Kludged[*]

October 5, 2010

Abstract

Is there reason to believe that our brains have evolved to make efficient decisions so that the details of the internal process are irrelevant? I develop a model which illustrates a limitation of adaptive processes: improvements tend to come in the form of *kludges*. A kludge is a marginal adaptation that compensates for, but does not eliminate fundamental design inefficiencies. When kludges accumulate the result can be perpetually sub-optimal behavior even in a model of evolution in which arbitrarily large innovations occur infinitely often with probability 1.

Keywords: kludge.

---

[*] I thank Andy Postlewaite and three anonymous referees for suggestions that improved (drastically) the quality of this paper. A previous version has been reviewed in NAJ economics. See http://www.najecon.org/naj/v14.htm

1 Introduction

In July of 2004, Microsoft announced that the release of Vista, the next generation of the Windows operating system, would be delayed until late 2006. Jim Allchin famously walked into the office of Bill Gates and proclaimed, "It's not going to work." Development of Windows had become unmanageable and Allchin decided that Vista would have to be re-written essentially from scratch.

> Mr Allchin's reforms address a problem dating to Microsoft's beginnings. . . . PC users wanted cool and useful features quickly. they tolerated – or didn't notice – the bugs riddling the software. Problems could always be patched over. With each patch and enhancement, it became harder to strap new features onto the software since new code could affect everything else in unpredictable ways.[1]

The Alternative Minimum Tax was introduced by the Tax Reform Act of 1969. It was intended to prevent taxpayers with very high incomes from exploiting numerous tax exemptions and paying little or no tax at all. Over time, the shortcomings of the AMT as a solution to the proliferation of exemptions have begun to appear. However, over this same time, the federal tax and budgeting system has come to depend on the AMT to the point that many observers think that changing the AMT, without complicated accompanying adjustments elsewhere, would be worse than leaving it as is.

Flat fish inhabit the sea floor, but for many this was not the original habitat. When their ancestors moved to the sea floor, they adapted by changing their orientation from swimming "upright" to on their sides. Given their existing bone structure, this was the only way to become "flat", but it rendered one eye useless. So, by a further adaptation many of today's species of flatfish migrate one eye to the opposite side of their body during development.[2]

As beautifully documented in the film *The March of the Penguins*, emperor penguins spend a nearly 9 month breeding and nurturing cycle which involves walking up to 100 KM away from any food source in order to avoid predators. The problem for penguins is that they are birds, and hence lay eggs; but they are flightless birds, so they find it inconvenient to move

---

[1] "Code Red: Battling Google, Microsoft Changes How it Builds Software." *The Wall Street Journal*, Robert Guth, September 2005.

[2] For a vivid account of the evolution of bony flat fish, see Dawkins (1986).

to areas where the eggs can be easily protected. They adapted not by rectifying either of these two basic problems,[3] but instead by compensating for them by an extremely costly and risky behavior.

Each of these examples represents a *kludge*: an improvement upon a highly complex system that solves an inefficiency but in a piecemeal fashion and without addressing the deep-rooted underlying problem. There are three ingredients to a kludge. First the system must be increasing in complexity so that new problems arise that present challenges to the internal workings of the system. Second, a kludge addresses the problem by patching up any mis-coordination between the inherited infrastructure and the new demands. Third, the kludge itself– because it makes sense only in the presence of the disease it is there to treat– intensifies the internal inefficiency, necessitating either further kludges in the future or else eventually a complete revolution.[4]

Microsoft Windows is a complex system whose evolution is guided by a forward-looking dynamic optimizer. It is not surprising therefore that, after two decades worth of kludges that accompanied the expansion from DOS to Windows to 32 bit and eventually 64 bit architecture, revolution was the final solution. In the case of the US Tax Code, or for that matter any sufficiently complex body of contracts that govern interactions among diverse interests, while the evolution may be influenced by forward-looking considerations, full dynamic optimization is more tenuous as a model of the long-run trade-offs.

But the story is very different for flat fish and penguins, and, to come to the point, for human brains, whether we are considering the evolution of the brain across generations or the development of the decision-making apparatus within the life a single individual. Here, progress is *adaptive.* An adaptive process is not forward-looking and certainly not governed by dynamic optimization. An adaptive process inherits its raw material from the past, occasionally modifies it by chance (mutation or experimentation), and selects among variants according to success *today.*

Nevertheless there is the possibility, not completely fanciful, that an adaptive process can produce complex systems that perform as well today as those that were designed by an optimizer given the same set of raw materials. Indeed, there is a tradition in economics that accepts the dis-

---

[3] Incidentally, it has happened in evolutionary history that oviparous (egg-laying) species have adapted to vivipary (giving birth to live offspring.) Some species of sharks are important examples. Vivipary enables a long internal gestation so that the developing offspring is protected and nourished within the body of the mother.

[4] See wikipedia for the history, usage, and pronunciation of the word kludge.

tinction between adaptation and optimization, but rationalizes a positive methodology based on unfettered optimization by an appeal to this unwritten proposition.[5]

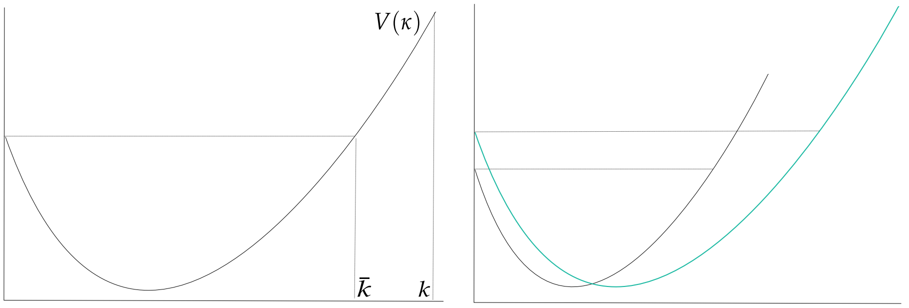
In this paper I present a model which probes that common story. I analyze a simple single-person decision problem. An *organism* is a procedure for solving this problem. I parametrize a family of such algorithms which includes the optimal algorithm in addition to algorithms that perform less well. An adaptive process alters the organism over time, favoring improvements. I show conditions under which no matter how long the adaptive process proceeds, an engineer, at any point in time, working only with the raw materials that presently make up the organism, could eliminate a persistent structural inefficiency and produce a significant improvement. In the model, kludges arise naturally and are the typical adaptations that improve the organism. A kludge always improves the organism at the margin, but also increases both its complexity and its internal complementarity and as a by-product makes it harder and harder for adaptation to undo these inefficiencies in the future.

Illustrative Example The following stylized example illustrates the main ideas. An organism of complexity k has k genes each of which can have value -1 or $+1$. Thus, its genetic code π is a k-length sequence such as $\pi = \{-1, +1, +1, \dots, -1\}$. The fitness of the organism is given by the function $V(\pi)$ which exhibits complementarity. A simple example would be $V(\pi) = V(\kappa)$ where κ is the number of $+1$ genes. If $V(0) = k, V(1) = 2k$, and $V(\kappa)$ is U-shaped then, holding k fixed, there are two local optima: a *negatively aligned* organism (all genes are -1) and a *positively aligned* organism (all genes are $+1$.) Refer to Figure 1.

