# **Gradualism, Natural Selection, and the Randomness of Mutation –**

# **Fisher, Kimura, and Orr, Connecting the Dots**

# **Matthew J. Maxwell and Elliott Sober**

**Abstract**: Evolutionary gradualism, the randomness of mutations, and the hypothesis that natural selection exerts a pervasive and substantial influence on evolutionary outcomes are pair-wise logically independent. Can the claims about selection and mutation be used to formulate an argument for gradualism? In his *Genetical Theory of Natural Selection*, R.A. Fisher made an important start at this project in his famous “geometric argument” by showing that a random mutation that has a smaller effect on two or more phenotypes will have a higher average fitness than a random mutation that has a larger phenotypic effect. Motoo Kimura’s demonstration that a gene’s probability of fixation depends on both the selection coefficient and the effective population size shows that Fisher’s argument for gradualism was mistaken. Here we analyze Fisher’s argument and explain how Kimura’s theory leads to a conclusion that Fisher did not anticipate. We identify a fallacy that reasoning about fitness differences and their evolutionary consequences should avoid. We then distinguish forward-directed from backward-directed versions of gradualism. The backward-directed thesis about a single mutation may be correct, but the forward-directed thesis is not. After that we consider a sequence of random mutations that all affect the same cluster of phenotypes and Allen Orr’s idea that there is an optimal sequence of mutation sizes, moving from larger to smaller as the population approaches a fixed optimum. This provides a likelihood justification for a forward-directed version of gradualism when there is a fixed optimum, but it also provides an argument against forward-directed gradualism if the optimum shifts substantially as the population evolves. Finally, we consider whether genome-wide association studies (GWASs) can furnish empirical evidence that bears on the truth of gradualism. The version of gradualism that GWASs directly bear on concerns the phenotypic effect size of a gene *after* it arises by mutation and has reached an appreciable frequency, not the phenotypic size of the mutation event itself.

**key words:** Fisher,gradualism, Kimura, mutation, Orr, natural selection

The following three theses about gradualism, natural selection, and mutation are pair-wise logically independent:

* Gradualism: when a population evolves from one state to another, it does so by accumulating a large number of small changes, not by accumulating a small number of large changes.[[1]](#footnote-1)
* Natural selection is “the main but not the exclusive cause” of evolution.
* Mutations are random.

Darwin believed all three, but he didn’t get very far at showing how natural selection acting on random mutations would lead gradualism to be true. He does come somewhat close to drawing a connection of this sort in *Variation of Animals and Plants Under Domestication* when he writes “as conspicuous deviations of structure occur rarely, the improvement of each breed is generally the result ... of the selection of slight individual differences” (Darwin 1868, pp. 234-235). This was a comment about artificial selection. In another passage, he says that “slight individual differences … suffice for the work [in both artificial and natural selection], and are probably the sole differences which are effective in the production of new species” (p. 192). Darwin does not explain why he affirms the sufficiency of small differences for both artificial and natural selection but asserts its necessity only for the case of natural selection in the production of new species.

In Section 1, we explain R.A. Fisher’s (1930) “geometric argument,” which shows that a random mutation is more apt to be advantageous if it is smaller than it would be if it were bigger. Fisher and others took this to be an argument for gradualism, where Darwin’s always/never formulation is replaced by “almost always” and “rarely.” In Section 2, we discuss Kimura’s (1983) and Orr and Coyne’s (1992) criticism of Fisher’s argument. We think these authors are right that Fisher’s argument fails, but not for the reason they present. We then describe how Kimura’s formula for calculating the probability of fixation of a gene, given its selection coefficient and the effective population size, is relevant to the problem that Fisher addressed. The surprising result is that a smaller mutation often has a lower average probability of fixation than a bigger mutation possesses. In Section 3, we distinguish forward-directed from backward-directed versions of gradualism. When applied to a single mutation, the former is false, but the latter may be true. In Section 4, we explain Allen Orr’s (1998) idea that there is an optimal sequence of mutation sizes when a cluster of phenotypes is affected by the same set of genes, and describe how this idea bears on gradualism. In Section 5, we consider how genome-wide association studies (GWASs) may be relevant to deciding whether gradualism is true. In Section 6, we conclude.

**1 Fisher’s geometric argument about mutation**

R.A. Fisher (1930, pp. 41-44) saw a connection among the three theses, which he expressed in his famous “geometric argument.” A mutation may affect 0, 1, 2, …, or n phenotypic traits. If it affects zero, then it makes no difference to the survival or reproductive success of the organism in which it occurs, in which case (individual) selection will not affect whether it increases in frequency or exits from the scene. Matters change when a mutation influences a single quantitative phenotype (for example, the organism’s blood pressure). The mutation has a fixed size and will be random in terms of whether it moves the organism from its present state to a point that is closer to the optimum or to a point that is further away; each possibility has probability ½. The expected change in phenotype (and in fitness) of a small mutation is the same as the expected change in phenotype (and in fitness) of a big mutation; both are equal to zero.



Figure 1 represents a big and a small mutation that each affect two quantitative phenotypic traits (for example, blood pressure and blood pH) where possible pairs of values are represented in the x-y plane. Both mutations are random, meaning that each will change the organism’s phenotypic state from point X to some point on the relevant circle, and the different possible phenotypic upshots of a mutation have the same probability density of occurring. Our labels “big” and “small” are shorthand for a comparative fact: one of the mutations is *bigger* than the other in terms of how much each changes the organism’s phenotype. Figure 1 includes a third circle, which is centered on the optimum and has the organism’s present location at X on that circle’s circumference. Notice that the probability of the small mutation’s improving fitness is greater than the probability that the big mutation will do so. The first probability is almost ½ while the second is about 1/3. In both cases, the probability equals the length of a mutation’s circumference that improves fitness divided by its total circumference. The formula for this probability is:

A random mutation’s probability of improving fitness =

Here r is the size of the mutation and d is the distance to the optimum.[[2]](#footnote-2) The result bears out Fisher’s (1930, p. 42) claim that when r=0, the probability is ½ and as r increases, the probability decreases, reaching 0 at r =2d. For r > 2d, the mutation must increase the organism’s distance from the optimum. For all r>0, a mutation is more apt to reduce fitness than to increase it. In this sense, the effect of mutation on *fitness* is not *random*; it is *biased*. What is truly unbiased in Fisher’s model is the effect of a mutation on *quantitative phenotypes*. It is well to note that the symmetry assumptions in Fisher’s conception of random mutations are idealizations.

