

RESEARCH ARTICLE

Optimizing Population Variability to Maximize Benefit

Leighton T. Izu^{1*}, Tamás Bányász^{1,2}, Ye Chen-Izu^{1,3,4}

1 Department of Pharmacology, University of California, Davis, Davis, California, United States of America, **2** Department of Physiology, University of Debrecen, Debrecen, Hungary, **3** Department of Medicine, University of California, Davis, Davis, California, United States of America, **4** Department of Biomedical Engineering, University of California, Davis, Davis, California, United States of America

* ltizu@ucdavis.edu



OPEN ACCESS

Citation: Izu LT, Bányász T, Chen-Izu Y (2015) Optimizing Population Variability to Maximize Benefit. PLoS ONE 10(12): e0143475. doi:10.1371/journal.pone.0143475

Editor: Roberto Andre Kraenkel, São Paulo State University, BRAZIL

Received: July 21, 2014

Accepted: November 5, 2015

Published: December 9, 2015

Copyright: © 2015 Izu et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data can be found DOIs: <http://dx.doi.org/10.6084/m9.figshare.1601024> and <http://dx.doi.org/10.6084/m9.figshare.1601027>.

Funding: This work was supported by grant number R01HL090880, National Institutes of Health, National Heart, Blood, and Lung Institute (<http://www.nhlbi.nih.gov/>), grant recipients: LTI and YC; grant number R03 AG031944 (YC), National Institutes of Health, National Institute on Aging (<http://www.nia.nih.gov/>); and University of California, Davis Startup Funds, University of California, Davis (<http://ucdavis.edu> (LTI)). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Variability is inherent in any population, regardless whether the population comprises humans, plants, biological cells, or manufactured parts. Is the variability beneficial, detrimental, or inconsequential? This question is of fundamental importance in manufacturing, agriculture, and bioengineering. This question has no simple categorical answer because research shows that variability in a population can have both beneficial and detrimental effects. Here we ask whether there is a certain level of variability that can maximize benefit to the population as a whole. We answer this question by using a model composed of a population of individuals who independently make binary decisions; individuals vary in making a yes or no decision, and the aggregated effect of these decisions on the population is quantified by a benefit function (e.g. accuracy of the measurement using binary rulers, aggregate income of a town of farmers). Here we show that an optimal variance exists for maximizing the population benefit function; this optimal variance quantifies what is often called the “right mix” of individuals in a population.

Introduction

Variability is inherent in all populations. In manufacturing, variability has been called “the enemy of mass production” [1] and decades of research and effort have gone into reducing variability between parts [2, 3]. On the other hand, variability between individuals is the *sine qua non* of biological evolution. Whether variability in a population is beneficial, detrimental, or inconsequential is not often clear. Agriculture provides some classic examples. The potato blight epidemic of 1845–1852 that caused massive starvation in Ireland [4] occurred in part because of overdependence on a narrow range of species that happened to be susceptible to disease. Yet the detrimental effects of narrow diversity are often outweighed by the benefits of product uniformity and high productivity. Intuitively, it is plausible that there might be an optimal level of variability or diversity among individuals that could maximize the total population benefit in both productivity and adaptability.

Competing Interests: The authors have declared that no competing interests exist.

Here, we analyze how the variance between individuals can affect some measure of benefit to the population as a whole. The model we use is a population composed of heterogeneous individuals that make binary decisions; individuals vary in making a yes or no decision, and the aggregate effect of these decisions on the population is quantified by a benefit function. Such a model can be used to describe various populations, such as a town of farmers with each individual deciding on whether or not to plant tulips that year, or a population of binary measuring devices with distributed thresholds. The benefit function is defined in a way to quantify some form of benefit to the population. The benefit function for the farmer example is the aggregate income of all farmers in town; for the measuring devices, it is the accuracy of measurement. Here we show that an optimal variance exists that maximizes the population benefit function.

Methods and Analysis

Mathematical framework

To make the quantitative link between variability in a population and population benefit we need to (a) model how individuals in a population behave and (b) quantify the effects of the individuals' behavior. We use a simple behavioral model where an individual makes a binary decision based on whether a signal, called L , exceeds a threshold. Variability is introduced by assuming that the threshold differs between individuals. The probability distribution of individual thresholds, or simply the *threshold distribution*, is $\phi(L, s)$, where s is the measure of variability.

The effect of all the individual's behavior is measured by the *population benefit function* $B(L, s)$, which depends on both the signal level (L) and the variability (s).

The threshold distribution $\phi(L, s)$ and the population benefit function $B(L, s)$ constitute our mathematical framework to study how variability affects population benefit. The problem we will solve is this: *For a given signal L and threshold distribution $\phi(L, s)$, what level of variability s maximizes the population benefit $B(L, s)$?*

Sloppy Rulers

To answer this question, we consider the problem of measuring the length of an object with a population of binary rulers. We start with this example for three reasons. First, in this case the meaning of $\phi(L, s)$ and $B(L, s)$ are easily understood. Second, this measurement problem requires us to develop a method, called the Sloppy Algorithm, that finds the optimal value of variability s that maximizes the benefit function $B(L, s)$ for any L . Third, there is a surprising equivalence between making accurate measurements with binary rulers and maximizing income in a town of farmers.

The problem is to measure an object of length L where $0 \leq L \leq 1$ using a population of N binary rulers of unit length. These binary rulers have only a single mark engraved *approximately* about the midpoint. Because the position of the mark varies from ruler to ruler they are called "sloppy rulers." While making measurements with such rulers might seem contrived and unrealistic, this process, in fact, always occurs when determining the least significant digit of any measuring device. For example, in an 8-bit analog-to-digital converter (ADC) with a 5 volt full-scale range, the quantal step size is $5V/256 = 19.5$ mV, which we can take as our unit length. L is now some voltage between 0 and 19.5 mV scaled between 0 and 1. The last bit will be set to 0 or 1 depending on whether L is less than or greater 0.5. In an ADC the scaled voltage at which the bit switches between 0 and 1 (in this case 0.5) is called the threshold voltage. By analogy, we now call the graduation mark on a sloppy ruler the threshold.

It is important that the threshold of each of the N rulers be distributed approximately about the midpoint. If the mark on all N rulers were exactly at $1/2$ (as in an ADC) and L happened to be 0.4 then all we can say is that the object has length slightly less than $1/2$. But as we will show, if the thresholds are randomly distributed in “proper” way then L can be accurately determined with any degree of resolution. The proper way will be the distribution with optimal variability.

The length of an object using the sloppy rulers is estimated as follows: Align the left edges of the object and ruler. Each ruler i casts its “vote” of $\lambda_i = 0$ (vote no) or $\lambda_i = 1$ (vote yes) depending on whether the right edge of the object is below or above the threshold at ρ_i , respectively. The *threshold distribution* ϕ is the distribution of threshold locations, ρ_i . We initially assume ϕ is the normal distribution with mean $\mu = 1/2$ and standard deviation $SD = s$. The estimate of L , $\bar{\lambda}$, equals the number of yes votes, n , multiplied by 1 (length of the ruler), divided by the total number of votes N , $\bar{\lambda} = (n \times 1)/N$. We allow the position of the graduation mark to range from $-\infty$ to ∞ (possibly running off the ruler). Because the ruler length is 1 , $\bar{\lambda}$ numerically equals the probability that L is greater than ρ , $P(L > \rho)$, which as $N \rightarrow \infty$, is

$$\lim_{N \rightarrow \infty} \bar{\lambda} = P(L > \rho) = \int_{-\infty}^L \phi(z; \mu = 1/2, s) dz = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\sqrt{2}(2L - 1)}{4s} \right) \right] \quad (1)$$

$$\equiv Q(L, s)$$

This equation gives the relationship between the estimate $\bar{\lambda}$ and the variability s .

We now define a benefit function. The benefit function reflects the worth or merit we ascribe to $\bar{\lambda}$ and so may be defined in many ways. When measuring length, a natural definition of benefit would indicate how close $\bar{\lambda}$ is to the true length L . One such population benefit function, B , is given by

$$B(L, s) = 1 - \frac{|Q(L, s) - L|}{\epsilon_{\max}} = 1 - 2|Q(L, s) - L|. \quad (2)$$

ϵ_{\max} is defined as the supremum (least upper bound) of L_1 (the 1-norm) errors

$$\epsilon_{\max} = \sup \{|Q(L, s) - L|, L \in [0, 1], s \geq 0\}. \quad (3)$$

From the definition of Q it follows that $\epsilon_{\max} = 1/2$ so $B(L, s)$ ranges from 0 to 1 . The population of rulers that has the *optimal level of variability*, s^* , estimates the length exactly. Perfect accuracy and maximal benefit are achieved when s^* solves the fixed point problem

$$Q(L, s^*) = L. \quad (4)$$

The Sloppy Algorithm

At first glance there appears to be a logical problem of needing to know L in the first place in order to get s^* . However, we escape this dilemma by proving that for any L , s^* can be found iteratively and that the estimate $\bar{\lambda}$ converges monotonically to L using the following “Sloppy Algorithm”:

1. Choose s_0 arbitrarily.
2. Compute $\bar{\lambda}_1 = Q(L, s_0)$.
3. For any $i > 1$,

$$\text{choose } s_i \text{ such that } Q(\bar{\lambda}_i, s_i) = \bar{\lambda}_i. \quad (5)$$

The solution to Eq (5) is

$$s_i(\lambda_i) = \frac{\sqrt{2}(2\lambda_i - 1)}{4\text{erf}^{-1}(2\lambda_i - 1)}. \quad (6)$$

4. The next value λ_{i+1} is computed from

$$\bar{\lambda}_{i+1} = Q(L, s_i) \quad (7)$$

Steps [3] and [4] are the recursion rules and are repeated as many times as desired.

Fig 1 illustrates how the Sloppy Algorithm works. The leftmost drawing in Fig 1A shows 10 rulers with thresholds (short vertical lines) normally distributed about the center with standard deviation of $s_0 = 0.1$. For this first step the value of s_0 can be any value >0 . For step 2 we compute λ_1 by counting the number of times the right edge of the object (gray bar), whose length is $L = 0.3$, lies to the left of all the thresholds. In this case $n = 1$ so $\lambda_1 = 0.1$. For step 3, we compute s_1 by solving $Q(\lambda_1, s_1) = \lambda_1$, using Eq (6) or reading off the graph of $s(L)$ in Fig 1B, which gives $s_1 = 0.312$. For step 4 a new set of rulers is created with thresholds drawn from a normal distribution with SD s_1 (middle set of rulers) and we get the new estimate λ_2 . In this case, the right edge of the object lies above the threshold of two rulers so $n = 2$ and $\lambda_2 = 0.2$. Repeating step 3 we get $s_2 = 0.356$ based on $\lambda_2 = 0.2$. Repeating step 4 (rightmost set of rulers) gives $n = 3$ so $\lambda_3 = 0.3$, which happens to be the correct length. Steps 3 and 4 can be repeated as many times as desired. Fig 1B shows how s_i and λ_{i+1} are linked stepwise and the convergence of the algorithm.

Fig 2A and 2B show the convergence of $\bar{\lambda}_i$ and s_i for 5 different initial values of s_0 that range from 10^{-5} to 1. In this case the true value of L is 0.723 (chosen arbitrarily) so $s^*(0.723) = 0.377$; their values are shown by the black dashed lines. Notice that regardless of the initial value of s_0 , within 2 or 3 iterations $\bar{\lambda}_i$ already converges to within a few percent of L .