In each period the organism is, with probability $0 < (1-q) < 1$ subject to random mutation. When a mutation occurs, each gene has an independent probability $\mu > 0$ of changing signs. If the new version of the organism is an improvement (higher V) then it replaces the existing organism. Otherwise the existing organism survives to the next period. In this model, when k is constant over time, with probability 1 a negatively aligned organism will eventually be replaced and never return. It is enough that $\bar{k}$ genes change sign where $V(\bar{k}) > k = V(0)$. But now suppose that k is increasing over time. In particular, suppose that a second form of mutation can occur. With complementary probability q, the complexity increases by 1 and the new gene is determined at random.

---

[5] The classic defense is Friedman (1966).



(a) Drastic mutation for fixed k.

(b) Increased complexity increases the size of a drastic mutation

Figure 1: Kludges in the illustrative example.

If the organism is initially negatively aligned, and the complementarity in V is strong, an increase in complexity survives only when the new gene is similarly aligned. An increase in complexity thus produces an incremental improvement but potentially reduces the probability of the drastic changes needed to overhaul the structure of the organism from negative to positive alignment. If so, then the negatively aligned organism is kludged.

In fact, if a drastic mutation always requires some minimum fraction of genes to switch, then as k grows over time, drastic mutation probabilities shrink exponentially fast. This means that the total probability that a drastic mutation ever occurs is strictly less than 1. Essentially, while at any point in time the probability of jumping between local optima is positive, the distance between the local optima is growing fast enough that the total probability remains bounded away from 1. In particular there is a positive probability that the kludges perpetually accumulate and ultimately prevent the efficient re-alignment from taking place.

The key features in this story are the complementarity in the fitness function V and the rate of increased complexity relative to mutations. This paper builds a model of a decision-making algorithm and its evolutionary process which give rise to these features. In the model, an organism evolves a procedure for processing and encoding data. When the procedure is poorly adapted to the environment, improvements can still come from increased complexity: the organism analyzes additional data but encodes it using the inefficient protocol. Because aggregating the data requires a common encoding, increased complexity increases the dependence on the protocol. Useful aggregation requires, at a minimum, that at least half of the data are encoded consistently with the protocol. Finally, unlike in the simple example, complexity grows at an endogenous rate in the model. As the marginal value of additional data decreases, complexity growth becomes very slow. Indeed the organism waits for roughly k^2 periods before increasing complexity from k to $k + 1$. Nevertheless this is fast enough that with positive probability the kludged protocol survives forever.[6]

1.1 Kludges In The Brain

Neuroscience views evolution as having progressively built higher functioning systems on top of more primitive systems, each time leaving the latter essentially intact.[7] The resulting patchwork often bears all of the hallmarks of a kludge. The brain itself is built of neurons, which as evidenced by their presence in jellyfish and snails are evolutionarily ancient building blocks of neurocircuitry. Unfortunately they don't scale well: they are energy-inefficient, noisy, and slow; hardly the tools one would have chosen to build a complex, conscious brain.

But the most glaring clashes come in the interaction between conscious decision-making and the more primitive systems that provide the data guiding our decisions. For example, the brain's system of passive, context-driven memory is a poor match for a conscious decision-making center which needs arbitrary access to memories. Hence, unlike a computer which recalls memories by their addresses, when we need to find lost keys we "retrace our steps." This is an attempt to manipulate the passive memory system, conjuring up the requisite context in hopes of shaking the lost memory out. "Blindsight" reveals the inefficient integration between primitive visual centers in the brain and more evolutionarily recent visual modules in the cortex. Blind subjects asked to locate an object held before them are, naturally, unable to make any guess with confidence. But when asked to reach for the object, some subjects can consistently locate it and grab it.

1.2 Organism Design

Structurally inefficient decision-makers present a problem not just for positive methodology, but normative as well. Much of welfare economics is founded on revealed preference and agent sovereignty. These principles

---

[6] Sandholm and Pauzner (1998) make a similar point in a coordination game with a growing population. There, even very slow population growth precludes escape from inefficient equilibria.

[7] Linden (2007) and Marcus (2008) are excellent popular accounts.

presume that the choices we observe reveal what benefits the agent. But when the adaptive process creates a wedge between observed behavior and the underlying objective the agent is designed to satisfy, there is a corresponding wedge between revealed preference and true preference. Put differently, if we grant that there is some underlying objective that guides the adaptive process, then at best we can view the organism as an agent whose efforts at achieving that objective are the result of a second-best solution designed by nature, the principal. We can no better infer that underlying objective from the choice behavior of the organism than we can identify the distorted choices made by an incentivized agent with the principal's first-best solution.

Indeed, this principal-agent metaphor is the basis of an increasingly popular methodology for behavioral economics. For example, Robson (2001) studies the biological rationale for hedonic utility. His model shows that utility can be understood as an optimal compensation scheme for an agent who has private information about the fitness consequences of various consumption bundles. Implicit is the interpretation that natural selection can be equivalently regarded as a fitness-maximizing principal with a freedom to design the agent's preferences limited only by asymmetric information. The bottom-line of such a model is an agent whose revealed preference exactly coincides with nature's first-best.

By contrast, interesting non-standard preferences can be generated by a similar model in which metaphorical nature is assumed to face additional constraints. For example, Samuelson and Swinkels (2006) consider a design problem in which the agent necessarily makes errors in information processing and nature's incentive scheme must trade off the value of incorporating the agent's private information about the local environment against the risks of granting too much leeway to imperfectly formed beliefs. Constrained to use the blunt instrument of utility to provide incentives to the agent, nature's optimal design necessarily induces behavioral biases, including self-control problems, menu-dependence, and present-bias. In Rayo and Becker (2007) nature is constrained to use outcome-contingent rewards to motivate the agent and there are limits to the granularity of this "happiness" instrument. The resulting optimal incentive scheme is equivalent to reference-dependent preference. In dynamic decision problems, certain behavioral responses can substitute for expanded memory and in Baliga and Ely (2009), nature's design economizes on memory capacity by utilizing this trade-off. The result is observationally equivalent to a sunk-cost bias.

In each of these models, the conclusions are driven by the particular constraint imposed on nature's representative, the principal. It follows that the same evolutionary argument can turn each one of these conclusions on its head. Nature is appropriately modeled as an optimizing principal only if natural selection can be assumed to operate long enough to reach an optimum. But then there should also have been ample time for nature to relax these constraints. In the language of incentives, because there is no intrinsic conflict of interest between principal and agent, in the long run nature simply "sells the firm" to the organism. Whatever residual effect of the constraints persists should have negligible costs.[8] Equivalently, observable behavior should be arbitrarily close to the first-best, and hence free from (costly) behavioral anomalies.

The present paper provides a defense of this methodology against such arguments. On the one hand, the model yields arbitrarily large innovations in the design of the organism. It follows that each "component" of the organism is optimally designed taking as given the existence of, and interactions with, other components. That is, the organism is optimal subject to certain design constraints. And on the other hand, these constraints need not be eliminated despite the fact that arbitrarily large innovations occur infinitely often. Indeed, their presence can have non-vanishing shadow costs even in the long run.