Fisher’s argument entails that

1. Pr(organism O’s fitness improves | O has a small mutation) >

Pr(organism O’s fitness improves | O has a big mutation).

However, proposition (1) does not entail that

E(Mutation M’s degree of fitness improvement | M is small & M improves fitness) >

E(Mutation M’s degree of fitness improvement | M is big & M improves fitness).

Here “E(X)” denotes the expected value of X. In fact, this inequality is false, which you can see by examining Figure 1. A proof is provided in Appendix 2. What is true is the following:

1. E(Mutation M’s degree of fitness improvement | M is Small) >

E(Mutation M’s degree of fitness improvement | M is Big).

A proof is provided in Appendix 3. Each of these expected values is *negative*, as shown in Figure 2.

Figure 1 illustrates Fisher’s argument when there are two phenotypes that are affected by a single mutation. What happens if the number of phenotypes is increased? On average, small mutations outperform large ones, but the gap between the two becomes more pronounced (Fisher 1930, Hartl and Taubes 1996, Orr 2000). For any n-dimensional space, the probability of an adaptive mutation is ½ when r = 0 and goes to 0 when r > 2d. Between these limits, however, the curves vary, as shown in Figure 3. As n increases, the graphs become more convex, indicating a more precipitous decline in the probability of adaptive mutation. Figure 3 leaves it open that different loci in an organism may differ in how many phenotypes they influence.

How does the earlier comparison of big and small mutations fit into Figure 3? Big has a larger value of r/d than Small does when they occur in the same population, so Big is always downhill from Small on whichever line you focus on in the figure. Inequalities (1) and (2) are true everywhere. However, when r/d is small, Big is much more downhill than Small the larger the value of n. Note also that the large advantage that Small has in large n populations attenuates rapidly as the organism gets closer to the optimum.

Orr (2000) associates “organismic complexity” with higher values of n. In complex organisms, big and small mutations are less apt to be advantageous than the big and small mutations in simpler organisms. Orr calls this the “evolutionary cost of complexity.” He concludes that it is harder for a species to adapt the more complex it is. If lineages evolve from simpler to more complex (or in the opposite direction), their location in Figure 3 will shift from one line to another. Orr’s point will not apply if a different measure of organismic complexity is adopted.

Fisher (1930, p. 44) concludes his presentation of the geometric argument by drawing an analogy:

The conformity of these statistical requirements with common experience will be perceived by comparison with the mechanical adaptation of an instrument, such as the microscope, when adjusted for distinct vision.  If we imagine a derangement of the system by moving a little each of the lenses, either longitudinally or transversely, or by twisting through an angle, by altering the refractive index and transparency of the different components, or the curvature, or the polish of the interfaces, it is sufficiently obvious that any large derangement will have a very small probability of improving the adjustment, while in the case of alterations much less than the smallest of those intentionally effected by the maker or the operator, the chance of improvement should be almost exactly one half.

The analogy is not perfect, since the derangement that Fisher describes concerns just one feature of a complex instrument, and Fisher’s asymmetry between bigger and smaller mutations applies only if derangements affect two or more features. What the argument does apply to is a derangement that is not just larger than a small one, but is large enough to guarantee overshooting the optimum. This fine point aside, the microscope analogy is something that Darwin could have embraced. In his *Autobiography* Darwin (1959, p. 59) says that he read William Paley’s 1802 *Natural Theology* while a student at Cambridge and was “charmed and convinced” by it. True, his theory of natural selection led him to reject Paley’s design argument, but Paley’s watch analogy could have served Darwin well in defending evolutionary gradualism. Perhaps Darwin hesitated to do so because he thought the watch analogy was poisoned by its association with intelligent design, so using it to defend his own theory would have compromised his one long argument.

**2 Probabilities of fixation of mutations that have different sizes**

Fisher seems to have regarded his geometric argument as providing a justification of gradualism. He says that “such large mutations occurring in the natural state would be unfavourable to survival, and as soon as the numbers affected attain a certain small proportion in the whole population, an equilibrium must be established in which the rate of elimination is equal to the rate of mutation" (Fisher 1930, p. 41). Fisher’s point is that big mutations are almost always driven to extinction, but that the population continues to have big mutations present at low frequencies owing to continuing mutational input. Clarke (1971) also interprets the geometrical argument as an argument for gradualism when he writes that Fisher “demonstrated that selection is very unlikely to favour extreme changes.” Orr (2009) similarly takes Fisher to be providing what eventually became a standard defense of gradualism.

Kimura (1983 p. 150) denies that Fisher’s argument succeeds as a defense of gradualism: “It seems to me that Clarke overlooked the point that mutations, if their effects are very small, are likely to have very small selective advantages and therefore have correspondingly small probabilities of fixation.” Orr and Coyne (1992, p. 728) say that “the important point is that a mutation's probability of fixation is directly proportional to its phenotypic effect. Thus, Kimura argues, even if large mutations are less likely to be favorable than small ones, they are nevertheless fixed more easily.”

Kimura and Orr and Coyne are right that there is something fishy about Fisher’s argument, but their comments don’t pinpoint where the argument goes wrong. Fisher’s argument correctly notes that the expected improvement from a big mutation is always less than the expected improvement from a small mutation. This *does* take into account the fact that when Big is advantageous, Big can get the organism closer to the optimum than Small does. Fisher did not ignore the fact that a big mutation can be advantageous, and then has a higher probability of evolving to fixation than a smaller advantageous mutation.