It can be shown that an ensemble of measuring devices is mathematically equivalent to a single device making multiple measurements. To illustrate this, we show, in Section 4 in S1 File, that a person, embedded in the Sloppy Algorithm iteration loop, can determine the absolute gray scale value of an image. The convergence to the correct gray scale value, shown in Fig E in S1 File, is similar to Fig 2.

The Sloppy Algorithm is not limited to ϕ being normally distributed and the right hand side of Eq (4) can be replaced by any strictly monotonically increasing function $f(L)$ bounded by 0 and 1. We prove that the Sloppy Algorithm converges everywhere monotonically when $f(L) = L$ and ϕ is the normal distribution in Section 1 in S1 File. Fig 2A and 2B show λ_i and s_i converging monotonically to the correct solution.

Using the Sloppy Algorithm to achieve high-resolution measurements from low-resolution instruments

It may be surprising that a population of sloppy rulers can make accurate, high resolution measurements even though each ruler, having just a single mark, is of the lowest possible resolution. Sloppy rulers, when combined with the Sloppy Algorithm, provide a useful method to obtain accurate, high-resolution measurements. This method differs from dithering [5, 6], which is also used to increase measurement resolution. Section 5 in S1 File explains how they differ mathematically and the different arenas where they are useful.

There are two costs in using sloppy rulers. First, high resolution measurements require many rulers. To get three-digit resolution requires $N \geq 1000$ rulers. Second, accuracy requires

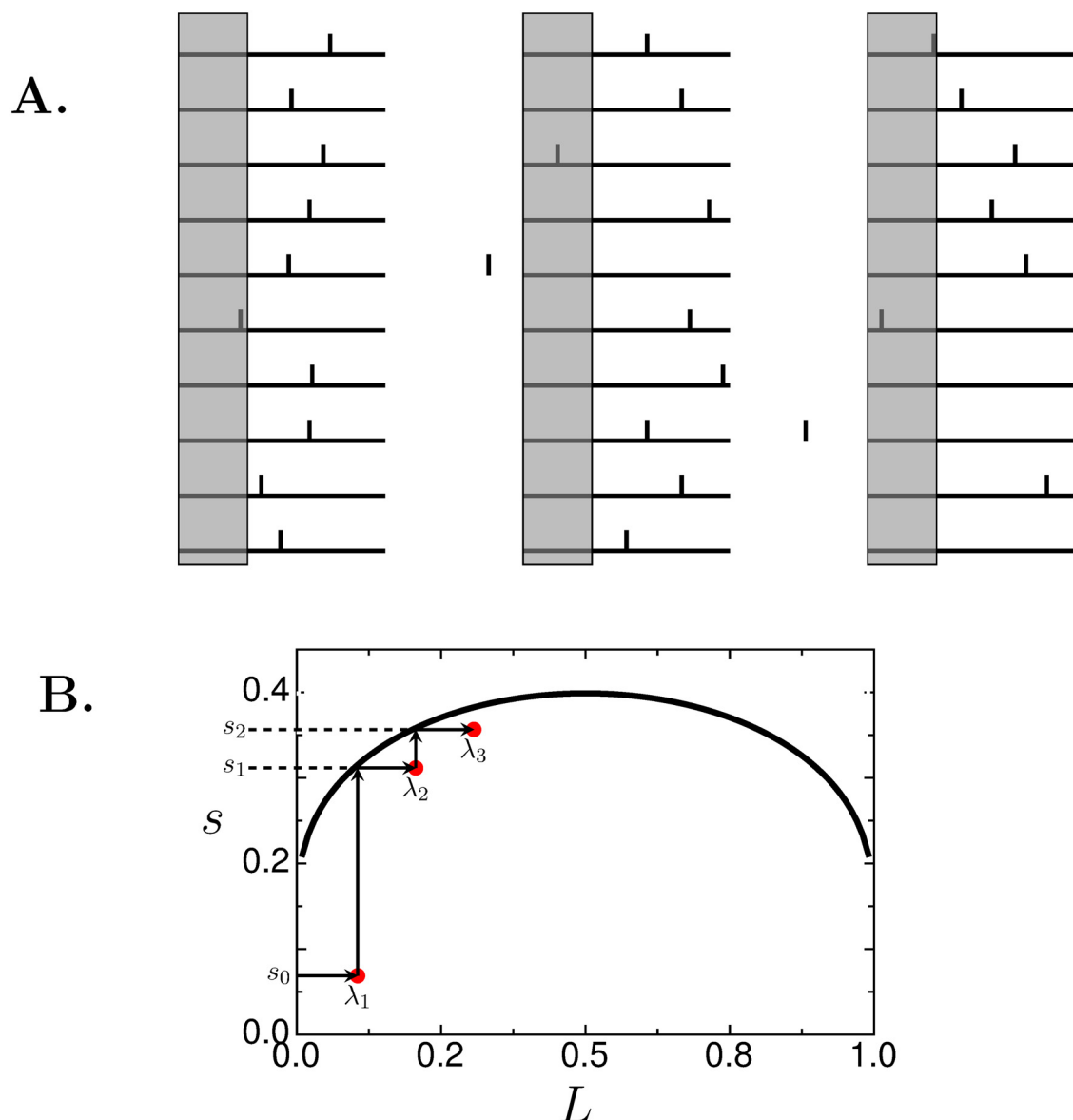


Fig 1. Schematic of recursion rules of the Sloppy Algorithm. (A), Width of gray bar is to be measured with rulers of unit length (horizontal lines). Vertical tick lines are the graduation marks. Note that a graduation mark can lie beyond the edge of the ruler as seen in the middle and rightmost set of rulers. (B), s , solves the fixed point problem $Q(L, s) = L$. Arrows show how s_i and λ_{i+1} , generated by this ruler example, are linked to each other and how the algorithm converges.

doi:10.1371/journal.pone.0143475.g001

using the correct s , which the Sloppy Algorithm finds iteratively. As Fig 2 shows, it can take from one to three iterations to get an accurate estimate of L .

Fig 3A shows what happens if the wrong value of s is used. The x -axis is the true length L and the y -axis is the estimate of L given by $Q(L, s)$. Every L requires a specific $s^*(L)$ that gives the correct estimate $Q(L, s^*) = L$ and this s^* is found using the Sloppy Algorithm. The black circles are these estimates and they fall exactly on the line $L = L$. Each solid curve shows the estimate made with a fixed value of s ; the estimates are all wrong except at $L = 1/2$ and at most two other values of L . The red curve, however, merits special attention because it shows a “short cut” to find reasonably accurate solutions without the cost of iterating.

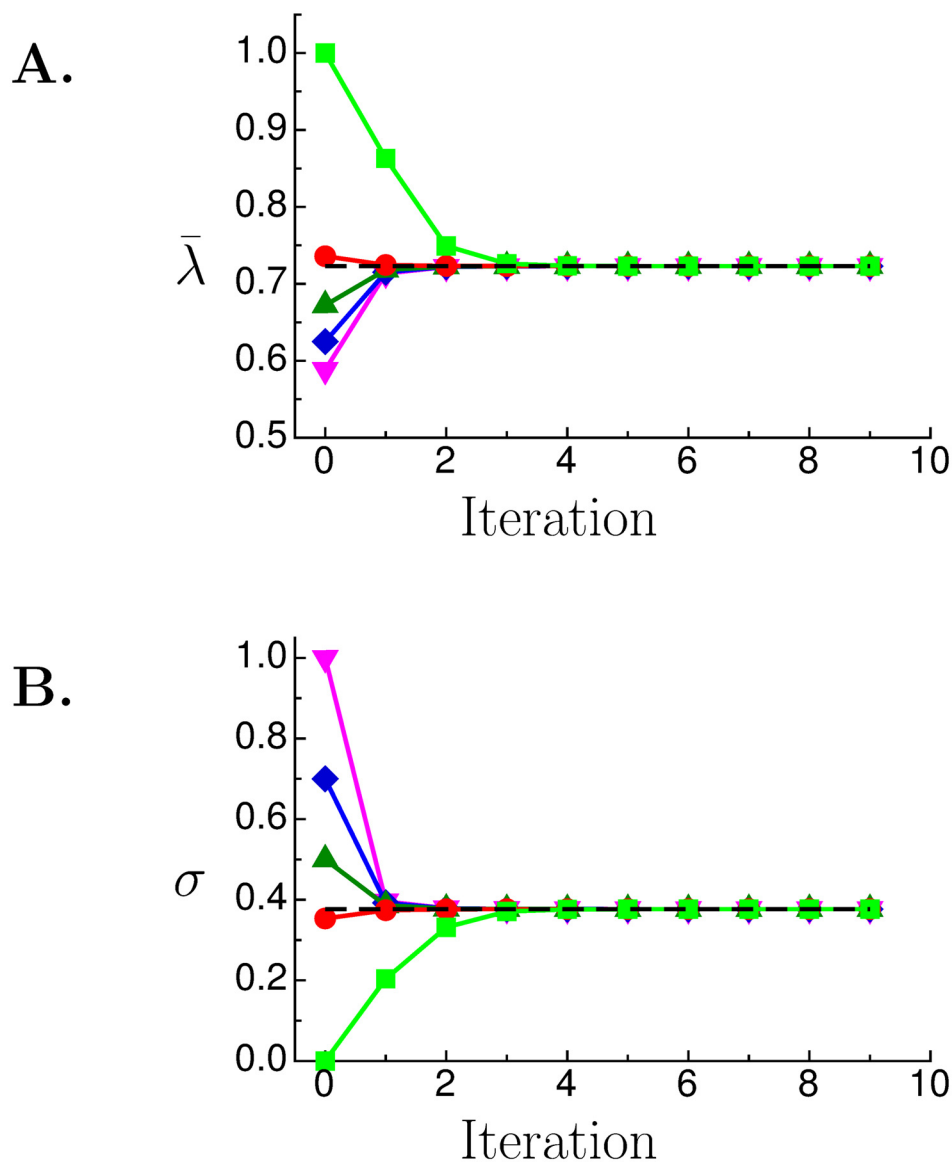


Fig 2. Convergence of the Sloppy Algorithm. Convergence of $\bar{\lambda}_i$ (A) and s_i (B) when the Sloppy Algorithm was used with different initial values of s ; $s_0 = 10^{-5}$, light green, square; magic number, red, circle; 0.5, green, up-triangle; 0.7, blue, diamond; 1.0, magenta, down-triangle. Dashed black line marks the correct values of $L = 0.723$ and $s^* = 0.377$.

doi:10.1371/journal.pone.0143475.g002

The “magic number distribution”

Although no *single* value of s produces an exact estimate for all values of L , we see that the red curve in Fig 3A is close to the unit slope line except near 0 and 1. This curve was generated using the standard deviation s_m called the “magic number.” We derive the value of the magic number, $s_m = \sqrt{2}/4 \approx 0.35$ in Section 2 in S1 File. We call the distribution $\phi(L, s_m)$ the *magic number distribution*.