2 The Model

An organism is designed to solve a fixed decision problem, instances of which are presented to the organism repeatedly over time. The decision problem has the following interpretation. A resource is available at a certain location. The location is realized independently in each period. Signals which reveal the location of the resource are available to the organism. The problem for the organism is to input these signals, interpret them, and then choose a location in attempt to extract the resource. At any given point in time, the organism is subject to a complexity constraint which imposes a technological trade-off between these activities. The optimal design of the organism at a given level of complexity will balance this trade-off. Over time, the organism will evolve by increasing complexity and the design of

---

[8] Of course there are physical constraints which could not be eliminated no matter how long nature is left to act. But the appropriate comparison here is between the outcome of the evolutionary process as modeled by a designer subject to physical and non-physical constraints and the design that is optimal subject only to the physical constraints. The argument here is that all residual internal design inefficiencies should, in the long run, have negligible bottom-line consequences.

the organism will adjust to the relaxed constraint. We describe the long run behavior of this evolutionary process.

The Decision Problem A resource is hidden at a location $\theta \in [-1, 1]$ which is drawn from the uniform distribution. The organism will choose a location a and search intensity i. The payoffs are

$$u(a, i, \theta) = i\left[2a\theta - a^2\right]. \tag{1}$$

These payoffs model the following trade-offs. First, for any given search intensity, the organism would like to choose a location a as close as possible to θ. Likewise, for any given a, the payoffs are increasing in search intensity. The complexity constraint (to be described in the next section) will imply a technological trade-off between these to be balanced against the trade-off in payoffs captured by Equation 1. Beyond capturing these trade-offs, the precise functional form is for computational convenience only. We refer to the expression $2a\theta - a^2$ as the payoff per-unit of intensity, or *average payoff*.

In the environment there is a collection of signals σ available to the organism which convey information about the current location of the resource. The structure of correlation betwen σ and θ is given by a fixed joint distribution to be specified later. New realizations of both are drawn from this distribution in each of an infinite sequence of periods, independently over time.

Formally, $\sigma = \sigma_1, \sigma_2, \ldots \in \{-1, +1\}^\infty$ is an infinite sequence of signals from $\{-1, +1\}$ and the organism is able to collect a finite sample of size k. The parameter k is referred to as the *precision* of the organism.

The organism has a *strategy* which selects a location to search as a function of the sample $\sigma_1, \ldots, \sigma_k$. At any point in time, this strategy is characterized by another finite sequence $\pi = \pi_0, \pi_1, \ldots, \pi_k \in \{-1, +1\}^{k+1}$ in accordance with the following rule. When the organism observes the sample $\sigma_1, \ldots, \sigma_k$ it applies this formula

$$a(\sigma_1, \ldots, \sigma_k | \pi) = \frac{\pi_0}{k+2} \sum_{j=1}^{k} \pi_j \sigma_j. \tag{2}$$

to select a location to search. To interpret this, note that the sign of $\pi_0 \pi_j$ expresses whether the organism acts as though signal σ_j is positively or negatively correlated with the location θ. The organism then obtains conditionally expected payoff

$$i\left[2a(\sigma_1, \ldots, \sigma_k | \pi) \cdot \mathbf{E}(\theta | \sigma_1, \ldots, \sigma_k) - a(\sigma_1, \ldots, \sigma_k | \pi)^2\right] \tag{3}$$

The sequence $\pi_0, \ldots, \pi_k$ can be viewed as the part of the genetic code of the organism that represents the organism's current "model" of the true correlation structure in the environment. Throughout the paper I will refer to π as the genetic code, noting that this is a slight misuse of terminology because both i and π determine the behavior of the organism and both will be subject to mutation.

We have assumed that the organism uses an algorithm from a particular class, namely those parametrized by Equation 2. When the parameters of the environment are defined next, we will show that the optimal algorithm is always a member this class.

Computation and Complexity The organism is subject to a complexity constraint. We model this by assuming that the organism has a fixed capacity x of "energy" or computational capacity with which to implement its strategy. The complexity cost includes the computational cost of processing signals plus the costs of search intensity. We model computational costs by assuming that each arithmetic operation (adding, multiplying by -1) in the formula Equation 2 has unit cost.[9] Thus, the strategy $a(\sigma_1, \ldots, \sigma_k | \pi)$ requires

$$\kappa(\pi) = k + \sum_{j=0}^{k} \mathbb{1}_{\{\pi_j = -1\}}$$

computational steps to implement,[10] and an organism of complexity x would then have a search intensity equal to the residual

$$i = x - \kappa(\pi). \tag{4}$$

The optimal design maximizes Equation 3 subject to this complexity "budget" constraint.

The Environment Recall that π describes the organism's model of the correlation between σ and θ. Now I describe the true correlation structure. The *environment* is described by a parameter λ which specifies for each j whether σ_j is positively or negatively correlated with θ. Formally, λ is an

---

[9] As will become clear below these computational costs are best thought of as representing the costs of transforming the data into a common format for aggregation and not literally the costs of doing arithmetic.

[10] or convenience we exclude the cost of dividing by $k + 2$ as this cost will be common to all algorithms in this class.

infinite sequence $\lambda = \lambda_1, \lambda_2, \ldots \in \{-1, +1\}^\infty$ that is determined at the beginning of time according to an i.i.d. process with $\text{Prob}(\lambda_j = 1) = l > 1/2$. Recall that in each period, θ is drawn from the uniform distribution. Next, conditional on the realized location θ, each signal σ_j is chosen independently with

$$\text{Prob}(\sigma_j = \lambda_j) = \hat{\theta}. \tag{5}$$

where $\hat{\theta} = (\theta + 1)/2$ is just $\theta \in [-1, +1]$ normalized to a probability. To understand this structure, first consider a signal j for which $\lambda_j = 1$. In this case, observing $\sigma_j = 1$ indicates that the resource is likely to be located further to the right, whereas $\sigma_j = -1$ indicates that the resource is likely to be located further to the left. However, when $\lambda_j = -1$, these inferences are reversed. For this reason, if $\lambda_j = -1$, then we say that the jth signal is *inverted*.

3 The Optimal Organism

The first step in the analysis is to solve the "engineering benchmark" problem of maximizing payoffs for any fixed level of complexity under the assumption that λ is known. This will be our point of comparison when we later analyze the evolutionary outcome when the organism must adapt to λ over time by natural selection.

Let us first ignore complexity costs and begin by characterizing the first-best algorithm for a given precision k. Inspection of the objective function in Equation 1 reveals that for any given sample of size k, the first-best strategy computes the conditional expectation $\mathbf{E}(\theta|\sigma_1, \ldots, \sigma_k)$ and selects location

$$a_k^*(\sigma_1, \ldots, \sigma_k) := \mathbf{E}(\theta|\sigma_1, \ldots, \sigma_k).$$

An explicit formula for the conditional expectation follows from standard properties of binomial sampling.[11]

$$\bar{\theta}_k := \mathbf{E}(\theta|\sigma_1, \ldots, \sigma_k) = \frac{1}{k+2} \sum_{j=1}^{k} \lambda_j \sigma_j. \tag{6}$$

---

[11] Write $\hat{\theta} = (\theta + 1)/2$. Then $\hat{\theta}$ is distributed uniformly on $[0,1]$ and observation of $\{\lambda_j \sigma_j\}_{j=1}^k$ is a binomial sampling process from $\{-1, 1\}$ with unknown probability $\hat{\theta}$ that $\lambda_j \sigma_j = 1$. It is a standard result that in this case the posterior distribution of $\hat{\theta}$ is a Beta distribution with parameters (ζ_1, ζ_2) where ζ_1 is one plus the number of j for which $\lambda_j \sigma_j = 1$ and ζ_2 is one plus the number of j for which $\lambda_j \sigma_j = -1$. The expectation of the Beta distribution is $\frac{\zeta_1}{\zeta_1 + \zeta_2}$. This yields Equation 6 after some algebra.