To see the flaw in Fisher’s argument,[[3]](#footnote-3) notice first that propositions (1) and (2) above are not about the probabilities of fixation that big and small mutations have once they occur. Rather, they describe how big and small mutations affect the phenotypes and fitnesses of organisms.

To figure out whether a bigger mutation B has a higher probability of going to fixation than a smaller mutation S, the relevant consideration is whether

(3) E[Pr(S goes to fixation | S occurs)] >

E[Pr(B goes to fixation | B occurs)]

The wrong thing to think about is whether

(4) Pr(S goes to fixation | E(fitness of S) = e1) >

Pr(B goes to fixation | E(fitness of B) = e2)

Here “E(-)” denotes expected value. To evaluate (3), you need to take account of each of the phenotypes that might arise from a mutation of given size, find its fitness, then compute its probability of going to fixation, and then find the average of those probabilities. To evaluate (4), your task is simpler; you consider the single average fitness for a mutation of given size, and compute the chance of fixation that a mutation with that average fitness value has. Inequalities (3) and (4) differ in their orders of operation. It is a fallacy to think that (3) follows from (4). What is more, (4) is true but irrelevant to the question of gradualism. And there is a surprise: (3) is often false for big and small mutations when the bigger mutation does not overshoot the optimum (proof in appendix 5).

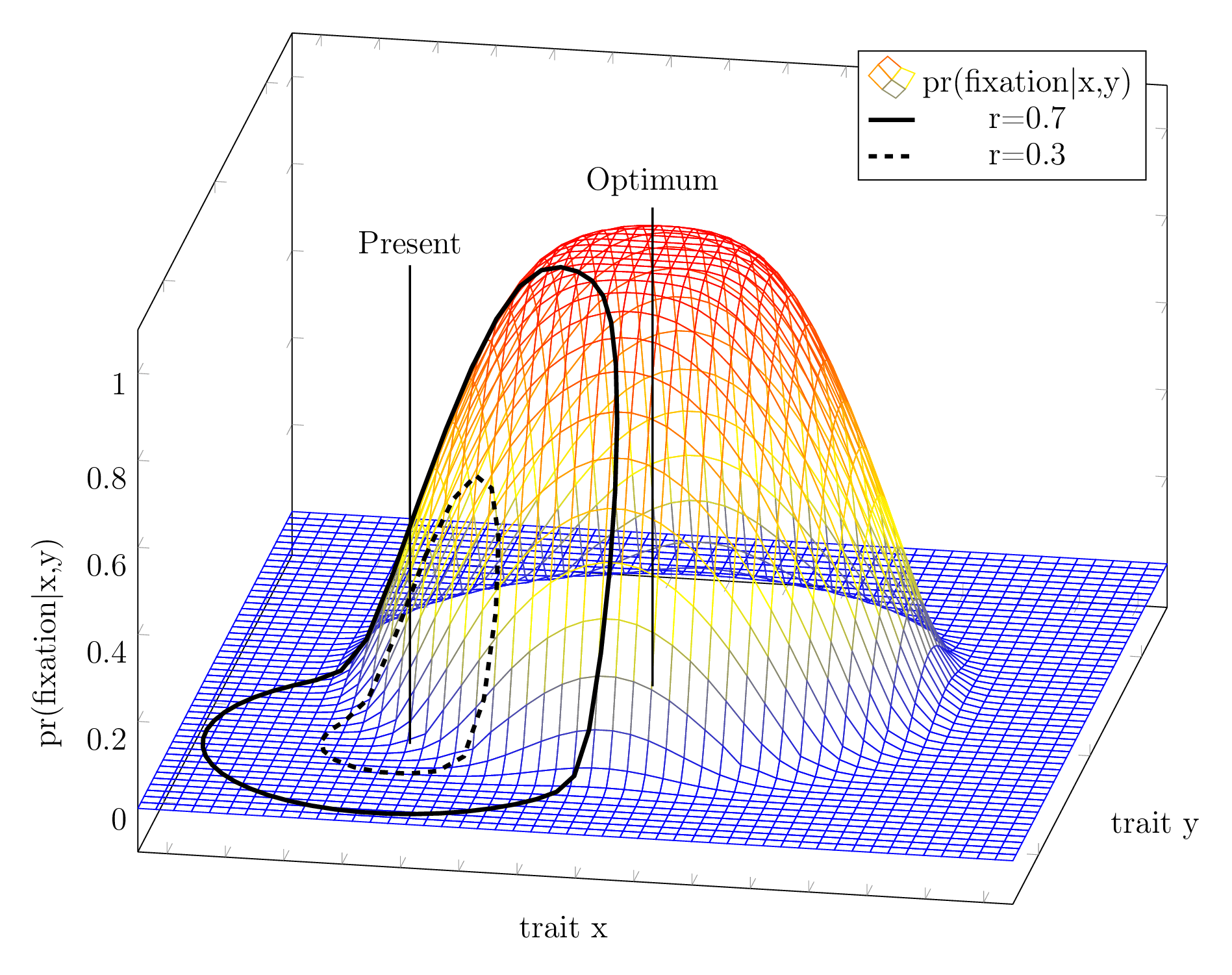
To think about the probability of fixation, given the different phenotypes that a mutation of given size might produce, you need to use a formula due to Kimura (1957, 1962) that says that a mutation M’s probability of fixation is where Ne is the effective population size and s is mutation M’s selection coefficient when it is introduced into a population that has wild type gene W at 100%. The selection coefficient is defined as follows:

s(M) = fitness(M) – fitness(W).

Figure 4 describes the probability of fixation for different values of the selection coefficient in a population of size 10 and as the population approaches infinity, but there are two properties of the curves shown that apply to all finite values of N: the probability of fixation of a mutation increases as its s value goes up, and the curve is convex when s is negative and concave when s is positive.