The magic number distribution is important because this single distribution could replace the infinite number of distributions (indexed by $s^*(L)$) needed to get *perfect* benefit. To assess how well the magic number distribution can do this we need to gauge how close a distribution

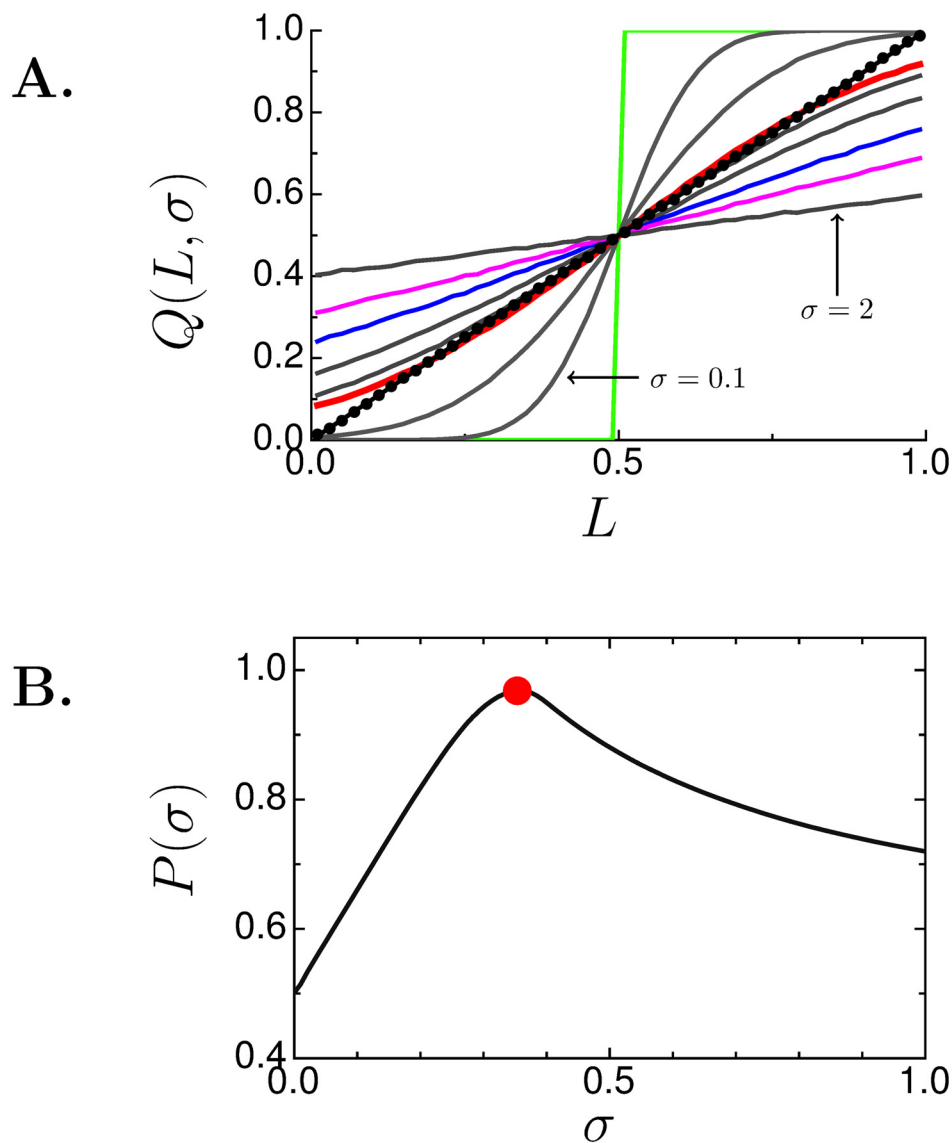


Fig 3. Sloppy rulers without Sloppy Algorithm. (A), length estimate ($Q(L, s)$) when s is fixed to 0 (light green), 0.1, 0.2, magic number (red), 0.4, 0.5, 0.7 (blue), 1 (magenta), and 2. Perfect fit falls on the diagonal line (circles). (B), performance $P(s)$ given by Eq (8). Maximum occurs at the magic number $\sqrt{2}/4 \approx 0.35$ (red circle).

doi:10.1371/journal.pone.0143475.g003

gets to being perfect for all L . This gauge is called the *performance*, $P(s)$, defined as the average of $B(L, s)$ over all values of L

$$P(s) = 1 - \frac{\int_0^1 |Q(L, s) - L| dL}{\int_0^1 \epsilon_{\max} dL} = 1 - 2 \int_0^1 |Q(L, s) - L| dL. \quad (8)$$

Like $B(L, s)$, $P(s)$ ranges between 0 (worst performance) and 1 (perfect performance).

$P(s)$, shown in Fig 3B, is maximized at the magic number (red circle). The performance at the magic number is 97% of the maximum value of 1. Perfect performance is impossible with a single distribution. However, the magic number distribution is the *optimal single distribution* that best approximates the infinitude of exact solutions obtained by the Sloppy Algorithm.

This means that a population of sloppy rulers drawn from the magic number distribution can make good estimates of L with high resolution (given enough rulers) *without* iterations. This is of practical importance as it might be very difficult to manipulate the underlying variability of the threshold distribution even once, much less iteratively. Manipulating the threshold distribution is easy with rulers or voltage comparators (used in ADCs) but when the individuals are people manipulating thresholds might be difficult though possible [7].

Because the magic number distribution gives a good estimate for almost all L , starting the Sloppy Algorithm loop using s_m as the initial guess produces an initial estimate λ_0 that will be almost always close to the correct value. This is shown by the red traces in Fig 2 where s_m was used as the starting guess.

Optimizing variability for maximum benefit: An economic example

Measurements with sloppy rulers might not appear to have any bearing on how diversity of crops or people affect population benefit but we will show there is a very close relationship between making accurate measurements and maximizing aggregate income of the population.

To better understand how variability affects population benefit and performance, imagine a town of farmers who independently choose to plant tulips or the less valuable beans. One might initially predict that every farmer working for his/her own best interest would plant tulips. However, when faced with the uncertainty that the tulip market might be glutted, some farmers might choose to grow beans given its certain marketability. Let us assume that a farmer overcomes his/her concern for a tulip glut and plants tulips when the demand for tulips exceeds his/her risk threshold. Some farmers are risk-takers, others are risk-averse. Assuming there is a distribution of risk threshold in the farmer population, we ask what is the optimal level of threshold variability that maximizes the aggregate income of all the farmers. The aggregate income might not be important to the individual farmer but is important to the population, that is, the town. The larger the aggregate income, the more the town flourishes.

Define the benefit \tilde{B} , as the revenue collected by a town based on the following assumptions: (a) the income to the town increases at a rate of b_1 per tulip up to the market demand D but receives no income for the tulips in excess of the demand. (b) Farmers not planting tulips plant beans whose value per farmer, b_2 , is less than that of tulips. Planting tulips incurs an *opportunity cost* by forgoing other activities such as planting beans. (c) A penalty, b_3 , is assessed per tulip for wasted resources in producing a number of tulips that is above or below the demand.

Suppose each farmer can produce β tulips; the town of N farmers can fulfill a maximum demand $D_m = N\beta$. The number of tulip farmers needed to exactly meet the demand D is $n^* = D/\beta$. Define L to be the ratio of demand to maximum demand, $L = D/D_m$ ($0 \leq L \leq 1$), so $D = \beta NL$. The normalized benefit B is defined as the ratio of benefit \tilde{B} to maximum demand

$$B(L, v) \equiv \frac{\tilde{B}(L, v)}{D_m} = [H(v)v - H(v-1)(v-1)]Lb_1 + \frac{b_2}{\beta}(1-Lv) - L|1-v|b_3 \quad (9)$$

H is the Heaviside function. v is the ratio of the number of farmers planting tulips, n , to the number needed to meet the demand, $v = n/n^*$. We get an important interpretation of v by rewriting it as $v = NQ(L, s)/(\beta LN)$, which is the *output to demand ratio*. The first term on the right hand side of Eq (9) is the income derived from tulips; it increases linearly until $v = 1$ then flattens because the town receives nothing by producing more tulips than demanded. In the second term, $1 - Lv$ is the fraction of farmers planting beans. We assume that the farmers earn

more by planting tulips than beans so $\beta b_1 > b_2$. The third term is the penalty for over- or under-producing tulips. Note that $B(L, \nu)$ is maximized when $\nu = 1$, when tulip output exactly matches demand. Fig 4C shows the *benefit landscape* $B(L, \nu)$. The thick white line at $\nu = 1$ demarcates the ridge of maximal benefit.

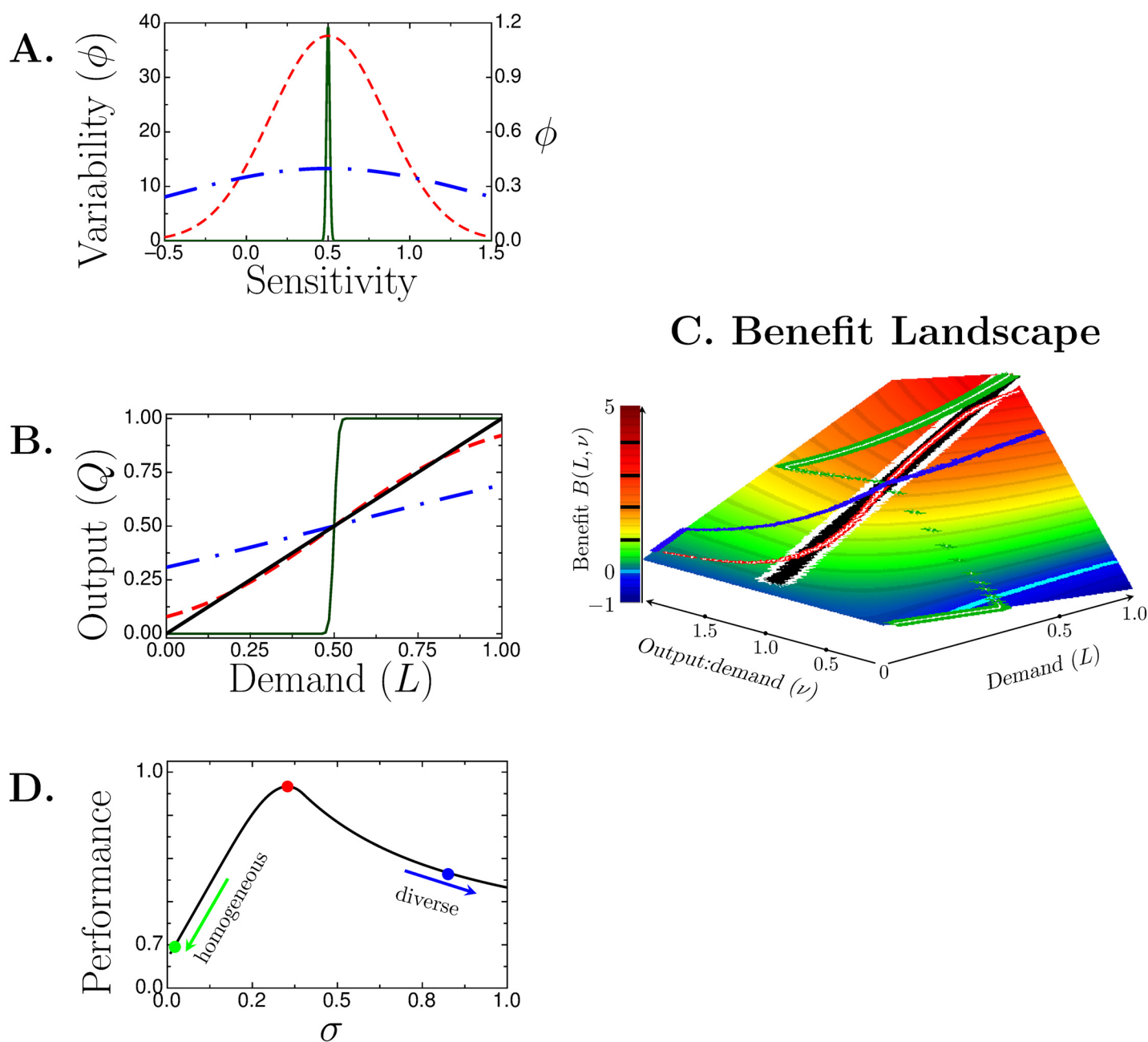


Fig 4. Farmer example. (A) Probability density function (ϕ) of farmers' sensitivities. $s = 0.01$, green (solid, left axis); magic number, red (dashed, right axis); 0.8, blue (dash-dot, right axis). (B) Farmers output relative to demand. Solid black line shows perfect matching between output and demand. Curves' colors and line patterns match those in A. (C), benefit landscape, $B(L, \nu)$. Thick white line lies on the ridge where B is maximized. Cyan curve near the L -axis is where $B = 0$; $B < 0$ for points below the curve (closer to L -axis) and $B > 0$ above the curve. Other colored curves are the benefits derived from farmer tulip output shown in panel (B). Black, Sloppy Algorithm; red, s = magic number; green, 0.01; and blue, 0.8. (D) Performance as a function of population variability s . Circles mark $s = 0.01$ (green), magic number (red), and 0.8 (blue).