Therefore the optimal average payoff is given by the random variable

$$2a_k^*\bar{\theta}_k - (a_k^*)^2 = \bar{\theta}_k^2$$

whose expectation with respect to σ is

$$\mathbf{E}_\sigma\left(\bar{\theta}_k^2\right) = \mathrm{var}(\bar{\theta}_k) \tag{7}$$

since the unconditional expectation of θ is equal to zero and therefore by the law of total probability so is $\mathbf{E}_\sigma\left(\bar{\theta}_k\right)$.

Thus, the variance of the estimator $\bar{\theta}_k$ will play an important role in the analysis. For future reference, in the appendix it is shown that this variance increases monotonically in the precision[12]

$$\mathrm{var}\left(\bar{\theta}_{k-1}\right) < \mathrm{var}\left(\bar{\theta}_k\right) < \mathrm{var}\left(\theta\right) = 1/3 \tag{8}$$

for all k and moreover that the incremental variance is asymptotically on the order of $1/k^2$:

$$\mathrm{var}(\bar{\theta}_k) - \mathrm{var}(\bar{\theta}_{k-1}) = \frac{1-\mathrm{var}(\bar{\theta}_{k-1})}{(k+2)^2} \tag{9}$$

Note for future reference that the covariance between $\lambda_j\sigma_j$ and $\lambda_{j'}\sigma_{j'}$ is positive.

$$\mathrm{cov}(\lambda_j\sigma_j, \lambda_{j'}\sigma_{j'}) > 0 \tag{10}$$

Minimizing Computational Costs We have shown that the first-best design for a given k achieves average payoff $\mathrm{var}(\bar{\theta}_k)$. Using this result, after incorporating complexity costs, we obtain the following formulation of the optimal design for a fixed level of complexity x.

$$\begin{aligned} \max_{\pi} \quad & i \cdot \mathrm{var}\,\bar{\theta}_k \\ \text{subject to} \quad & a(\cdot|\pi) = \bar{\theta}_k \\ & x = i + \kappa(\pi) \end{aligned}$$

A solution to this problem necessarily economizes the complexity constraint by choosing the genetic code π to minimize the computational costs of implementing the estimator $\bar{\theta}_k$. This cost-minimization problem is the central idea of the model.

---

[12] Intuitively, higher precision means a better estimate implying that the estimate is more sensitive to the variation in the state.

Compare the class of algorithms in Equation 2 and the optimal estimator in Equation 6. It is easy to see that there are two types of organism that achieve the optimum. A *positively-aligned* organism is one with $\pi_0 = +1$ and $\pi_j = \lambda_j$ for $j = 1, \ldots, k$. A *negatively-aligned* organism is one with $\pi_0 = -1$ and $\pi_j = -\lambda_j$ for $j = 1, \ldots, k$. In both cases, the resulting strategy $a(\sigma_1, \ldots, \sigma_k | \pi)$ is identically equal to $\bar{\theta}_k$.

On the other hand, these two design protocols typically differ in their computational costs. The budget constraint for a positively aligned organism reduces to

$$x = i + \sum_{j=1}^{k} \left(\frac{3 - \lambda_j}{2} \right)$$

whereas a negatively aligned organism yields the following budget constraint

$$x = i + \sum_{j=1}^{k} \left(\frac{3 + \lambda_j}{2} \right). \tag{11}$$

Thus, the gene π_0 adds flexibility to the design of the organism potentially allowing it to economize on computation.[13]

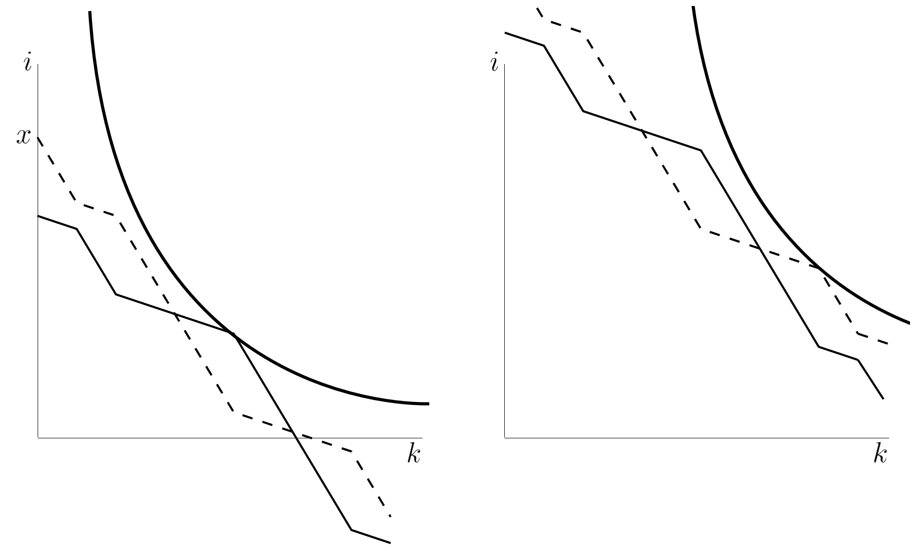
The diagrams in Figure 2 illustrate the optimal design for a fixed complexity x. The budget lines capture the tradeoff between intensity and precision for positively- (dashed) and negatively- (solid) aligned organisms respectively. Adding the jth unit of precision requires one or two additional computational steps, depending on the alignment and the value of λ_j. The "indifference curve" is the set of pairs (i, k) which achieve the same expected total payoff $i \operatorname{var}(\bar{\theta}_k)$.

Figure 2(a) shows a case in which the optimal organism is negatively aligned. As the organism increases in complexity, the budget lines shift up, potentially switching the alignment of the optimal organism. This is illustrated in Figure 2(b). Indeed, the optimal alignment depends on the sign of the moving average

$$L(k) := \frac{1}{k} \sum_{j=1}^{k} \lambda_j.$$

---

[13] Up to now we have assumed that the optimal design must implement $\bar{\theta}_k$ and searched for the algorithm that minimizes computational costs. We can see from the budget equations that the optimum indeed must implement $\bar{\theta}_k$. If instead $\pi_0 \pi_j \neq \lambda_j$ then the jth signal is biasing the estimate and also consuming computational steps. A better design would reduce precision by one and increase intensity.



(a) Low x. Negative alignment (solid line) is optimal.

(b) Higher x. Budget lines shift upward and now positive alignment is optimal.

Figure 2: Optimal organism for a fixed level of complexity x.

If it is positive, then the fraction of inverted signals up to k is less than $1/2$, and the optimal organism will be positively aligned. The negatively aligned organism is optimal in the alternative case.

Recall that we have assumed that $l > 1/2$. This implies that for sufficiently complex organisms, positive alignment is optimal.[14] A convenient way to visualize this is to consider k sufficiently large so that $L(k) \approx 2l - 1$ and the two budget lines are approximately

$$x \approx i + k\,(2 - l)$$

for positive alignment and

$$x \approx i + k\,(1 + l)$$

for negative alignment. This is illustrated in Figure 3.