To apply Kimura’s formula to a big and a small mutation, you need to consider the range of s values a big mutation might have (one for each of its possible phenotypic upshots) and do the same for a small mutation. The range of s values for a small mutation is centered on a negative number; the range of s values for a big mutation is centered on an even more negative number. When s=0 the probability of a mutation’s going to fixation is 1/(2N) in a diploid population. That’s very small, when, for example, the population is of size N=100. This means that when s values get more and more negative, the probabilities of fixation shrink from 1/200 to something smaller. Thus, a big mutation risks producing more extreme negative s values than a small mutation, but these negative s values have a low “cost.” On the other hand, a big mutation can produce more extreme positive values than a small mutation, and the probability of fixation goes up rapidly as s becomes positive, so the “benefits” that a big mutation might enjoy can be substantial. Small plays it safe while Big is a risk-taker. Big has a larger variance in s values than Small does.

Figure 5: Probability of Fixation for Two Sizes of Mutation (N=10)



To construct Figure 5, we took Figure 1 (the one with the smiley face) and added a third dimension, which is the probability density of fixation given a mutation’s upshot for the two phenotypes, where the pair of phenotypes generated by a mutation determines its s value. Here we’re assuming as Fisher did that the fitness of a phenotype pair is its distance from the optimum. Kimura did not challenge this assumption. You can see from Figure 5 that the small mutation and the big mutation have pretty much the same low probabilities of fixation when they move the organism away from the optimum. On the other hand, when they move the organism towards the optimum, the big mutation picks up larger probabilities of fixation than the small mutation does. The upshot is that the big mutation shown in the figure has a higher average probability of fixation than the small one.

This raises the question of what the “optimum” size is for a mutation in terms of its probability of reaching fixation. If you assume, as Fisher did, that fitness is a linear function of distance to the optimum, then the answer, for two dimensions, is that the optimum size is 73% of the distance from the organism’s present state to the optimum (proof in appendix 5). That’s a BIG mutation!

Orr (2000) assumes a Gaussian relationship between phenotypic distance and fitness distance, and obtains a similar result. For two dimensions, his result is that the optimum mutation size is 65% of the distance to the optimum, which is slightly smaller than our value. He explains his intermediate mutation size, saying:

The reason is clear. Although smaller mutations enjoy greater probabilities of being favorable, they result in small increases in fitness upon substitution. Conversely, although larger mutations yield larger gains in fitness, they are less likely to be favorable. The optimal mutation walks a thin line at which these forces balance.

That is, gains in fitness from moving towards the optimum are greater than losses from moving the same distance away. However, it follows directly from the assumption that the relationship between phenotypic distance and fitness distance is Gaussian that gains in fitness arising from movement towards the optimum are greater than losses in fitness that come from moving away (for any but the closest of starting locations).  That is, Orr’s result is almost completely explained by the Gaussian assumption. What we have shown is that by dropping the Gaussian assumption and using Fisher’s linearity assumption instead, it’s still false that smaller mutations are more apt to evolve to fixation than larger ones.  The Gaussian assumption is not necessary (though it may be realistic) and the issue is not how to counterbalance the fact that smaller mutations have a higher probability of being advantageous.  Orr misses the greater part of the explanation; small gains to fitness result in large increases to probability of fixation while small losses to fitness result in little decrease. The non-linearity of the fixation curve shown in Figure 4 is key.

It might be objected that our use of Kimura’s formula is “unfair” to Fisher, either because Fisher was interested in adaptive evolution in very large populations or because Kimura’s model rejects assumptions that Fisher made. Our reply to the first objection is that the focus on extremely large populations does not save Fisher’s argument.  We noted above that for all finite values of N, the fixation curve is concave when s is positive and convex when s is negative.  The probability of fixation of an advantageous mutation does not go to 1 as N goes to infinity; this is because Kimura’s formula models the introduction of a single (token) mutant allele into a population.  For any finite selection coefficient, there's a non-zero chance that the allele fails to be passed to the second generation.  Given that, a mutation’s probability of eventually going to fixation can’t equal 1, no matter how big the population is.

With respect to the second objection, Stolzfus (2017; see his “Supporting Quotations”) argues that Fisher took evolution to be the shifting of gene frequencies that are already present in a population, not the introduction of new alleles followed by their fixation.  Our reply is that this distinction, taken by itself, is not enough to vindicate Fisher’s argument.  Rather, we must additionally assume that the population is infinite and all alleles in the population have frequencies bounded away from zero. With those assumptions, it follows that all advantageous alleles increase in frequency while all deleterious alleles decline. Fisher’s geometric argument is then a sidebar, merely noting that a small mutation has a higher probability of creating advantageous alleles than a big mutation. The way big mutations can pick up larger positive s values than small mutation are able to do (as shown in Figure 5) is negated in this model. In addition, there is no account in this Fisherian picture of how mutation events might be able to create new alleles that have frequencies bounded away from zero in an infinite population.

As it happens, Chapter IV of Fisher 1930 *does* treat individual (token) mutations and their probability of spreading through a finite population.  This chapter anticipates Kimura’s formula by relating the relative fitness of an allele to the probability of extinction of that allele in succeeding generations, as represented by a Poisson series; *a fortiori*, Fisher recognized that there is a non-zero chance that a newly introduced allele is not passed on to the second generation.  While Fisher may have believed he could ignore mutations in his geometric argument, Kimura’s assumptions were not alien to him.  It is surprising that Fisher did not deploy this idea in order to discover the relationship between mutation size and probability of fixation in finite populations, no matter how large they are.