doi:10.1371/journal.pone.0143475.g004

To link the benefit function to the farmer threshold distribution, suppose that each farmer independently decides whether to plant tulips or beans. The farmers' decisions are based on their risk threshold, which we assume is normally distributed about the mean of $D_m/2$ with standard deviation $s \cdot D_m$. Eager farmers start planting as soon as there is any demand while sluggards hold off until the demand becomes high. The fraction of farmers whose sensitivity is less than L is $Q(L, \mu = 1/2, s)$. The number of tulip planters is then $n = NQ(L, s)$ (where $\mu = 1/2$ is understood) so $v = NQ(L, s)/NL$, which equals the ratio of the town's tulip output to the demand. Because $B(L, v)$ is maximized when $v = 1$, when output equals demand, it follows that the optimal variability s^* solves $Q(L, s^*) = L$. This fixed point problem is identical to that of the ensemble of sloppy rulers (Eq (4)). This means that the problem of getting accurate measurements with sloppy rulers is identical to the problem of efficient labor allocation in this simple economy.

Just as a population of sloppy rulers whose graduation marks have optimal s^* variability can estimate the value of L exactly, a population of farmers whose sensitivities have variability s^* will produce exactly the amount of tulips as the market demands. For every demand L there is a unique optimal variability in the farmers' sensitivity $s^*(L)$ that results in a perfect matching of output to demand. If the threshold distribution can be altered with appropriate rewards or punishments [7], it is possible to use the Sloppy Algorithm to adjust s so that the collective output of the town's farmers exactly matches the demand. However, this may be neither practical nor necessary. As with the sloppy rulers, if the farmers' threshold distribution is the magic number distribution then their collective tulip output would almost exactly match the demand and thereby nearly maximize the benefit to the town. Importantly, this matching of output to demand occurs spontaneously by individual farmers making independent decisions, without needing a manager to dictate how many farmers in the town should plant tulips.

The benefits of optimal variability

To see how variability amongst farmers affects the collective benefits, consider three towns having farmers with different risk threshold distributions shown in Fig 4A. The farmers' output (Q , black line) relative to the demand (L) is shown in Fig 4B. In the town with a homogeneous population (green curves in Fig 4A and 4B), the farmers work in lock step and produce no tulips when the demand is low and too much when the demand is high. The benefit to the town is shown by the green trace on the benefit landscape in Fig 4C. Although this town produces no tulips when the demand is very low ($L < 0.2$) the town derives some positive benefit by planting beans (indexed by b_2). For more moderate demand ($0.2 < L < 0.5$) the penalty for not producing tulips (the b_3 term) outstrips the income from beans and the town suffers a deficit ($B < 0$, green curve dips below $B = 0$ indicated by the cyan-colored curve). Once the demand exceeds 0.5, all the farmers in this town plant tulips. The benefit increases but overproduction occurs (Fig 4B). The net benefit is submaximal because overproduction is penalized and, because all farmers are planting tulips, the town forfeits income from beans.

In contrast, the town of farmers with an overly broad distribution of risk thresholds (Fig 4A, blue dot-dashed curves) produces too much when the demand is low and too little when the demand is high (Fig 4B). The benefit this town derives is submaximal (blue curve, Fig 4C) because it is penalized for overproduction at low demand and also penalized for underproduction of valuable tulips because too many farmers plant the less valuable beans.

The town of farmers with the magic number distribution (red, dashed curve, panel Fig 4A) outputs tulips in near perfect accord with demand (Fig 4B) and the benefit (red-white curve in Fig 4C) is close to the theoretical perfect line except when the demand is very low because then too many farmers plant tulips.

To quantify how the town's benefit changes with s , we define the performance $P(s)$ as the ratio of the integral of the benefit (Eq (9)) over all L for a given s to the maximal achievable benefit,

$$P(s) = \frac{\int_0^1 B(L, v(s)) dL}{\int_0^1 B(L, v(s^*(L))) dL}. \quad (10)$$

$P(s)$ ranges from 0 (zero income) to 1 (maximal achievable income). $P(s)$ is shown in Fig 4D. The town with the magic number distribution (red circle) performs the best with $P(s_m) = 98\%$. By contrast, the homogeneous town ($s = 0.01$, green circle) has a performance 67%, and the overly diverse town ($s = 0.8$, blue circle) has a performance of 83%.

Optimal variability depends on the population benefit function

We can intuitively understand how a group of people who function well to accomplish one goal might perform poorly when given another goal. Here we quantify how performance changes when the benefit function is redefined. The key result of this section is that the optimal level of variability depends on the benefit function.

Let us define a new benefit function

$$B(L, v) = \frac{f(L)}{L} v - \frac{v^2}{2}, \quad (11)$$

where L and v have their same meaning as in the farming example. Differentiating B with respect to v shows that B is maximized along $v = f(L)/L$. This implies that the population output is optimal when the variability s^* solves $Q(L, s^*) = f(L)$. For both the ruler and farming cases, $f(L)$ equaled L . Now suppose that $f(L)$ is the cubic function $f(L) = L - (\gamma/2)L \cdot (L - 1/2) \cdot (L - 1)$ for $0 \leq \gamma \leq 4$ (these constraints keep $f(L) \geq 0$). s^* is found for any L using the Sloppy Algorithm.

The benefit landscape is shown in Fig 5A. The black curve shows the ridge where B is maximal. To achieve maximum benefit s^* must be used, that is, for every demand L a different population must be used. As with the farmers, population diversity might not be practically changeable. Therefore we would like to see if there is a single distribution that produces near maximal benefit for all L . In other words, is there a magic number distribution?

In the case where $f(L) = L$ and ϕ is normal, we could determine the magic number s_m analytically. In other cases, such as the one considered now, s_m is defined as the value that maximizes the performance $P(s)$ (Eq (10)) and is found numerically.

Here, it turns out that $s_m = 0.242$ and $\phi(L, s = 0.242)$ is shown in by the red trace in Fig 5B. The output $Q(L, s_m)$ (red trace in panel C) for this magic number distribution matches very closely the ideal output $Q(L, s^*)$ (black trace). The benefit, shown by the red-white curve in Fig 5A, deviates significantly from maximal only when the demand is low. The performance of this magic number distribution is 0.94.

Suppose the town of farmers with the magic number distribution ($s_m = \sqrt{2}/4 \approx 0.35$) that fared well with the benefit function defined by Eq (9) is now evaluated with the new benefit function (Eq (11)). How would this town fare? This town's distribution and output are shown by the blue traces in Fig 5B and 5C. The benefit (blue curve in Fig 5A) deviates markedly from maximal. This town's performance was 0.98 for the earlier benefit function (Eq (9)), but for the new benefit function, its performance is a paltry 0.037. For this new benefit function, this town is too diverse ($s_m = 0.35$). A modest reduction in the variability to $s_m = 0.242$ (making the town's farmers slightly more homogeneous) will boost the town's performance 25-fold to 0.94.

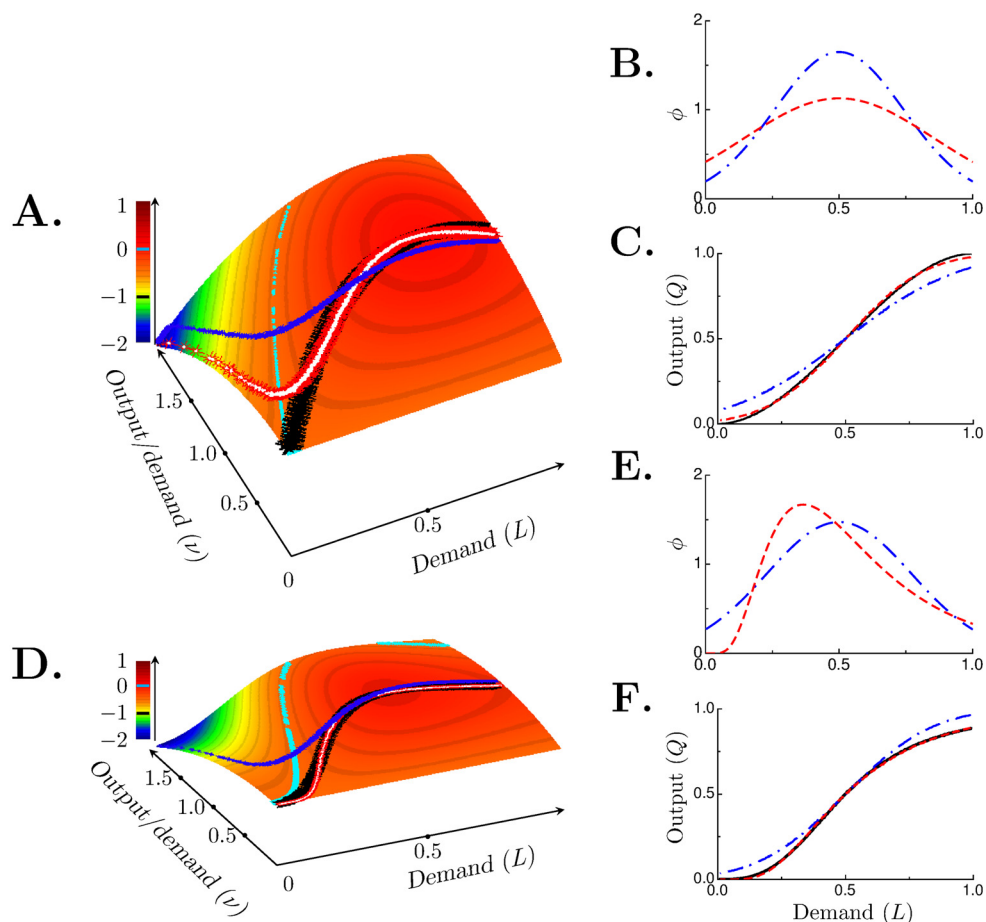


Fig 5. Benefit function determines optimal distribution. (A), benefit landscape defined by Eq (11) when $f(L)$ is cubic ($\gamma = 4$). (B), normal distributions with $s = 0.242$ (red, dashed curve) or 0.354 (blue, dot-dashed curve). The former is the magic number for this benefit function while the latter is the magic number of the benefit function in the tulip farmer case. (C) shows the corresponding outputs of these populations; black curve is $f(L)$. Sloppy Algorithm output matches $f(L)$ exactly. Heavy black curve in panel (A) lies along the ridge of maximal benefit. Red (with white dots added for clarity) and blue curves on the benefit landscape are the benefit corresponding to the outputs in panel (B). (D), benefit landscape defined by Eq (11) when $f(L)$ is sigmoidal. (E), magic number lognormal (red, dashed curve) and magic number normal (blue, dot-dashed curve) population decision models. The magic numbers are $s_m = 0.559$ (shape factor) for the lognormal and $s_m = 0.271$ (standard deviation) for the normal distribution. (F), outputs from corresponding distributions; black curve is $f(L)$. Heavy black curve in panel (D) mark the ridge of maximal benefit. Red (with white dots) curve is the benefit for the lognormal distribution and the blue curve is for the normal distribution. Cyan-colored curves in (A) and (D) show where $B(L, \nu) = 0$; points closer to the demand axis are positive.

doi:10.1371/journal.pone.0143475.g005

This example shows that the optimal level of variability in a population depends strongly on the benefit function. A population that is optimally diverse for one benefit function might do very poorly when the definition of benefit changes.