---

[14] What is important is that a particular structure is eventually optimal forever, not that it is necessarily positive alignment. The model could easily be extended to allow the eventually preferred alignment to be randomly determined. This would capture the point that the flexibility of the organism is beneficial *ex ante* because the preferred alignment cannot be predicted in advance.

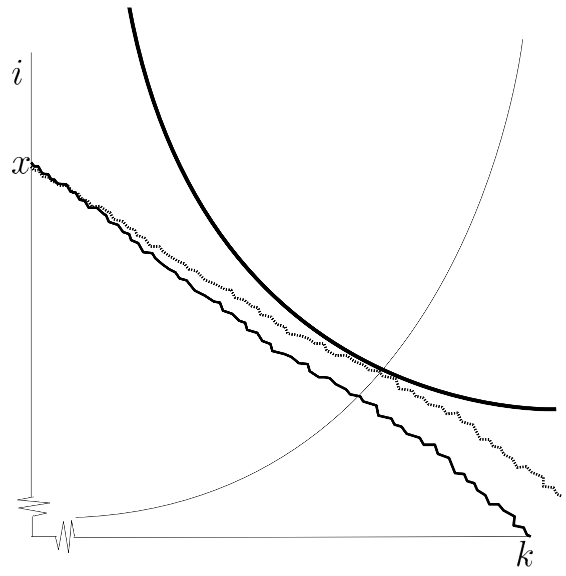


Figure 3: Optimal alignment for large k.

4 Kludge

To this point we have been analyzing the design of the organism from the perspective of an engineer who knows λ. The focus now turns to a model of the adaptive process whereby the genetic code π is instead shaped over time by the force of natural selection.

First, note that for sufficiently complex organisms, positive alignment yields a greater budget. Once this is the case, any negatively aligned organism is attempting to implement the optimal decision rule via an inefficient protocol. For this reason and reasons developed further below, we refer to such an organism as a *kludge.*

Definition 1. *Suppose that the fraction of inverted signals up to k is less than* 1/2*, i.e.*

$$\frac{1}{k}\sum_{j=1}^{k}\lambda_j > 0.$$

Then we say that an optimal negatively aligned organism with precision k is a kludge.

We can quantify the inefficiency of a kludge of complexity x. A switch to positive alignment would produce an organism of the same precision but strictly higher intensity. Indeed the intensity and therefore the payoff can be increased by a number which (on average) increases linearly in k.

However, this measure may be hard to interpret as it depends on a cardinal interpretation of payoffs. As an alternative, let us define the following ordinal concept of inefficiency of an organism. Say that the organism is *asymptotically structurally inefficient* if there is a specific gene (here π_0) such that at some point in time, and forever thereafter, there exists an improvement which involves altering this gene, and yet no such change ever occurs.[15]

5 The Adaptive Process

The final ingredient in the model is a description of the process by which the organism evolves. I adopt a simple model of mutation and natural selection designed to capture the effects of a general class of adaptive processes. The specific assumptions are chosen mostly for expositional and analytical convenience.

The organism $\mathcal{O}$ is parametrized by the pair $\mathcal{O} = (x, \pi)$. We consider adaptation of a single organism over time. Let $\mathcal{O}_t$ denote the organism existing at time t. Each period t, the organism $\mathcal{O}_t = (x, \pi)$ is evaluated according to its overall payoff given by

$$V(\mathcal{O}_t) = \mathbf{E}_{\theta,\sigma}\left(x - \kappa(\pi)\right)\left[2\theta a(\sigma_1, \ldots, \sigma_k|\pi) - a(\sigma_1, \ldots, \sigma_k|\pi)^2\right].$$

In each period, with positive probability, a variation occurs which results in a new version $\mathcal{O}'$ of the organism. We apply the following simplistic model of natural selection. If the variant $\mathcal{O}'$ is more fit, i.e. $V(\mathcal{O}') > V(\mathcal{O}_t)$, then the variant replaces the existing version and survives to date

---

[15] A virtue of this definition is that it excludes "marginal inefficiencies" where at any point in time some inefficiencies are present, but every inefficiency, once it appears, is eventually eliminated. For example, we may imagine that the most recently developed features of the organism might begin in an inefficient state, but eventually as the organism matures, these features are improved to their optimal state and align optimally with the rest of the organism. By contrast, asymptotic structural inefficiency identifies persistent mis-alignments. It would be desirable to sharpen the definition even further by considering dynamic efficiency issues. Without going into the details of such a definition, I note that the kludges in this paper represent static as well as dynamic inefficiencies. In addition to outperforming a kludge at each point in time, positively aligned organisms also grow in intensity and precision faster than kludges.

$t+1$, that is $\mathcal{O}_{t+1} = \mathcal{O}'$. If not, then the existing version survives, i.e. $\mathcal{O}_{t+1} = \mathcal{O}_t$.[16]

Two types of variations can occur which we call *expansion* and *mutation*. First, with probability q, the organism undergoes expansion. This involves either increasing complexity x or increasing the length of the genetic code π, by one unit. Assume that the organism always expands on the dimension which results in a higher payoff and that the "gene" π_{k+1} at the added locus $k+1$ is set optimally given the existing π.[17]

To illustrate and to preview formal arguments to come, suppose at time t, the organism is a kludge with complexity x and precision k. Let $V_{x,k}$ denote the overall payoff of such an organism. If expansion occurs and $V_{x+1,k} > V_{x,k+1}$ then the organism will increase complexity by one unit, otherwise it will increase precision (by increasing π.)

To trace out the evolution of a kludge, it is useful to define the following sequence of numbers $\{x_k\}_{k=1}^{\infty}$. First, set $x_1 = 1$. Next, for $k > 1$, define x_k to be the smallest $x \geq x_{k-1}$ such that $V_{x,k} > V_{x+1,k-1}$. As the organism expands, it will improve along the relevant margin as illustrated in Figure 4 (beginning from some initial level of precision $\bar{k}$.)

On the other hand[18], with probability $(1-q)$ the organism instead undergoes a mutation which alters the genetic code π. Mutations are modeled as follows. A subset of the "loci" $\{0, 1, \ldots, k\}$ is selected and the genes π_j at those loci change sign. We will assume that all subsets have positive probability.[19] Apart from this, we can be quite general in terms of the probabilities with which various subsets are selected. One simple model would be the following case of independent mutations: there is a fixed mutation probability $\mu > 0$ and each gene π_j changes sign with independent probability μ.

---

[16] It is obviously a simplification to use expected fitness as the criterion, but this is an approximation to an environment in mutations arrive infrequently. In particular, we can think of each period as representing a length of time sufficiently long that the actual performance approximates the average.

[17] These assumptions keep the analysis simple. A general model would set the sign of π_{k+1} randomly. However, whenever the sign is mis-aligned, (i.e. $\pi_{k+1}\pi_0 \neq \lambda_{k+1}$) this would reduce the payoff of the organism, so there is no loss in ignoring these events. A process in which the organism can expand on either dimension with positive probability delivers the same conclusion as this one but with added notational burden.

[18] It is an innocuous notational simplification that some variation occurs in every period with probability 1.

[19] This is used only for the benchmark result in Proposition 1.