**3 Forward and backward formulations of gradualism**

Here are two formulations of gradualism:

(F1) Pr(gene G is now at fixation | gene G arose earlier by a small mutational change) >

Pr(gene G is now at fixation | gene G arose earlier by a big mutational change)

(B1) Pr(gene G arose earlier by a small mutational change | gene G is now at fixation) >

Pr(gene G arose earlier by a big mutational change | gene G is now at fixation)

We call these “F1” and “B1” because they both involve a single mutation. Notice that F1 involves forward-directed probabilities while the probabilities in B1 are backward-directed. We have shown that F1 is often false when mutations are random, but that doesn’t mean that B1 is false as a claim about random mutations.

Here we pause to reflect on a familiar rule of thumb ─ that when evolution is “controlled” by natural selection, fitter traits increase in frequency and less fit traits decline. This rule may seem to be called into question by the falsehood of F1, coupled with the truth of proposition (2) – that E(Mutation M’s degree of fitness improvement | M is Small) > E(Mutation M’s degree of fitness improvement | M is Big). To see why the rule is not impugned by these two findings, consider the fact that Big and Small are not traits that increase in frequency under natural selection in the models considered here. When a mutation of any size occurs, what is transmitted to the next generation is the changed *state* of the gene (and its associated phenotypes) not the *size of the mutation* that gave rise to that state. The ideas of Fisher and Kimura that we are discussing do not concern the evolution of genes that control the size of mutations elsewhere in the genome. It is a fallacy to think that Fisher’s results and the rule of thumb show that forwards gradualism is true.

Although propositions F1 and B1 are logically independent of each other, they are joined at the hip by the odds formulation of Bayes’s theorem, which applies to the example at hand as follows:

(Odds) =

x

Odds has the form: (backward ratio) = (forward ratio) x (ratio of priors). We know that the forward ratio is often less than 1. If we knew what the ratio of priors is, we could then see whether the backward ratio is greater than 1. However, it’s unclear how one could estimate the prior ratio.

**4 A sequence of mutations that affect the same set of phenotypes**

Both Fisher and Kimura considered whether a bigger mutation has a higher probability of going to fixation than a smaller one on the assumption that mutations are random in their phenotypic effects. Orr (1998) took the important step of thinking about a series of mutations, each affecting the same cluster of phenotypes. Assuming that the phenotypic optimum suddenly shifts to a new location and then stays put, Orr derived an exponential probability distribution of optimal mutation sizes. Here “optimal” means the sequence of mutation sizes that has the highest probability of going to fixation.

To illustrate Orr’s idea, suppose two phenotypes are now at fixation, where the two are caused by the presence of n genes. Those genes arose by mutations that then went to fixation. Suppose these mutation/fixation events are spaced out temporally, so that one gene arises by mutation and then goes to fixation, after which another does the same. Using Fisher’s linearity assumption, which entails that the optimal mutation size is 73% of the distance d to the optimum, the optimal sequence of mutation events of different phenotypic sizes will be:

(S) The first mutation is of size (73%)d, the second is of size 73%(27%)d, the third is of size

73%(27%)2d, …, and the nth is of size 73%(27%)n-1d.

If n is large, the result is a small number of very large mutations followed by a large number of very small ones.

It remains to relate Orr’s idea to the contrast between forward and backward probabilities. Here are the two options:

(Fn) For each S\* ≠ S,

Pr(phenotypes P1…Pm are now at fixation and each is affected by the same n genes | those n genes originated by n mutation events that followed S) >

Pr(phenotypes P1…Pm are now at fixation and each is affected by the same n genes | those n genes originated by n mutation events that followed S\*)

(Bn) For each S\*≠ S,

Pr(those n genes originated by n mutation events that followed S | phenotypes P1…Pm are now at fixation and each is affected by the same n genes) >

Pr(those n genes originated by n mutation events that followed S\* | phenotypes P1…Pm are now at fixation and each is affected by the same n genes)

The forward-directed thesis Fn is true. Fn does not entail Bn. To evaluate whether Bn is true requires information about prior probabilities. Note also that Fn also does not entail F1 when n=1. That’s because F1 affirms gradualism with respect to a single mutation, but Fn does not.

Fn does not say that the genes that now encode a phenotype *probably* arose by a series of mutation events that conform to the S sequence. Rather, it says that the observation of a cluster of phenotypes that is encoded by the same n genes favors the hypothesis that those genes originated by an S sequence of mutation events over the hypothesis that it arose by any other specific sequence S\* of mutations. Here “favoring” is used in the sense of the Law of Likelihood (Hacking 1965, Edwards 1972, Royall 1997, Sober 2008, 2015).

What does Fn say about gradualism? Sequence S is more gradualistic than many alternative sequences, but S is less gradualistic than many others. Indeed there are infinitely many alternative S\* sequences in both categories. Rather than trying to place Fn in one of two categories (extreme gradualism and extreme saltationism), we think it is better to simply recognize that sequence S is an optimal mixture of bigger and smaller. That optimum will include a few large mutations and a lot of small ones for large n, and so it is “closer” to extreme gradualism than it is to extreme saltationism.

The truth of Fn leaves open how often the assumptions that go into its derivation are satisfied in the real world. We note, in this connection, the result’s reliance on the assumption of pleiotropy. How often do single genes have multiple phenotypic effects? The answer depends on how one individuates genes and phenotypes.

Another assumption is that the optimum does not change as the population evolves. If the phenotypic optimum moves steadily away from the organism’s phenotypic location as the population evolves, the optimal sequence of mutation sizes will be a series of large phenotypic jumps provided that the moving target is moving fast enough. This is the situation envisaged by Van Valen’s (1973) red queen hypothesis.