Optimal variability depends on the threshold distribution

In this section we show that performance depends not only on the benefit function but also on the type of underlying threshold distribution. Up to now we assumed that the threshold distribution was normal. However, the lognormal distribution occurs frequently in biology [8, 9]. For example, mRNA [10] and protein levels [11, 12] are distributed lognormally.

Furthermore, the magnitude of a cellular response is often a sigmoidal function of the signal amplitude. We therefore use the benefit function given above (Eq (11)) where $f(L)$ is now the sigmoidal function $L^n/(0.5^n + L^n)$. We compare how two threshold distributions, lognormal and normal, fare with this benefit function. The Sloppy Algorithm will find the optimal shape factor and the standard deviation, for the lognormal and normal distribution respectively, that maximizes the benefit function. When these optimal values are used the output is given by the black trace in Fig 5F and the benefit is maximal as shown by the black curve in Fig 5D.

As before we would like to replace the continuum of threshold distributions that give perfect benefit ($B(L, s^*) = 1$) with a single distribution that gives good performance. Magic numbers were calculated for both models. For the lognormal distribution, the magic number for the shape factor is $s_m = 0.559$; for the normal distribution, the magic number is $s_m = 0.242$. The red trace in Fig 5E is the lognormal magic number distribution and the blue trace is the normal magic number distribution. The output for the lognormal distribution (red trace in Fig 5F) is very close to the optimal (black trace) while the output of the normal distribution (blue trace) deviates considerably. The corresponding benefits are shown by the red-white and blue curves in Fig 5C. Consonant with near perfect output, the performance of the population with the lognormal distribution is 0.999. The normal distribution population's performance is only 0.727.

What accounts for the differences in performances of these two distributions? The lognormal distribution works well because the function $Q(L, s_m)$ using the lognormal “looks like” the sigmoidal $f(L)$ while $Q(L, s_m)$ using the normal distribution does not “look like” $f(L)$. By “looks like” we mean close in the L_∞ norm (see Section 3 in S1 File).

This example shows that for a given benefit function, there might be no single population distribution of a particular type (normal in this example) that performs well. However, another type of population distribution (lognormal here) might work very well. Whether it is feasible to shape the population decision model will depend, of course, on the nature of the population.

Discussion

Reframing the diversity debate

What is the impact of variability or diversity in a population? This question has profound implications for industry, agriculture, bioengineering, and also for social institutions. Reasonable and compelling arguments have been made to support the beneficial effects of either diversity or homogeneity in a population. For example, some have called for supporting a diverse population of computer operating systems to thwart the widespread infection by computer viruses and worms [13, 14]. Others have countered that having a few operating systems allows programmers to concentrate their efforts to reduce vulnerabilities to cyber-attacks in these few [15]. Similarly, agriculture and forestry reap the benefits of product uniformity, high yields, and ease of management by using a small range of species. But catastrophic failure can occur should these few species succumb to disease or pests [4, 16–18]. An economy based on a broader diversity of crops is unlikely to fail catastrophically but might not enjoy the benefits of high productivity. The question on the value of diversity in society has sparked intense and often contentious social and political debates. The landmark Supreme Court decision of *Regents of the University of California v. Bakke* (438 U.S. 265 (1978)) hinged on the principle that a diverse student body is essential to the quality of higher education. But as with computers and agriculture, either broad or narrow diversity may bring trade-offs. A study by Putnam showed that broadly diverse communities had lower trust amongst its citizens than less diverse ones [19], but greater creativity had also been seen in more diverse groups [20, 21]. Another study by Watson et al. [22] showed that in the workplace, heterogeneous groups initially had

more difficulty working together but after seventeen weeks they surpassed the homogeneous group at problem identification and generating alternative solutions.

Is it better to have diversity or homogeneity? The often contentious debate about the value of diversity in the workplace, schools, and society often stems, in part, from the Procrustean attempt to force the effects of diversity into a good-or-bad dichotomy. We argue that this is the wrong question. Instead, the proper and useful question is “What level of variability maximizes population benefit?”

To our knowledge only a few studies have addressed this question. Sih and Watters [23] measured the mating activity in a population of water striders as a function of the proportion of inactive, or moderately aggressive, or hyper-aggressive males in a group. They found that mating activity (a type of benefit) was a hump-shape function of the aggressiveness level, which is qualitatively similar to the performance curves we show in Figs 3B and 4D. Groups dominated by inactive males did not mate frequently and groups dominated by hyper-aggressive males drove females away.

New insights on the existence and persistence of variability

In biological systems variability is ubiquitous and sometimes puzzling. Phenotypic variations occur even in genetic clones. Johannsen showed that genetically identical bean plants produced beans with broadly distributed sizes [24]. Recent studies extend Johannsen’s bean size variation to findings of broad variations in protein levels between isogenic cells [10–12] and in behaviors such as latency to enter S-phase [25] and apoptosis time [11]. Phenotypic variability is puzzling because in a static and spatially homogeneous environment, natural selection is supposed to favor the most fit, eliminate the rest, thus *narrowing* the range of variability [26–28]. However, the world is neither static nor homogeneous over long time. In a fluctuating and unpredictable environment a species that maintains a broad set of phenotypes, a strategy known as “bet-hedging”, might have a better chance of survival over multiple generations [28–31]. Nevertheless, could phenotypic variability be important on short time scales? Our work now provides a new insight into why variability is important to the population even in the lifetime of an individual.

The key to understanding this is to shift the focus from individuals to the population. This shift of focus allows us to see so called “cheaters” in a new light. Yeast cannot directly metabolize sucrose but when grown in a sucrose medium they express invertase, which cleaves sucrose into metabolizable glucose [32, 33]. Because invertase is expressed on the cell membrane, the glucose produced is free to diffuse into the surrounding medium and be used by other cells. Some cells, called “cheaters”, do not express invertase but benefit from glucose produced by other cells [34]. Because cells incur a metabolic cost to produce invertase, cheaters have a reproductive advantage over invertase producers [34]. Thus, one might expect the population to eventually converge to a homogeneous population of cheaters. But, of course, a population composed solely of cheaters would soon starve. Conversely, if all cells produced invertase there would be a glut of glucose and less reproduction. There is a clear analogy between the yeast and farmers. The town maximizes its income when there is an optimal diversity of farmers so there are enough farmers planting tulips to meet the market demand but not so much as to glut the market. The town is penalized when too many farmers plant tulips because fewer farmers are left to plant marketable beans. Likewise, there is no need for all yeast cells expend energy making invertase and produce a glut of glucose; the “cheaters” can devote their energy to reproduction. If we define the population benefit function as the growth rate, there would be an optimal mix of slow-growing invertase producers and fast-growing “cheaters” that maximizes the growth rate of the colony as a whole. D.S. Wilson [26] writes that differences between

individuals within a population may be “end product of natural selection” rather than “the raw material on which natural selection acts.” In other words, evolution may be selecting for organisms that maintain an optimal level of phenotypic variability. Genetic noise in gene translation or transcription [35–37] may also serve to maintain this level of variability.

The mathematical framework we developed here provides a tool for answering the question “What is the optimal level of diversity in a population that maximizes benefit?” We hope our work prompts new ways of thinking about and analyzing the effect of variability on a population.

Supporting Information

S1 File. Mathematical proof for convergence Sloppy Algorithm, convergence demonstrations, and differences between the Sloppy Algorithm and dithering. This document has five sections. Section 1 has a proof for the convergence of the Sloppy algorithm. Section 2 shows how the magic number is explicitly computed when $\phi \in \mathcal{N}$ and $f(L) = L$. Section 3 gives the condition when a useful magic number can be found. Section 4 gives an example where a person is part of the Sloppy Algorithm and has the task of determining the absolute gray-scale value. Section 5 describes the differences between sloppy rulers and dithering. (PDF)

Acknowledgments

This work was supported by NIH grants R01 HL090880 (to LTI and YC) and R03 AG031944 (YC) and start-up funds from the University of California, Davis. We thank Professor Michael Savageau (UC Davis) for discussion and comments on this manuscript.

Author Contributions

Conceived and designed the experiments: LTI TB YC. Performed the experiments: LTI YC. Analyzed the data: LTI. Contributed reagents/materials/analysis tools: LTI. Wrote the paper: LTI YC. Developed Sloppy Algorithm and convergence proof: LTI. Pointed out critical flaw and proposed solution: TB.

References

1. Box G, Bisgaard S. Statistical Tools for Improving Designs. Mechanical Engineering. 1988; 110:32–40.
2. Goh TN. The Role of Statistical Design of Experiments in Six Sigma: Perspectives of a Practitioner. Quality Engineering. 2002; 14:659–671. doi: [10.1081/QEN-120003565](https://doi.org/10.1081/QEN-120003565)
3. Nordström F, Gawad P, Nowarski A. The science of manufacturing. ABB Review. 2006;p. 6–11.
4. Adams M W AHE, Rossman EC. Biological Uniformity and Disease Epidemics. BioScience. 1971 Nov; 21(21):1067–1070. doi: [10.2307/1295991](https://doi.org/10.2307/1295991)
5. Etchenique R, Aliaga J. Resolution enhancement by dithering. Am J Physics. 2004; 72:159–163. doi: [10.1119/1.1613274](https://doi.org/10.1119/1.1613274)
6. Lipshitz SP, Wannamaker RA, Vanderkooy J. Quantization and dither: A theoretical survey. J Audio Eng Soc. 1992; 40:355–375.
7. McNichol D. A Primer of Signal Detection. Psychology Press;.
8. Limpert E, Stahel WA, Markus A. Log-normal distributions across the sciences: Keys and clues. BioScience. 2001; 51(5):341–352. doi: [10.1641/0006-3568\(2001\)051%5B0341:LNDATS%5D2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051%5B0341:LNDATS%5D2.0.CO;2)
9. Furusawa C, Suzuki T, Kashigawa A, Yomo T, Kaneko K. Ubiquity of log-normal distributions in intracellular reaction dynamics. Biophysics. 2005; 1:25–31. doi: [10.2142/biophysics.1.25](https://doi.org/10.2142/biophysics.1.25)
10. Bengtsson M, Ståhlberg A, Rorsman P, Kubista M. Gene expression profiling in single cells from the pancreatic islets of Langerhans reveals lognormal distribution of mRNA levels. Genome Res. 2005; 15:1388–1392. doi: [10.1101/gr.3820805](https://doi.org/10.1101/gr.3820805) PMID: [16204192](https://pubmed.ncbi.nlm.nih.gov/16204192/)

11. Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK. Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature*. 2009; 459:428–432. doi: [10.1038/nature08012](https://doi.org/10.1038/nature08012) PMID: [19363473](https://pubmed.ncbi.nlm.nih.gov/19363473/)
12. Newman JR, Ghaemmamghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, et al. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*. 2006; 441:840–846. doi: [10.1038/nature04785](https://doi.org/10.1038/nature04785) PMID: [16699522](https://pubmed.ncbi.nlm.nih.gov/16699522/)
13. Eichin MW, Rochlis JA. With microscope and tweezers: An analysis of the internet virus of November 1988. In: *Proceedings of the IEEE Symposium on Research in Computer Security and Privacy*. IEEE. IEEE Computer Society Press; 1989. p. 326–343.
14. Stamp M. Risks of monoculture. *Communications of the ACM*. 2004 March; 47(3):120. doi: [10.1145/971617.971650](https://doi.org/10.1145/971617.971650)
15. Schneider FB, Birman P. The monoculture risk put into context. *IEEE Security and Privacy*. 2009 Jan/Feb; p. 15–17.
16. Ullstrup AJ. The impacts of the southern corn leaf blight epidemics of 1970–71. *Ann Rev Phytopathol*. 1972; 10:37–50. doi: [10.1146/annurev.py.10.090172.000345](https://doi.org/10.1146/annurev.py.10.090172.000345)
17. Bishir J, Roberds JH. On number of clones needed for managing risk in clonal forestry. *Forest Genetics*. 1999; 6(3):149–155.
18. Carnus JM, Parrotta J, Brockerhoff E, Arbez M, Jactel H, Kremer A, et al. Planted Forests and Biodiversity. *Journal of Forestry*. 2006-03-01T00:00:00; 104(2):65–77.
19. Putnam RD. *E Pluribus Unum: Diversity and Community in the Twenty-first Century* The 2006 Johan Skytte Prize Lecture. *Scandinavian Political Studies*. 2007; 30:137–174. doi: [10.1111/j.1467-9477.2007.00176.x](https://doi.org/10.1111/j.1467-9477.2007.00176.x)
20. Alesina A, La Ferrara E. Ethnic diversity and economic performance. *Journal of Economic Literature*. 2005; 43(3):762–800. doi: [10.1257/002205105774431243](https://doi.org/10.1257/002205105774431243)
21. Page S. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools and Societies*. Princeton University Press;.
22. Watson WE, Kumar K, Michaelsen LK. Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of Management Journal*. 1993 Jun; 36(3):590–602. doi: [10.2307/256593](https://doi.org/10.2307/256593)
23. Sih A, Watters JV. The mix matters: behavioural types and group dynamics in water striders. *Behaviour*. 2005; 142:1417–1431. doi: [10.1163/156853905774539454](https://doi.org/10.1163/156853905774539454)
24. Johannsen W. The genotype conception of heredity. *AmNaturalist*. 1911 Mar; 45(351):129–159. doi: [10.1086/279202](https://doi.org/10.1086/279202)
25. Albeck JG, Mills GB, Brugge JS. Frequency-modulated pulses of ERK activity transmit quantitative proliferation signals. *Mol Cell*. 2013; 49(2):249–261. doi: [10.1016/j.molcel.2012.11.002](https://doi.org/10.1016/j.molcel.2012.11.002) PMID: [23219535](https://pubmed.ncbi.nlm.nih.gov/23219535/)
26. Wilson DS. Adaptive Individual Differences within Single Populations. *Philosophical Transactions: Biological Sciences*. 1998; 353(1366):pp. 199–205. doi: [10.1098/rstb.1998.0202](https://doi.org/10.1098/rstb.1998.0202)
27. Pruitt JN, Riechert SE. How within-group behavioural variation and task efficiency enhance fitness in a social group. *Proc R Soc B*. 2011; 278:1209–1215. doi: [10.1098/rspb.2010.1700](https://doi.org/10.1098/rspb.2010.1700) PMID: [20943687](https://pubmed.ncbi.nlm.nih.gov/20943687/)
28. Seger J, Brockman HJ. What is bet-hedging. In: Harvey PH, Partridge L, editors. *Oxford Surveys in Evolutionary Biology*. Oxford University Press; 1987. p. 182–211.
29. Kussell E, Leibler S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*. 2005; 309(5743):2075–2078. doi: [10.1126/science.1114383](https://doi.org/10.1126/science.1114383) PMID: [16123265](https://pubmed.ncbi.nlm.nih.gov/16123265/)
30. Veening JW, Smits WK, Kuipers OP. Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol*. 2008; 62. doi: [10.1146/annurev.micro.62.081307.163002](https://doi.org/10.1146/annurev.micro.62.081307.163002) PMID: [18537474](https://pubmed.ncbi.nlm.nih.gov/18537474/)
31. Beaumont HJ, Gallie J, Kost C, Ferguson GC, Rainey PB. Experimental evolution of bet hedging. *Nature*. 2009; 462(7269):90–93. doi: [10.1038/nature08504](https://doi.org/10.1038/nature08504) PMID: [19890329](https://pubmed.ncbi.nlm.nih.gov/19890329/)
32. Carlson M, Botstein D. Two differentially regulated mRNAs with different 5' ends encode secreted with intracellular forms of yeast invertase. *Cell*. 1982; 28(1):145–154. doi: [10.1016/0092-8674\(82\)90384-1](https://doi.org/10.1016/0092-8674(82)90384-1) PMID: [7039847](https://pubmed.ncbi.nlm.nih.gov/7039847/)
33. Özcan S, Vallier LG, Flick JS, Carlson M, Johnston M. Expression of the SUC2 Gene of *Saccharomyces cerevisiae* is Induced by Low Levels of Glucose. *Yeast*. 1997; 13(2):127–137. doi: [10.1002/\(SICI\)1097-0061\(199702\)13:2%3C127::AID-YEA68%3E3.3.CO;2-R](https://doi.org/10.1002/(SICI)1097-0061(199702)13:2%3C127::AID-YEA68%3E3.3.CO;2-R) PMID: [9046094](https://pubmed.ncbi.nlm.nih.gov/9046094/)
34. Gore J, Youk H, van Oudenaarden A. Snowdrift game dynamics and facultative cheating in yeast. *Nature*. 2009; 459:253–256. doi: [10.1038/nature07921](https://doi.org/10.1038/nature07921) PMID: [19349960](https://pubmed.ncbi.nlm.nih.gov/19349960/)
35. Blake WJ, Balazsi G, Kohanski MA, Isaacs FJ, Murphy KF, Kuang Y, et al. Phenotypic consequences of promoter-mediated transcriptional noise. *Mol Cell*. 2006 Dec; 24(6):853–865. doi: [10.1016/j.molcel.2006.11.003](https://doi.org/10.1016/j.molcel.2006.11.003) PMID: [17189188](https://pubmed.ncbi.nlm.nih.gov/17189188/)

36. Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. PNAS. 2001; 98:8614–8619. doi: [10.1073/pnas.151588598](https://doi.org/10.1073/pnas.151588598) PMID: [11438714](https://pubmed.ncbi.nlm.nih.gov/11438714/)
37. Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. Nature. 2010 Sep; 467(7312):167–173. doi: [10.1038/nature09326](https://doi.org/10.1038/nature09326) PMID: [20829787](https://pubmed.ncbi.nlm.nih.gov/20829787/)

Supporting Information: File S1

Optimizing Population Variability to Maximize Benefit

Leighton T. Izu, Tamás Bányász, and Ye Chen-Izu

1 Convergence proof for Sloppy Algorithm

Notation and assumptions

Our goals in these appendices are:

1. Prove that the Sloppy Algorithm converges monotonically everywhere when ϕ is normally distributed and $f(L) = L$.
2. We will adapt the proof to the case where $f(L)$ is sigmoidal and ϕ is lognormal.
3. We will show how the magic number is derived when ϕ is normal and $f(L) = L$.
4. We show what it means to have a useful magic number.

Notation: (1) Strictly monotonically increasing, SMI. (2) Monotone convergence theorem, MCT. (3) $\mathcal{N}(\mu, s)$ is the normal distribution with mean μ and standard deviation s . (4) $\mathcal{LN}(m, s)$ is the lognormal distribution with median m and shape factor s .

Assumptions 1 and 2 hold throughout these appendices.

Assumption 1: Let $f(L)$ be strictly monotonically increasing (SMI) on $L = [0, 1]$ and let it be bounded by $0 \leq f(L) \leq 1$.

Assumption 2: Let $\phi(x, s)$ be a probability distribution function (pdf) and let $Q(L, s)$ be

$$Q(L, s) = \int_{\Gamma(L)} \phi(t, s) dt \tag{S1}$$

$$= \frac{1}{2} [1 + q(L, s)]. \tag{S2}$$

The following rules define the Sloppy Algorithm.

1. Choose s_0 arbitrarily.
2. Compute $Q(L, s_0) = f_1$.
3. For any $i \geq 1$

$$\text{choose } s_i \text{ such that } Q(\lambda_i, s_i) = f_i \text{ where } \lambda_i \text{ solves } f(\lambda_i) = f_i. \quad (\text{S3})$$

4. Compute the next f_{i+1} using

$$Q(L, s_i) = f_{i+1}. \quad (\text{S4})$$

Remarks. Monotonicity of $f(L)$ is needed for the inverse function $f^{-1}(x)$ to exist. $Q(L, s)$ is monotonic in L because Q is the integral of a pdf, which is everywhere non-negative.

General lemmas

Lemma 1 (Dominance). *If $f_i < f(L)$ then $f_{i+1} > f_i$.*

Proof: The proof hinges on the monotonicity of $f(L)$ and $Q(L, s)$. The monotonicity of $f(L)$ tells us that

$$\text{If } f_i \equiv f(\lambda_i) < f(L) \text{ then } \lambda_i = f^{-1}(f_i) < L. \quad (\text{S5})$$

From (S3) we get $f_i = Q(\lambda_i, s_i)$ and from (S4) we get $f_{i+1} = Q(L, s_i)$. Suppose to the contrary that $f_{i+1} \leq f_i$ then $f_i = Q(\lambda_i, s_i) \geq Q(L, s_i) = f_{i+1}$. Because Q is monotonic in L it follows that $\lambda_i \geq L$. However, because of the monotonicity of $f(L)$ it follows [statement (S5)] that $f(\lambda_i) = f_i \geq f(L)$, which contradicts our assumption. Therefore, $f_{i+1} > f_i$ as claimed.

□

Lemma 1'. *If $f_i > f(L)$ then $f_{i+1} < f_i$.*

Proof is similar to Lemma 1.