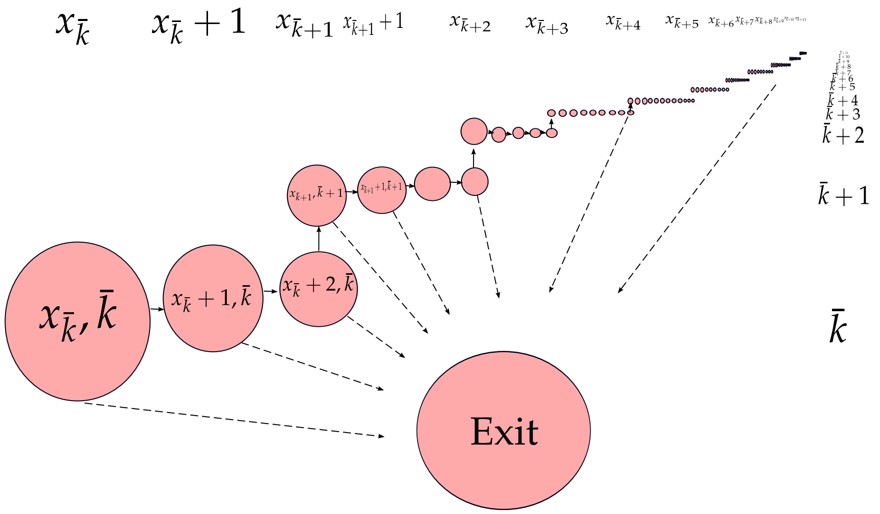


Figure 4: Escape Process.

Returning to Figure 4, the crucial mutation events will be those that result in a new organism that outperforms a kludge, illustrated by the diagonal arrows. Recall that a kludge is negatively aligned despite positive alignment being optimal. At a minimum, therefore, an improvement would require changing the alignment, i.e. the sign of π_0. Note however, that the kludge has all other genes π_j negatively aligned. Thus a changing only of π_0 would result in a complete misalignment of π. This means that in addition to a change in sign of π_0, a mutation must also change the sign of sufficiently many π_j so that the organism makes an accurate estimate of the location of the resource and benefits from the reduced computational burden of positive alignment. The size of such a mutation, measured by the number of genes which would be required to change sign, will determine the probability that the organism escapes being kludged.

The key to our result is in showing that the size of this mutation is bounded below by a constant fraction of k. Thus, the probability of escape will be shrinking as the organism expands in complexity. The independent mutation model would imply that this probability shrinks exponentially fast, and this will enable us to show that, with positive probability, an escape never occurs.

While the independent mutation model is useful for building intuition, it imposes more structure than is needed for the results. Our assumption abstracts the key condition which is the following generalized large-deviation property governing the asymptotic probabilities of large mutations.

Assumption 1 (Large Deviation Property)**.** *Let ω_k be a probability distribution over subsets of $\{0, 1, \ldots, k\}$. We say that the family of distributions $\{\omega_k\}_{k \in \mathbb{N}}$ satisfies a* large-deviation condition *if there exists $\mu \in (0, 1)$ and a function $M : \mathbb{N} \to (0, 1)$ such that $M(k)$ is an upper bound for the probability under ω_k of selecting a subset with cardinality larger than μk and*

$$\limsup_{k} \frac{M(k+1)}{M(k)} < 1.$$

This is a large-deviation property which asymptotically bounds the probability of mutations to a large fraction m (greater than μ) of genes from any given large subset. This probability must shrink to zero at a sufficiently fast rate. Note that by a standard result from large-deviation theory, the independent mutation model is a special case. We now turn to a formal analysis of the adaptive process.

6 Analysis

First, we consider an instructive benchmark case in which $q = 0$. In this case, the complexity of the organism is fixed and cannot increase. Then, because the mutation probabilities are strictly positive, with probability 1 the organism will be optimally adapted after some finite timespan.

Benchmark with $q = 0$ Consider an arbitrary organism $\mathcal{O}$ of complexity x. Let $\mathcal{O}^*$ be an optimal organism of the same complexity. There is a lower bound on the probability of a mutation large enough to transform any such $\mathcal{O}$ into $\mathcal{O}^*$. In the worst case, a change to the entire genetic structure will be required and the probability of such a large mutation is strictly positive by assumption. When $q = 0$, the organism will never increase in complexity and so this remains forever a lower bound on the probability of reaching an optimally adapted organism in a single step. It follows that with probability 1 the optimal organism will appear eventually. Moreover an optimal organism can never be replaced if the complexity of the organism cannot increase.

Proposition 1. *When $q = 0$, with probability 1 the organism is eventually optimally adapted, regardless of the initial complexity.*

The proposition shows that any asymptotic inefficiency that arises when $q > 0$ is not due to a simple problem of local optima. The model allows for arbitrarily large mutations at any point in time. Thus, any improvement, of any fixed size, if available for sufficiently long, will eventually be realized. On the other hand, this argument does not apply to improvements which require larger and larger mutations. Potentially, the organism can gradually improve at the margin by increasing in complexity, all the while intensifying the complementarity among its components. This would mean that substantial improvements decline in probability. Whether such improvements will be realized will depend on the rate at which this probability declines.

The main result of the paper concerns the case of $q > 0$ and asymptotic structural inefficiency.

Theorem 1. *Suppose $\mu < 1/2$. When $q > 0$ there is a positive probability that the organism will be forever kludged and thus asymptotically structurally inefficient.*

6.1 Proof of Theorem 1

Recall that we have assumed that $l > 1/2$. The parameter l determines the probability that each $\lambda_j = +1$. As discussed above, what matters for the optimal design of the organism is the sign of the moving average

$$L(k) = \frac{1}{k} \sum_{j=1}^{k} \lambda_j.$$

Because $l > 1/2$, by the strong law of large numbers, there is probability one on the set of environments in which there exists some $\bar{k}$ such that $L(k)$ is positive for all $k > \bar{k}$. Throughout the proof, we fix such an environment λ and integer $\bar{k}$ and consider the stochastic process of evolution in that environment.

Let $\mathcal{O}$ be a kludge. We wish to bound the probability that the organism will ever be replaced by one that is not a kludge. Since expansion can only lead to a kludge of larger complexity or precision, the crucial events will involve mutations. We say that a *beneficial mutation* occurs if a mutation produces an organism $\mathcal{O}'$ which has strictly larger payoff. Let $\eta_{x,k}$ be the probability that a kludge of complexity x and precision k will undergo a beneficial mutation. Consider the following simplified stochastic process. Let the states of the process correspond to the levels of complexity and precision (x, k) of the organism. Refer to Figure 4. At each state (x, k), three transitions are possible. First, with probability q, an expansion occurs and the organism transits to the next state which is $(x+1, k)$ in the case of $x < x_{k+1}$ (the horizontal transitions in the figure) or $(x, k+1)$ in the case of $x = x_{k+1}$ (the vertical transitions.) Second, with probability $(1-q)\eta_{x,k}$, the organism "escapes" and the process terminates. Finally, with the remaining probability, the process continues and the state is unchanged at (x, k).

The process begins in state $(x_{\bar{k}}, \bar{k})$. We refer to this process as the *escape process.* It is a simplification of the true evolutionary process because it does not consider the further evolution that occurs after a beneficial mutation. We will analyze the probability that escape *never* occurs and this will be a lower bound for the probability that the organism remains kludged forever.