**5 GWASs to the rescue?**

Darwin, Fisher, and Kimura were in no position to estimate the values of the probabilities deployed in the Backward or Forwards formulations of Gradualism, but perhaps contemporary biology is able to do so. Genome-wide association studies (GWASs) may provide useful guidance, since they estimate the (phenotypic) effect sizes of numerous genetic loci.[[4]](#footnote-4) Do these studies show that extant alleles with smaller phenotypic effects are more common than alleles with larger? These studies are carried out in connection with loci that are polymorphic, and so they fail to bear directly on formulations of gradualism that are only about genes at fixation. However, the gap can be bridged. The first step is to broaden our conception of gradualism, so that it makes a claim about all genes that have reached appreciable frequencies, not just the ones that have reached fixation. The second step is the assumption that if GWASs support gradualism with respect to genetic polymorphisms, then they also support gradualism concerning genes that are at fixation. The third step is the inference that a locus that now has small effect on a phenotype, also had small effect on the phenotype as the genes at that locus evolved. The first step seems unproblematic, but the second and third are open to question.

GWASs don’t study single *alleles*; rather, they study how much effect a *locus* has on a given phenotype by looking at the different genotypes that exist at that locus. Roughly, the question is how much variation there is in a given phenotype among individuals that have different genotypes at a given locus, while controlling, as much as possible, for variation at other loci. GWAS papers often say that effect sizes are usually “small” (see, for example, Simons *et al.* 2018 and Visscher *et al.* 2010). This entails that “small” effect size is usually more common than “large,” but this dichotomous classification is less helpful for our purpose than one in which effect size come in more than two discrete categories or is treated as a continuous variable.

Park *et al.* (2010) consider three phenotypes – height, Crohn’s disease, and BPC cancers. They note that GWASs cannot detect cases in which effect sizes are very small, as even a very large sample size will mean that some of these will not be statistically significant. There is no such bias at the other end, however, as big mutations are easily detected in GWASs. Park *et al.* provide a statistical technique for correcting for this bias, and draw the conclusion that for all three phenotypes, loci with smaller effects are more common than loci with larger. It is interesting that two of the phenotypes they study are diseases, whereas the third is not. Lopez-Cortegano *et al.* (2019) studied Crohn’s disease and inflammatory bowel disease and showed that a log-normal model has a very good AIC score. The log-normal curve that fits the data best (their Figure 1a) shows that the expected frequency of the alleles at a locus is a declining function of its effect size. They did not compare the log-normal model to other models, however. Loos *et al.* (2013) present data in which alleles that increase heart rate are more common the smaller their effect. Marouli *et al.* (2017) did a GWAS on human adult height and report that “variants with a larger effect size on height variation tend to be rarer.”

Boyle *et al.* (2017) go further and argue that many complex traits are “omnigenic” − that is, they are influenced by many, many genes with small effect sizes. Early GWAS studies found that significant variants had small effects, and moreover, much of the heritability of diseases and traits was unaccounted for by these variants (Manolio *et al*. 2009). Only 5% of the variance in height, for example, was accounted for by the 40 most significant loci, while heritability was estimated at 80%. Most of the heritable variance, therefore, must be accounted for by loci with small effect sizes. Boyle *et al.* 2017 estimate that perhaps 62% of all common SNPs are associated with a non-zero effect on height. Accounting for linkage disequilibrium, this means that 3.8% of SNPs causally affect height, suggesting that upwards of 100,000 loci have independent effects on height. Perhaps obviously, most of these have a very small effect (median effect size 0.14mm). Similar results hold for other complex traits, such as schizophrenia, Crohn’s disease, and rheumatoid arthritis.

How do present GWAS studies bear on the issue of gradualism? We suggest that the formulation to consider is this:

(F-GWAS) Pr(most of the phenotypes in GWASs are such that most loci now have small

average effects on those phenotypes | the phenotypes and genotypes that evolved in

the past were such that most loci had small average effects on the phenotypes) >

Pr(most of the phenotypes in GWASs are such that most loci now have small

average effects on those phenotypes | the phenotypes and genotypes that evolved in

the past were such that most loci had large average effects on the phenotypes).

The law of likelihood interprets this inequality to mean that GWAS data favor one hypothesis about the past over another. F-GWAS is a forward inequality, but it differs from Fn, which is also a forward inequality. Fn concerned Orr’s top/down theoretical argument about mutation sizes. F-GWAS isn’t about mutation size at all. It’s about the phenotypic effect sizes that genotypes had once they appeared by mutation. Once again, the conclusion is not that a gradualist hypothesis is *probable*.

If F-GWAS is true, can it be extended so that it addresses the question of mutation size? It can, if mutations that have a given phenotypic effect size usually give rise to genotypes that have the same phenotypic effect size. However, this is not inevitable. Suppose there are two alleles, A1 and A2, at a locus in a population and they simultaneously mutate to B1 and B2, respectively, and those new B alleles then go to fixation. This means that there were three genotypes before and three genotypes after. Suppose the average heights of individuals with these six genotypes were as follows:

H(A1A1) = 3.0 cm, H(A1A2) = 3.01 cm, H(A2A2) = 3.02 cm

H(B1B1) = 6.0 cm, H(B1B2) = 6.01 cm, H(B2B2) = 6.02 cm

The mutations are big, but the average effect sizes at the locus, before and after, are small.

In the light of the apparent GWAS evidence in favor of a forward formulation of gradualism, there are two caveats. The first is the familiar point that *a phenotype is the joint product of genes and environment*.  If present and past environments are different, there is no assurance that a locus that had small average phenotypic effects in the past will probably have small average effects in the present. The second caveat is that the GWASs conducted to date may represent a biased sample of phenotypes.

**6 Conclusion**

Fisher’s geometric argument seems to show that natural selection acting on a random mutation that influences two or more phenotypes suffices to make gradualism true. We have shown that there are two formulations of gradualism ─ one forward-directed, the other backward-directed. The forward-directed thesis is false and it does not follow from Fisher’s geometric argument. The backward-directed thesis may be true, but Fisher’s argument does not show that it is. Fisher’s argument, in the first instance, shows that a small mutation has a higher probability of improving an organism’s fitness than a bigger mutation possesses. The argument does not address the issue of how the average probability of fixation of a smaller mutation compares with the average probability of fixation of a bigger mutation. What is needed here is Kimura’s formula, but that has a surprising consequence: bigger mutations often have higher average probabilities of reaching fixation than smaller mutations have.