Lemma 2 (Uniqueness). *$f(L)$ is the unique accumulation point.*

Proof: Suppose there is $f(\tilde{L})$ which is the limit of the sequence of the Sloppy Algorithm and we assume without loss of generality that $f(\tilde{L}) < f(L)$. From (S4) we have

$$\lim_{i \rightarrow \infty} Q(L, s_i) = \lim f_{i+1} = f(\tilde{L}).$$

And from (S3) we have

$$\lim_{i \rightarrow \infty} Q(f^{-1}(f_i), s_i) = \lim f_i = f(\tilde{L}).$$

Equating these two expressions give

$$Q(L, s) = Q(f^{-1}(f(\tilde{L})), s).$$

But if $f(\tilde{L}) < f(L)$ then by statement S5 it follows that

$$f^{-1}(f(\tilde{L})) = \tilde{L} < f^{-1}(f(L)) = L.$$

But because Q is monotonic in L then $Q(L, s) > Q(f^{-1}(f(\tilde{L})), s)$ contradicting the equality. Therefore, $f(\tilde{L}) \geq f(L)$.

We can argue similarly that $f(\tilde{L}) \leq f(L)$ therefore, $f(\tilde{L}) = f(L)$ and thus proving uniqueness.

□

Remark: We proved Lemmas 1 and 2 without specifying f or ϕ so they hold generally.

Convergence of the Sloppy Algorithm when $\phi = \mathcal{N}$ and $f(L) = L$

The Sloppy Algorithm will not, in general, converge monotonically for any pair $(f(L), \phi)$. In the case where $f(L) = L$ and ϕ is the normal distribution, the Sloppy Algorithm converges monotonically everywhere. The proof given below can be adapted to study convergence of the Sloppy Algorithm for any $(f(L), \phi)$ pair.

Theorem 1 (Convergence Theorem). *Let $f(L) = L$ on $L \in [0, 1]$. Let ϕ be the normal distribution $\mathcal{N}(L, 0, s)$ with mean of zero and standard deviation s . $Q(L, s)$ is defined by (S2) where Γ is the range from $1/2 - L$ to ∞ . Then $\{f_i\}$ defined by recursion rules (S3) and (S4) converges monotonically to $f(L)$ everywhere.*

Note that in this case $f_i = \lambda_i$.

Proof: The key to the proof and understanding whether the sequence $\{f_i\}$ converges monotonically depends on the shape of the $s^*(L)$ curve. $s^*(L)$ solves the fixed point problem

$$f(L) = Q(L, s^*(L)). \quad (\text{S6})$$

For the pair $(f(L), \phi) = (L, \mathcal{N}(L, s))$, $Q(L, s)$ is

$$Q(L, s) = \frac{1}{s\sqrt{2\pi}} \int_{1/2-L}^{\infty} e^{-z^2/2s^2} dz \quad (\text{S7})$$

$$= \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\sqrt{2}(2L-1)}{4s} \right) \right]. \quad (\text{S8})$$

$s^*(L)$ is found by solving (S6) with $f(L) = L$,

$$s^*(L) = \frac{\sqrt{2}(2L-1)}{4 \operatorname{erf}^{-1}(2L-1)}, \quad (\text{S9})$$

which is shown in Figure A. Note that as $L \rightarrow 1/2$, $\operatorname{erf}^{-1}(2L-1) \rightarrow \frac{\sqrt{\pi}}{2}(2L-1)$ so $s^*(1/2) = 1/\sqrt{2\pi} \approx 0.4$. We now have the pieces needed to complete the proof.

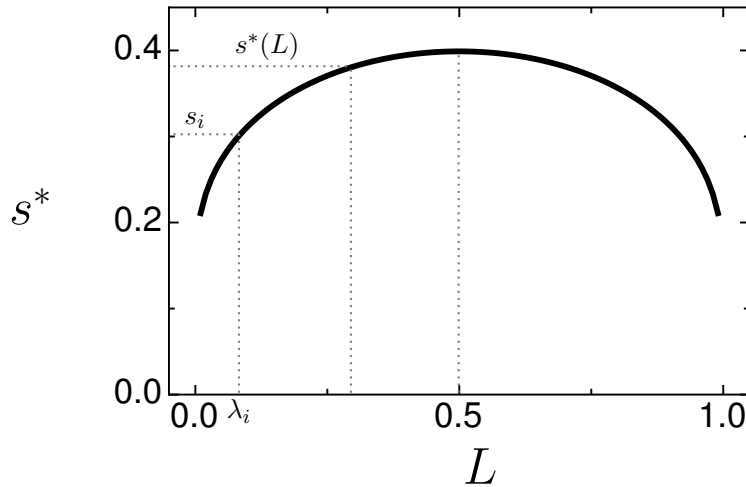


Figure A: Optimal standard deviation s when ϕ is normally distributed and $f(L) = L$.

Case 1A: Suppose λ_i and L lie on the same side of $1/2$, say, $\lambda_i < L < 1/2$ (Figure A). From Lemma 1 we know that $\lambda_{i+1} > \lambda_i$ but we do not yet know

whether λ_{i+1} is greater than or less than L . Define z as

$$z(s) = \frac{\sqrt{2}(2L - 1)}{4s}. \quad (\text{S10})$$

Recall that $\lambda_{i+1} = Q(L, s_i) = Q(z_i)$ (step (S4) in the Sloppy Algorithm). Because $s_i < s^*(L)$ (see Figure A) and because $L < 1/2$ (which makes the numerator < 0) it follows that $z_i < z^*(L)$. Because Q is SMI in z , it follows that $\lambda_{i+1} = Q(z_i) < Q(z^*(L)) = L$. Thus $\{\lambda_i\}$ is a SMI sequence bounded above by L so by the monotone convergence theorem (MCT), the sequence converges to some $\tilde{\lambda}$. However, Lemma 2 tells us that $\tilde{\lambda} = L$.

Case 1A': The cases where $L < \lambda_i < 1/2$, $1/2 < L < \lambda_i$, and $1/2 < \lambda_i < L$ can be handled in the same way to show $\{\lambda_i\}$ is a SM increasing (decreasing) bounded above (below) by L .

Case 1B: Suppose $\lambda_i < 1/2 < L$. We want to show that there is some $k > i$ for which $\lambda_k > 1/2$. Suppose no such k exists so for all $k > i$, $\lambda_k \leq 1/2$. From Lemma 1 we know that $\lambda_{i+1} > \lambda_i$. Therefore $\{\lambda_k\}$ is a SMI sequence bounded by $1/2$ so by the MCT $\lambda_k \rightarrow \tilde{\lambda}$. However, Lemma 2 demands that $\tilde{\lambda} = L$ therefore, contrary to our assumption, there must have been some k where $\lambda_k > 1/2$. Beyond this k , the situation is identical to Case 1A or 1A'.

Because i was arbitrary, it follows that any sequence $\{\lambda_i\}$ generated by the Sloppy Algorithm converges monotonically to L .

□

Figure B shows the monotonic convergence of the Sloppy Algorithm. Each colored path represents a different L . Note that for $L > 1/2$, $\lambda_{i+1} > \lambda_i$ while the opposite is true for $L < 1/2$.

The Sloppy Algorithm when $\phi = \mathcal{LN}$ and $f(L)$ is sigmoidal

Let

$$f(L) = \frac{L^n}{K^n + L^n} \text{ where } 0 \leq L \leq 1. \quad (\text{S11})$$

Substituting $\phi = \mathcal{LN}$ into (S2) and integrating between 0 and ℓ gives

$$Q(\ell, m, s) = \frac{1}{2} \left[1 + \operatorname{erf} \left(\frac{\ln(\ell) - m}{\sqrt{2s}} \right) \right].$$

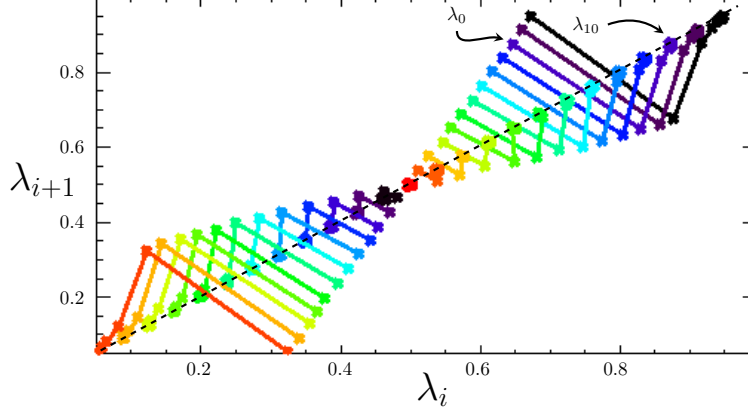


Figure B: **Global monotonic convergence when ϕ is normal and $f(L) = L$.** The $i + 1$ -st estimate of λ is plotted against the i -th estimate. Curved arrows point to λ_0 and λ_{10} for $L = 0.875$. Note that when $L > 1/2$, $\lambda_{i+1} > \lambda_i$ for all i , which indicates monotonic convergence. For $L < 1/2$, the opposite holds. In all cases, the initial value of s was 1.

$Q(\ell, m, s) = f(\ell)$ is satisfied when $s(\ell)$ is given by

$$s(\ell, m) = \frac{\ln(\ell) - m}{\sqrt{2}\text{erf}^{-1}(2f(\ell) - 1)}. \quad (\text{S12})$$

Let $\bar{\ell}$ solve

$$2f(\bar{\ell}) - 1 = 0. \quad (\text{S13})$$

Then m must be

$$m = \ln(\bar{\ell}). \quad (\text{S14})$$

The plot of $s(\ell)$ is shown in Figure C.

Case 2A: $\lambda_i < L < \bar{\ell}$. From Lemma 1, $f_i < f_{i+1}$ but we don't know if f_{i+1} is greater than or less than $f(L)$. By definition $f_{i+1} = Q(L, s_i)$ and $Q(L, s^*) = f(L)$. From Figure C we see that $s_i > s^*(L)$. Let

$$z_i = \frac{\ln(L/\bar{\ell})}{\sqrt{2}s_i}.$$

Because $L < \bar{\ell}$, the numerator is < 0 so $z_i > z^*$. Therefore, $f_i < f(L) < f_{i+1}$ and we do not get monotonic convergence.

Case 2B: $\bar{\ell} < L < \lambda_i$. By Lemma 5, $f_{i+1} < f_i$ but we don't know whether f_{i+1} is greater than or less than $f(L)$. From Figure C we see that $s^* < s_i$. Therefore, $z^* > z_i$ (now the numerator is > 0) therefore, $f_{i+1} = Q(L, s_i) <$

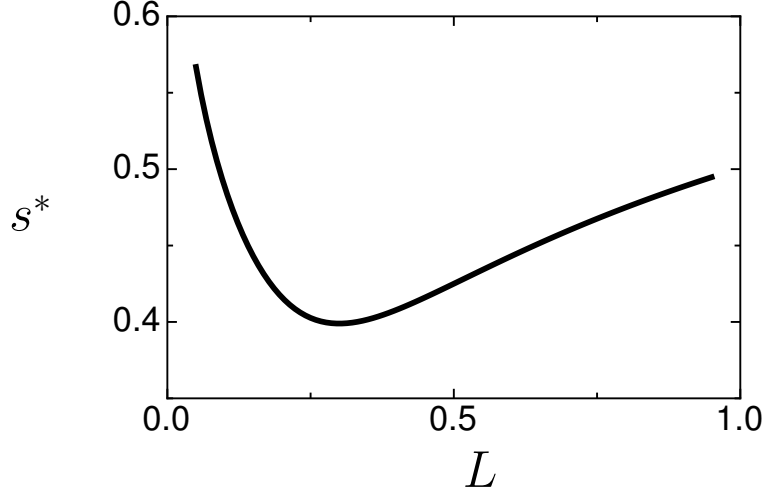


Figure C: **Optimal shape factor s when ϕ is lognormal and $f(L)$ is sigmoidal.** $s^*(L)$ is given by eqn. (S12)

$Q(L, s^*) = f(L)$. Thus, $f_i > f(L) > f_{i+1}$ and again we do not get monotonic convergence.