Lemma 1. *The probability that the organism remains asymptotically structurally inefficient is bounded below by the probability that the escape process never terminates. This probability is positive if and only if*

$$\sum_{k=1}^{\infty} \sum_{x=x_k}^{x_{k+1}} \eta_{x,k} < \infty.$$

Proof. We apply a standard technique from the theory of countable-state Markov chains[20]. With each non-terminal state (x, k), we associate a variable $Z_{x,k}$. We consider the system of equations which expresses each $Z_{x,k}$ as a linear combination of other variables where the weights are the non-terminal transition probabilities. This yields the following system.

$$Z_{x,k} = qZ_{x+1,k} + (1-q)(1-\eta_{x,k})Z_{x,k} \qquad x \in \{x_k, \ldots, x_{k+1}-1\} \tag{12}$$

$$Z_{x,k} = qZ_{x,k+1} + (1-q)(1-\eta_{x,k})Z_{x,k} \qquad x = x_{k+1} \tag{13}$$

The probability is positive that the system will never reach the terminal state if and only if the system possesses a bounded, non-zero solution. To establish this, we set $Z_{x_{\bar{k}},\bar{k}} = 1$, rewrite the system as follows

$$Z_{x+1,k} = \left(1 + \frac{(1-q)\eta_{x,k}}{q}\right) Z_{x,k} \qquad x \in \{x_k, \ldots, x_{k+1}-1\} \tag{14}$$

$$Z_{x,k+1} = \left(1 + \frac{(1-q)\eta_{x,k}}{q}\right) Z_{x,k} \qquad x = x_{k+1}, \tag{15}$$

and recursively substitute to obtain

$$Z_{x,k} = \prod_{l=\bar{k}}^{k-1} \prod_{y=x_l}^{x_{l+1}} \left[1 + \frac{(1-q)}{q}\eta_{y,l}\right] \cdot \prod_{y=x_k}^{x-1} \left[1 + \frac{(1-q)}{q}\eta_{y,k}\right]. \tag{16}$$

---

[20] See (Billingsley, 1995, Theorems 8.4 and 8.5)

The solution is bounded iff

$$\prod_{k=1}^{\infty} \prod_{x=x_k}^{x_{k+1}} \left[1 + \frac{(1-q)}{q} \eta_{x,k} \right] < \infty \tag{17}$$

which is equivalent to[21]

$$\sum_{k=1}^{\infty} \sum_{x=x_k}^{x_{k+1}} \left[\frac{(1-q)}{q} \eta_{x,k} \right] < \infty. \tag{18}$$

The latter is equivalent to the condition in the statement of the lemma. □

We now present two lemmas which establish a bound on the series identified in Lemma 1.

Lemma 2. *There exists a function* $M(k)$ *such that* $\eta_{x,k} \leq M(k)$ *for all* k *and* x, *and*

$$\limsup_k \frac{M(k+1)}{M(k)} < 1. \tag{19}$$

Lemma 3. *There exists a function* $C(k)$ *such that* $x_{k+1} \leq C(k)$ *for all* k *and*

$$\lim_k \frac{C(k+1)}{C(k)} = 1. \tag{20}$$

Combining lemmas 2 and 3 enables us to conclude the proof of Theorem 1 as they establish the bound

$$\sum_{k=1}^{\infty} \sum_{x=x_k}^{x_{k+1}} \eta_{x,k} \leq \sum_{k=\bar{k}}^{\infty} C(k) M(k) < \infty$$

which by Lemma 1 is enough to prove the theorem. □

We now turn to the proofs of Lemma 2 and Lemma 3.

Proof of Lemma 2. We will show that for any x, the probability $\eta_{x,k}$ is smaller than the probability that at least $k/2$ genes are selected for mutation.

Consider a mutation which resulting in the genetic code π with $\pi_0 = +1$. Let A be the subset of $\{\pi_1, \ldots, \pi_k\}$ which also undergo mutation, and assume that A consists of fewer than $k/2$ elements. We have

$$\pi_j = \begin{cases} \lambda_j & \text{for } \pi_j \in A \\ -\lambda_j & \text{for } \pi_j \in \{\pi_1, \ldots, \pi_k\} \setminus A \end{cases}$$

---

[21] Note that for any sequence of positive numbers R_n, $1 + \sum_1^x R_n \leq \prod_1^x (1 + R_n) \leq \exp(\sum_1^x R_n)$.

because the organism was negatively aligned prior to the mutation. Let B be a subset of $\pi_j \in \{\pi_1, \ldots, \pi_k\} \setminus A$ that has the same cardinality as A. (This is possible because A has fewer than $k/2$ elements.)

The expected average payoff to the new organism is

$$\mathbf{E}_\sigma \left[2a(\sigma|\pi) \cdot \mathbf{E}(\theta|\sigma) - a(\sigma|\pi)^2 \right]$$

which, expanding the definitions of $a(\sigma|\pi)$ and $\mathbf{E}(\theta|\sigma)$ is less than

$$\mathbf{E}_\sigma \left[\left(\frac{2}{k+2} \right)^2 \sum_{j=1}^{k} \pi_j \sigma_j \sum_{j=1}^{k} \lambda_j \sigma_j \right]$$

which we can decompose into three terms

$$\mathbf{E}_\sigma \left\{ \left(\frac{2}{k+2} \right)^2 \sum_{j=1}^{k} \lambda_j \sigma_j \left[\sum_{j \in A} \lambda_j \sigma_j - \sum_{j \in B} \lambda_j \sigma_j - \sum_{j \notin A \cup B} \lambda_j \sigma_j \right] \right\}.$$

Note that because the sequence $\{\lambda_j \sigma_j\}$ is i.i.d. conditional on θ, by de Finetti's theorem it is an exchangeable sequence and therefore,

$$\mathbf{E}_\sigma \left\{ \sum_{j=1}^{k} \lambda_j \sigma_j \left[\sum_{j \in A} \lambda_j \sigma_j - \sum_{j \in B} \lambda_j \sigma_j \right] \right\} = 0.$$

And recalling that for any j, j', the covariance of $\lambda_j \sigma_j$ and $\lambda_{j'} \sigma_{j'}$ is positive (Equation 10), the remaining term

$$\mathbf{E}_\sigma \left[- \sum_{j=1}^{k} \lambda_j \sigma_j \cdot \sum_{j \notin A \cup B} \lambda_j \sigma_j \right]$$

is negative. Thus, the expected average payoff of the new organism is negative. The total expected payoff is therefore lower than that of the original organism. This shows that a mutation is beneficial only if it affects more than $k/2$ genes. By the large deviation property, the probability of such a mutation is bounded by $M(k)$ where

$$\limsup_k \frac{M(k+1)}{M(k)} < 1.$$

$\square$

Proof of Lemma 3. Recall that x_k is defined to be the smallest $x \geq x_{k-1}$ such that $V_{x,k} > V_{x+1,k-1}$. Because a kludge is negatively aligned and negative alignment achieves the optimal average payoff $\operatorname{var} \bar{\theta}_k$ we can apply the budget formula in Equation 11, to express x_k as follows.

$$\left[x_k - 1 - \sum_{j=1}^{k} \frac{3 + \lambda_j}{2} \right] \operatorname{var} \bar{\theta}_k = \left[x_k - \sum_{j=1}^{k-1} \frac{3 + \lambda_j}{2} \right] \operatorname{var} \bar{\theta}_{k-1}$$

up to an integer. Rearranging, we can obtain the bound

$$x_k \leq \frac{3 \operatorname{var} \bar{\theta}_k}{\operatorname{var} \bar{\theta}_k - \operatorname{var} \bar{\theta}_{k-1}} + 2k$$

From Equation 8 and Equation 9 the first term is bounded by a constant χ multiplied by $(k+2)^2$ so that

$$x_k \leq C(k) := (k+2)^2 \chi + 2k$$

with

$$\lim \frac{C(k+1)}{C(k)} = \lim \frac{(k+3)^2 \chi + 2(k+1)}{(k+2)^2 \chi + 2k} = \lim \frac{(k+3)^2}{(k+2)^2} = 1.$$

□

7 Extensions

The model is purposefully simple. However a number of natural and instructive extensions are possible. Here I will sketch two.