The situation changes dramatically when a phenotype is caused by more than one gene. Using Orr’s idea of an optimal sequence of mutation sizes, and assuming that the optimal phenotype does not change as the population evolves, we claimed that the observation of two or more phenotypes that are caused by a single large cluster of genes favors the hypothesis that those genes had mutation sizes that conform to exponential distribution S over the hypothesis that those genes had mutations sizes that have any other distribution. This is a likelihood justification of gradualism. However, this argument for gradualism transmutes to an argument against if the optimal phenotype speeds away from the population as the population evolves.

It is interesting how the contrast between forward and backward gradualism is affected by the shift from considering the mutation size of a single gene that causes multiple phenotypes to considering the mutation sizes of several genes that cause that same multiplicity. Forwards gradualism as a claim about a single gene is false, given Fisher’s framework, but the hypothesis of forwards gradualism\* is favored by the observation that numerous genes affecting two or more phenotypes are at fixation, given Orr’s framework. We put an asterisk in this last sentence to note that the sequence of mutation sizes we have in mind is not 100% small. This interpretation of Orr’s idea involves forward probabilities.

We then considered how data from GWASs bear on gradualism. This bottom/up approach has no commitment to the top/down theories of Fisher, Kimura, or Orr, and so the assumptions of near-universal pleiotropy and a stationary optimum used in the latter aren’t needed to assess the former. The present picture coming from GWAS data seems to favor a forward formulation of gradualism over a forward formulation of saltationism, but this result is not about mutation sizes; it concerns the phenotypic effect sizes of genes *after* they arise by mutations. Again the conclusion is not that a version of gradualism is *probably* true; rather, the claim involves a *likelihood* comparison.

This paper has examined several versions of gradualism. Gradualism can be a claim about the size of a single mutation that goes to fixation, or about the sizes of several mutations that go to fixation sequentially, or it can be about the phenotypic effect size that genes have after they arise by mutation and achieve an appreciable frequency, in which case it isn’t about mutation size at all. Cutting across this three-fold distinction is the distinction between forward and backward probabilities. Finally, we noted a distinction between two styles of argument for gradualism; top/down arguments seek to provide theoretical reasons to expect a version of gradualism to be true; bottom/up arguments seek to defend a version of gradualism by appeal to observations. These and other possibilities merit consideration.

**Acknowledgments:** We are grateful to Alan Attie, David Baum, Brianna Bixler, Hayley Clatterbuck, Joe Felsenstein, Maël Lemoine, Bret Payseur, Maxim Shpak, Mike Steel, Benedek Valko, and Donald Waller for useful discussion. We also thank the anonymous referee of this journal for helpful suggestions.

**References**

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell*, 169(7), 1177-1186.

Clarke, B. (1971). “Natural selection and the evolution of proteins.” *Nature* 232: 487.

Darwin, C. (1859). *On the Origin of Species by Means of Natural Selection.* London: Murray.

Darwin, C. (1868). *The Variation of Animals and Plants Under Domestication*. London: Murray, vol. 2.

Darwin, C. (1958). *Autobiography*, N. Barlow (ed.), London: Collins.

Edwards, A. (1972). *Likelihood*. Cambridge University Press.

Eldredge, N. & Gould, S. J. (1972). Punctuated equilibria: an alternative to phyletic gradualism. *Models in paleobiology*, *1972*, 82-115.

Fisher, R.A. (1930). *The Genetical Theory of Natural Selection.* Oxford University Press.

Hacking, I. (1965). *The Logic of Statistical Inference*. Cambridge University Press.

Hartl, D. L., & Taubes, C. H. (1996). “Compensatory nearly neutral mutations: selection without adaptation.” *Journal of Theoretical Biology*, *182*(3): 303-309.

Kimura, M. (1957). “Some Problems of Stochastic Processes in Genetics.” *Annals of Mathematical Statistics* 28: 882-901.

Kimura, M. (1962).  “On the Probability of Fixation of Mutant Genes in a Population.” *Genetics* 47: 713–719.

|  |
| --- |
|  |
| Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press. |

Li, S., (2011). Concise Formulas for the Area and Volume of a Hyperspherical Cap. *Asian Journal of Mathematics and Statistics*, *4*(1): 66-70.

Loos RJF, et al. Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders. *Nat Genet.* 2013 Jun;45(6):621-31. doi: 10.1038/ng.2610. Epub 2013 Apr 14. PMID: 23583979; PMCID: PMC3696959.

López-Cortegano, E., & Caballero, A. (2019). GWEHS: A Genome-Wide Effect Sizes and Heritability Screener. *Genes*, *10*(8), 558.

Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., ... & Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265), 747-753.

Marouli, E., Graff, M., Medina-Gomez, C. *et al.* (2017). “Rare and low-frequency coding variants alter human adult height.” *Nature* 542:186–190.

Park, J. H., Wacholder, S., Gail, M. H., Peters, U., Jacobs, K. B., Chanock, S. J., & Chatterjee, N. (2010). Estimation of effect size distribution from genome-wide association studies and implications for future discoveries. *Nature genetics*, 42(7), 570-575.

Orr, H. A. (1998). The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, *52*(4), 935-949.

Orr, H. A. (2000). “Adaptation and the cost of complexity.” *Evolution*, *54*(1): 13-20.

Orr, H. A. (2009). “Darwin and Darwinism: the (alleged) social implications of *The Origin of Species*.” *Genetics*, *183*(3), 767-772.

Orr, H. A., & Coyne, J. A. (1992). The genetics of adaptation: a reassessment. *The American Naturalist*, *140*(5), 725-742.

Royall, R. (1997). *Statistical Evidence – a Likelihood Paradigm*. Chapman and Hall.