Figure D shows the nonmonotonic convergence for both $L < 1/2$ and $L > 1/2$.

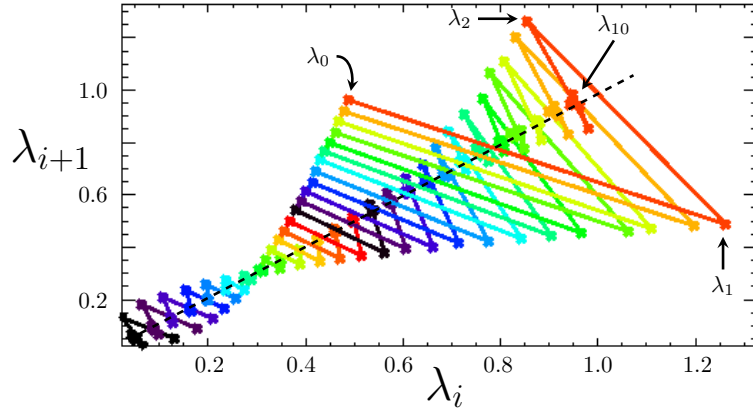


Figure D: **Nonmonotonic convergence of the Sloppy Algorithm when ϕ is lognormal and $f(L)$ is sigmoidal.** Plot is similar to Figure B. $f(L)$ is given by eqn. (S11) where $K = 0.3$. Although the Sloppy Algorithm converges, the convergence is nowhere monotonic.

Remarks: Using the same kinds of arguments we can show that for the cubic function $f(L) = L - (\gamma/2)L \cdot (L - 1/2) \cdot (L - 1)$ (for $\gamma \in [0, 4]$) and $\phi \in \mathcal{N}(0, s)$ the Sloppy Algorithm converges monotonically everywhere. However, when $\phi \in \mathcal{LN}(\ln(1/2), s)$ then the Sloppy Algorithm converges monotonically only for $L > 1/2$; convergence is oscillatory for $L < 1/2$.

2 Magic number for $\phi \in \mathcal{N}$ and $f(L) = L$

$s^*(L)$ solves the fixed point problem $Q(L, s^*) = L$. Writing $z = 2L - 1$ the fixed point problem becomes

$$\operatorname{erf}\left(\frac{\sqrt{2}z}{4s^*}\right) = z.$$

Because $L \in [0, 1]$ then $z \in [-1, 1]$. We use the approximation $\operatorname{erf}(x) \approx x$ for $x \in [-1, 1]$. Then the fixed point problem becomes

$$\frac{\sqrt{2}z}{4s^*} \approx \operatorname{erf}\left(\frac{\sqrt{2}z}{4s^*}\right) = z$$

from which it follows that

$$s^* \approx \frac{\sqrt{2}}{4} \equiv s_m.$$

Because this approximation holds over all $L \in [0, 1]$ it follows that s_m almost solves the fixed point problem for all L , that is, $Q(L, s_m) = L$.

3 Condition for having a useful magic number

When the benefit $B(L, \nu)$ is defined as (eqns. (9) and (11) in the main text)

$$\begin{aligned} B(L, \nu) \equiv \frac{\tilde{B}(L, \nu)}{D_m} &= [H(\nu)\nu - H(\nu - 1)(\nu - 1)] Lb_1 \\ &+ \frac{b_2}{\beta}(1 - L\nu) - L|1 - \nu|b_3 \end{aligned} \quad (\text{S15})$$

or

$$B(L, \nu) = \frac{f(L)}{L}\nu - \frac{\nu^2}{2}, \quad (\text{S16})$$

then $B(L, \nu)$ is maximized when $\nu = f(L)/L$. $s^*(L)$ solves the fixed point problem $Q(L, s^*(L)) = f(L)$ so B always is maximized when $s^*(L)$ is used. For arbitrary s , $Q(L, s) \neq f(L)$ so $\nu(s) = Q(L, s)$ will not maximize B . We'd like to replace the continuum $s^*(L)$ with a single magic number s_m that almost maximizes $B(L, \nu)$ for all L .

Clearly, the closer $Q(L, s_m)$ approximates $f(L)$ for *all* $L \in [0, 1]$, the closer $B(L, \nu)$ will be it to its maximum value. In other words, $Q(L, s_m)$ should “look

like” $f(L)$ in the sense that $Q(L, s_m)$ is close to $f(L)$ everywhere. The natural metric for this is the maximum norm,

$$d(Q(L, s_m), f(L)) = \max (|Q(L, s_m) - f(L)| \text{ for all } L \in [0, 1]) . \quad (\text{S17})$$

If $d(Q(L, s_m), f(L))$ is small then $B(L, \nu(s_m)) \approx B(L, \nu(s^*))$, meaning the benefit is nearly maximized.

The reason the normal distribution gave such poor performance when $f(L)$ was sigmoidal and the lognormal distribution gave excellent performance is because $Q_{\mathcal{N}}(L, s_m)$ (the integral in (S2) when $\phi \in \mathcal{N}$) does not look sigmoidal whereas $Q_{\mathcal{LN}}(L, s_m)$ ($\phi \in \mathcal{LN}$) looks remarkably sigmoidal.

4 Determining the grayscale level using the Sloppy Algorithm

Algorithms that work well in a computer simulation can fail miserably outside of a simulation. We tested whether the Sloppy Algorithm would work when a human was part of the iteration loop. The problem was to see if the Sloppy Algorithm could be used by a person to determine the absolute, as opposed to a relative, magnitude of a quantity. Examples of this kind of task is determining the brightness of a variable star by eye (<http://www.aavso.org/>) or the weight of an ox [1].

The specific problem task was to determine the gray scale value of an image. The image, a square displayed on the computer monitor, had a gray scale value between 0 and 255. Next to the test square, a comparison square of equal size whose gray level was randomly chosen from a normal distribution with a mean level of 128 and standard deviation s_i was shown. The person had to decide whether the test square was brighter or dimmer than the comparison square. After making N comparisons, $\bar{\lambda}_i$ was calculated from $\bar{\lambda}_i = n/N$, where n was the number of times that a test square was judged brighter than the comparison square and i is the iteration number. This is step [4] (eqn. (S4)) in the Sloppy Algorithm. Based on $\bar{\lambda}_i$ a new s_i that solved eqn. (S3) was determined. Fig. E shows results from three tests (from 2 subjects). The dashed lines mark the correct gray level L . The initial s value was set to $1 \times \Delta$ (filled circle) or $0.1 \times \Delta$ (filled square) where $\Delta = 256$ equals the range of possible gray values. These initial s values were chosen so that the first estimate $\bar{\lambda}_1$ would be far from L (the program “knew” the value of L but the person did not) thereby

allowing us to see how the estimates converged to L . The convergence to the correct grayscale value is similar to that seen in Figure 2 in the main text except that the convergence is nonmonotonic. Nonmonotonicity arises from the finite number of decisions (N) that were made; simulations show that the convergence becomes monotonic as $N \rightarrow \infty$. We used $N = 150$ to get good estimates of $\bar{\lambda}_i$ but making such a large number of decisions (150×6 iterations = 900 decisions) is tiring. Therefore, we tested whether setting s_0 to the magic number would hasten the convergence. The results (open triangle) in this case show that even on the first iteration the estimate (71.4) is already close to the correct value (70).

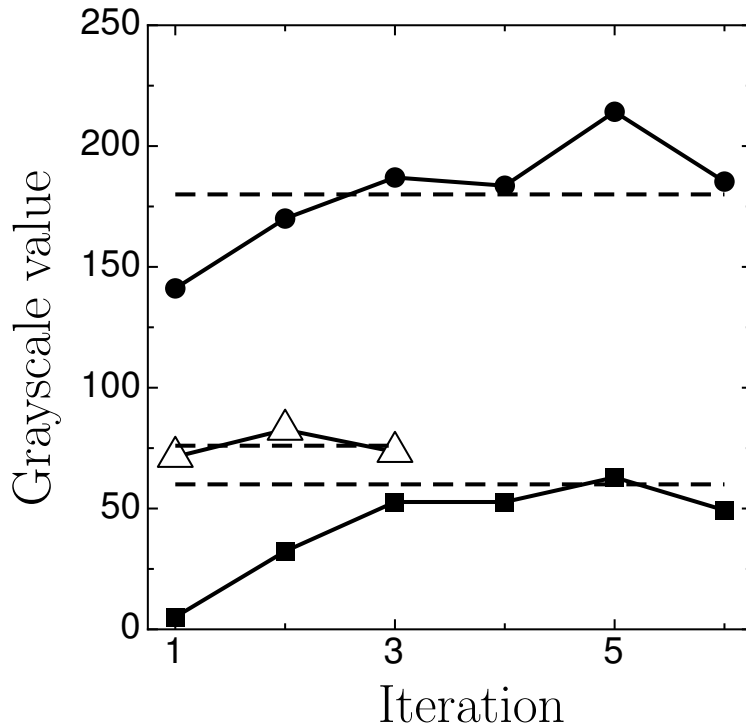


Figure E: **Evolution of grayscale estimation using Sloppy Algorithm.** Dashed lines indicate correct grayscale values. Subject 1, squares and triangles; subject 2, circles. Convergence is immediate when initial s equalled the magic number (triangles).

For this example one observer makes N decisions while in the main text each of the N rulers makes one decision. These two approaches are mathematically equivalent.

5 Sloppy Algorithm and sloppy rulers versus dithering

Sloppy rulers when combined with the Sloppy Algorithm can make accurate, high-resolution measurements even though each sloppy ruler has the lowest possible resolution. Dithering is a technique that can also improve measurement resolution [2] and has long been used to reduce quantization errors of analog-to-digital conversion [3]. Noise is essential in both dithering and sloppy rulers. However, sloppy rulers and dithering are different mathematically and in their arenas of application. In dithering the output signal is the average of both positive and negative excursions over many quantized states (256 states in an 8-bit analog-to-digital converter) centered around the input signal. By contrast, sloppy rulers average over only two states, zero and one.

This difference in what quantities are averaged is important in determining what the optimal noise level, s , should be to get accurate measurements. For dithering any s larger than half of the quantization step size will produce an accurate output [2]. For sloppy rulers, there is a unique s for each input value L that produces an accurate output, which the Sloppy Algorithm finds. Choosing s arbitrarily produces estimates of L shown in Fig. 3A in the main text. The y -axis is the estimate of L ; only by happenstance does the estimate match the true value of L .

The Sloppy Algorithm and dithering are useful in different systems. Sloppy rulers represent a wide class of systems that make binary decisions. Such systems include yes-or-no voting in politics, all-or-none protein expression in cells, choice of crops to plant. Dithering, on the other hand, is useful when there are many signal levels as in analog-to-digital converters and in smoothing out pixelation in images.

References

- [1] Galton F. Vox Populi. *Nature*. 1907;75:450–451.
- [2] Etchenique R, Aliaga J. Resolution enhancement by dithering. *Am J Physics*. 2004;72:159–163.
- [3] Lipshitz SP, Wannamaker RA, Vanderkooy J. Quantization and dither: A theoretical survey. *J Audio Eng Soc*. 1992;40:355–375.