3489):1501–1506, 1961.

[25] Ernst Mayr. The objects of selection. *Proceedings of the National Academy of Sciences*, 94(6):2091–2094, 1997.

[26] Jerome S Menet, Joseph Rodriguez, Katharine C Abruzzi, and Michael Rosbash. Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. *eLife*, 1:e00011, 2012.

[27] R. Milo, R. Phillips, and N. Orme. *Cell Biology by the Numbers.* Garland Science, 2016.

[28] Bartul Mimica, Benjamin A Dunn, Tuce Tombaz, V P T N C Srikanth Bojja, and Jonathan R Whitlock. Efficient cortical coding of 3d posture in freely behaving rats. *Science*, 362(6414):584589, 2018.

[29] Michaela Mller-McNicoll and Karla M Neugebauer. Good cap/bad cap: how the cap-binding complex determines RNA fate. *Nature Structural & Molecular Biology*, 21(1):9–12, 2014.

[30] Bruno A. Olshausen and David J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381(6583):607–609, 1996.

[31] H. Allen Orr. Fitness and its role in evolutionary genetics. *Nature Reviews Genetics*, 10(8):531–539, 2009.

[32] Justus O. Ortega, Stan L. Lindstedt, Frank E. Nelson, Sharon A. Jubrias, Martin J. Kushmerick, and Kevin E. Conley. Muscle force, work and cost: a novel technique to revisit the Fenn effect. *Journal of Experimental Biology*, 218(13):2075–2082, 2015.

[33] Ross Overbeek, Michael Fonstein, Mark DSouza, Gordon D. Pusch, and Natalia Maltsev. The use of gene clusters to infer functional coupling. *Proceedings of the National Academy of Sciences*, 96(6):2896–2901, 1999.

[34] Xaq Pitkow and Markus Meister. Decorrelation and efficient coding by retinal ganglion cells. *Nature Neuroscience*, 15(4):628635, 2012.

[35] Dario L. Ringach and Brian J. Malone. The Operating Point of the Cortex: Neurons as Large Deviation Detectors. *The Journal of Neuroscience*, 27(29):7673–7683, 2007.

[36] Biswa Sengupta, Martin Stemmler, Simon B. Laughlin, and Jeremy E. Niven. Action Potential Energy Efficiency Varies Among Neuron Types in Vertebrates and Invertebrates. *PLoS Computational Biology*, 6(7):e1000840, 2010.

[37] E P Simoncelli and B A Olshausen. Natural image statistics and neural representation. *Annual review of neuroscience*, 24(1):1193–216, 2001.

[38] Eero P Simoncelli. Vision and the statistics of the visual environment. *Current Opinion in Neurobiology*, 13(2):144–149, 2003.

[39] Evan C. Smith and Michael S. Lewicki. Efficient auditory coding. *Nature*, 439(7079):978–982, 2006.

[40] W. A. Sparrow and K. M. Newell. Metabolic energy expenditure and the regulation of movement economy. *Psychonomic Bulletin & Review*, 5(2):173–196, 1998.

[41] Niko Tinbergen. On aims and methods of ethology. *Zeitschrift fr Tierpsychologie*, (20):410–433, 16 March 1963.

[42] Sewall Wright. Evolution in mendelian populations. *Genetics*, 16(2):97–159, 1931.