First, the model of natural selection does not have any serious competition across organisms. And indeed in its simple form here, competition could eliminate the inefficiency as long as we are assured that at least one individual in the population begins positively aligned. This will be true with large probability in a large population. However, it is easy to enrich the decision problem so that with probability 1 any individual will be kludged and hence so will all individuals in the population. What follows is a sketch of such an extension.

Suppose that now there are two ways in which an organism can profitably increase in complexity. First, it may improve its procedure for solving a given resource extraction problem as in the present model. Second now, the organism may evolve the ability to extract a new kind of resource in addition to the old. Each problem that the organism is currently equipped to solve has its own apparatus identical in structure to the one modeled here, however independently evolving. And the organism always benefits from increasing in complexity in this way. Then the result from this paper that an individual problem will be kludged with positive probability implies that every member of the population, even the most fit, will be kludged in the way it solves a positive, and bounded, fraction of problems.

The second extension uses competition not as a robustness test, but to draw new conclusions from the model. Suppose that organisms compete for resources within a niche. There is a hierarchy of niches ordered by the complexity necessary to exploit each niche. At the beginning, simple organisms inhabit the simplest niche and evolve to adapt to that niche. Soon, efficient simple organisms emerge and grow the saturation point. This creates an incentive for organisms to increase in complexity in order to move into the unoccupied more complex niche. This process continues. The result of this process would be that some "species" stop increasing in complexity while others are perpetually driven to increasing complexity. Those that stop, the simplest organisms, will eventually become efficiently adapted to their niche as in Proposition 1. By contrast, the leading edge organisms will be better described by Theorem 1, potentially remaining kludged forever.

A Appendix

Here we prove the statistical result given in Equation 9 and Equation 8.

Define $\tau_j = \lambda_j \sigma_j$. Note that the $\mathbf{E}\tau_j = 0$ and so by Equation 6, $\mathbf{E}\bar{\theta}_k = 0$ and hence $\mathrm{var}(\bar{\theta}_k) = \mathbf{E}_\sigma\left(\bar{\theta}_k^2\right)$. We calculate

$$\mathbf{E}\left(\tau_j | \tau_1, \ldots, \tau_{j-1}\right) = \mathrm{Prob}(\tau_j = 1 | \tau_1, \ldots, \tau_{j-1}) - \mathrm{Prob}(\tau_j = -1 | \tau_1, \ldots, \tau_{j-1})$$

and by Equation 5 and the law of total probability,

$$\begin{aligned} &= \mathbf{E}\Big(\frac{\theta + 1}{2} | \tau_1, \ldots, \tau_{j-1}\Big) - \mathbf{E}\Big(\frac{1-\theta}{2} | \tau_1, \ldots, \tau_{j-1}\Big) \\ &= \frac{\bar{\theta}_{k-1} + 1}{2} - \frac{1 - \bar{\theta}_{k-1}}{2} \\ &= \bar{\theta}_{k-1}. \end{aligned} \tag{21}$$

From Equation 6,

$$\bar{\theta}_k = \left(\frac{k+1}{k+2}\right)\bar{\theta}_{k-1} + \frac{\tau_k}{k+2},$$

so

$$\begin{aligned}\operatorname{var}(\bar{\theta}_k) &= \mathbf{E}\left[\left(\frac{k+1}{k+2}\right)^2 \bar{\theta}_{k-1}^2 + 2\left(\frac{\tau_k}{k+2}\right)\left(\frac{k+1}{k+2}\right)\bar{\theta}_{k-1} + \frac{1}{(k+2)^2}\right] \\ &= \mathbf{E}_{\tau_1,\ldots,\tau_{j-1}}\mathbf{E}\left[\left(\frac{k+1}{k+2}\right)^2 \bar{\theta}_{k-1}^2 + 2\left(\frac{\tau_k(k+1)}{(k+2)^2}\right)\bar{\theta}_{k-1} + \frac{1}{(k+2)^2}|\tau_1,\ldots,\tau_{j-1}\right]\end{aligned}$$

Applying Equation 21

$$= \mathbf{E}_{\tau_1,\ldots,\tau_{j-1}}\left[\frac{1}{(k+2)^2} + \frac{2(k+1)+(k+1)^2}{(k+2)^2}\bar{\theta}_{k-1}^2\right]$$

and we have

$$\begin{aligned}\operatorname{var}(\bar{\theta}_k) - \operatorname{var}(\bar{\theta}_{k-1}) &= \mathbf{E}\left[\frac{1}{(k+2)^2} - \left(1 - \frac{2(k+1)+(k+1)^2}{(k+2)^2}\right)\bar{\theta}_{k-1}^2\right] \\ &= \mathbf{E}\left[\frac{1}{(k+2)^2} - \frac{1}{(k+2)^2}\bar{\theta}_{k-1}^2\right] \\ &= \frac{1-\operatorname{var}(\bar{\theta}_{k-1})}{(k+2)^2}.\end{aligned}$$

establishing Equation 9. □

Now to show Equation 8, note that the sequence of random variables $\bar{\theta}_k$ converges in distribution to the random variable θ and therefore $\operatorname{var}(\bar{\theta}_k) \to \operatorname{var}(\theta)$. Equation 9 implies that the former is monotonically increasing hence $\operatorname{var}(\bar{\theta}_k) < \operatorname{var}(\theta)$ for all k.

References

BALIGA, S., AND J. C. ELY (2009): "Mnemonomics: Sunk Cost Bias as a Memory Kludge," working paper.

BILLINGSLEY, P. (1995): *Probability and Measure*. Wiley and Sons.

DAWKINS, R. (1986): *The Blind Watchmaker*. W.W. Norton.

FRIEDMAN, M. (1966): "The Methodology of Positive Economics," in *Essays in Positive Economics*. University of Chicago Press.

LINDEN, D. (2007): *The accidental mind*. Belknap Press.

MARCUS, G. (2008): *Kluge: The haphazard construction of the human mind*. Houghton Mifflin Harcourt.

RAYO, L., AND G. BECKER (2007): "Evolutionary Efficiency and Happiness," *Journal of Political Economy*, 115(2), 302–337.

ROBSON, A. (2001): "Why Would Nature Give Individuals Utility Functions?," *Journal of Political Economy*, 109, 900–914.

SAMUELSON, L., AND J. SWINKELS (2006): "Information, Evolution and Utility," *Theoretical Economics*, 1(1), 119–142.

SANDHOLM, W., AND A. PAUZNER (1998): "Evolution, Population Growth, and History Dependence," *Games and Economic Behavior*, 22, 84–120.