Simons, Y. B., Bullaughey, K., Hudson, R. R., & Sella, G. (2018). A population genetic interpretation of GWAS findings for human quantitative traits. *PLoS biology*, *16*(3), e2002985.

Sober, E. (2008). *Evidence and Evolution*. Cambridge University Press.

Sober, E. (2015). *Ockham’s Razors – A User’s Manual.* Cambridge University Press.

Stoltzfus, A. (2017). Why we don’t want another “Synthesis”. *Biology Direct*, *12*(1), 1-12.

Van Valen, L. (1973). “A New Evolutionary Law.” *Evolutionary* Theory 1: 1–30.

Visscher, P. M., McEvoy, B., & Yang, J. (2010). From Galton to GWAS: quantitative genetics of human height. *Genetics research*, *92*(5-6), 371-379.

**Appendix 1: the probability that a random mutation will improve fitness**

In Figure 1, let *r* be the radius of a circle centered on X and *d* be the radius of the circle centered on the optimum.[[5]](#footnote-5) We use the law of cosines to find *w,* the distance from the phenotype resulting from to the optimum:

As we want to know the probability that this mutation is advantageous () and given that all values are positive distances,

Rearranging and cancelling terms we find,

Taking the arccosine to find the angles,

As both the positive and negative angles are equal, the total angle will be . Dividing by the full circle, we get the proportion of the full circle which improves fitness,

Note that when , there will be no solutions to this formula. In this case the circle centered on X will fully encompass the circle centered at the optimum, and thus all mutations will decrease fitness.

**Appendix 2: the expected degree of improvement, given that the mutation improves fitness**

We begin again with the law of cosines, where the mutation is of size *r*, but this time we solve for *w,* the fitness distance resulting from a mutation .

The improvement (decrease in fitness distance), , from mutation will be equal to . Thus,

To find the expected improvement given that the mutation improves fitness we integrate over the resulting in advantage and average over all resulting in advantage. For the full circle we would integrate from/to but as both the positive and negative portions are equal and we are only concerned with proportion, we simply integrate from 0 to the positive angle and divide by rather than .

Figure 5 shows the ratio of the expected improvement in fitness for a big mutation of size (0.1)f and the expected improvement in fitness of a small mutation of size 0.1, where f>1. As can be seen, the ratio is not always greater than 1. The ratio begins to drop off when the big mutation is larger than d and begins to “overshoot” the optimum. Eventually, this overshooting becomes so deleterious that small is more advantageous than big.

**Appendix 3: a random mutation’s expected improvement in fitness**

To find the expected fitness from a random mutation, we take the average of all possible mutations of size *r*. This is done by first integrating over all possible (0 to ) and then dividing by the full circle ():

This is shown in Figure 6.

**Appendix 4: the full formula for (area of spherical cap)/(area of sphere) for an n-dimensional sphere**

We begin with the formula for a spherical cap and sphere (Li 2011). *n* is the number of dimensions, *r* is the radius, and I the regularized incomplete beta function.

We divide the formula for the area of the spherical cap by the area of the full n-sphere to find the proportion of the sphere covered by the cap:

The areas of the sphere, cancel, leaving us with:

Using trigonometric identities and the results from Appendix 1, we find, :

(Appendix 1)

Thus our final result for the probability of an advantageous mutation in n-dimensions is:

As in Appendix 1, this formula will not be valid for . At all will take the organism further from the optimum (and thus be disadvantageous). Hartl and Taubes (1996) arrive at the same result, but express it as a fraction of integrals:

**Appendix 5: Probability of Fixation for a Mutation of Size *r***

The selection coefficient, *s*, for a mutation *B* is given by:

where *WA* is the fitness of the wild type and *WB* the fitness of the mutant. From Appendix 2, let the fitness distance to optimum of the wild type be 1, and be the distance to optimum for the mutant. Halving the distance to optimum doubles the fitness, reducing distance by 2/3 triples it, and so on, thus:

Choosing the wild type as our reference, we let , thus:

Substituting these values into the definition of s:

Substituting for *w* by the formula from Appendix 2:

The probability of fixation for a randomly mating diploid population of size is given by Kimura (1962):

Substituting the above value for *s* into Kimura’s formula and using the ‘exp’ notation for clarity [],

To find the expected probability of fixation for a random mutation of size *r*, we integrate the probability densities from the above formula over all mutations and divide by all possible mutations (effectively summing the probability densities for all possible mutations and dividing by the number of possible mutations to find an average probability):

Letting , we plot this curve in Fig 7. This plot is valid for all . The greatest probability of fixation for occurs when . Thus a mutation smaller than or larger than will have a lower average probability of fixation. This value is remarkably stable, going to 0.7304 as . Thus smaller mutations do not have a higher average probability of fixation than bigger mutations.

1. This formulation will be refined later in the paper. Note that the gradualism discussed here is not about rates of evolution. Thus it is not in conflict with the punctuated equilibrium thesis defended by Eldredge and Gould (1972). [↑](#footnote-ref-1)
2. Hartl and Taubes (1996) and Orr (2000) mention this result but do not supply a proof; a proof can be found in Appendix 1. [↑](#footnote-ref-2)
3. Here we set aside the fact that Fisher assumes that that fitness is linearly related to the phenotypic distance from the organism’s present state to the optimum, and the fact that Fisher considers only a single optimum, rather than taking account of the possibility that there are multiple adaptive peaks (Orr 1998). [↑](#footnote-ref-3)
4. GWASs don’t allow you to measure the prior probabilities described in Odds. GWASs examine mutations that have managed to evolve to measurable frequencies, not the frequencies of mutations of various sizes, many of which may promptly go extinct. [↑](#footnote-ref-4)
5. Fisher (1930) uses d/2 as the diameter of the circle centered on the optimum, which simplifies the mathematics slightly, but at the cost of clarity. [↑](#footnote-ref-5)