6.891: Computational Evolutionary Biology

R.C. Berwick & a cast of thousands
Today: the forces of evolution, III

The forces of evolution, III

Does selection maximize fitness?
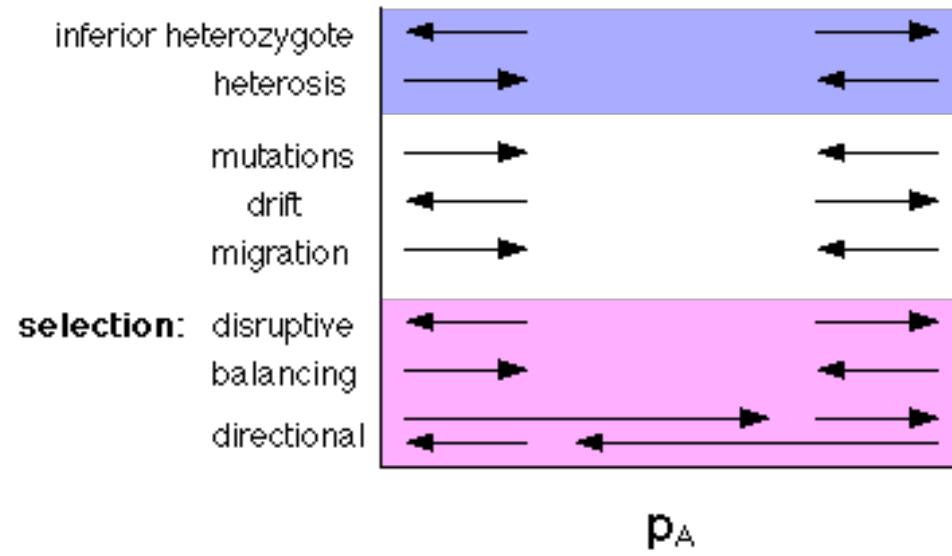Does sex make you fitter?
    Avoiding the ratchet
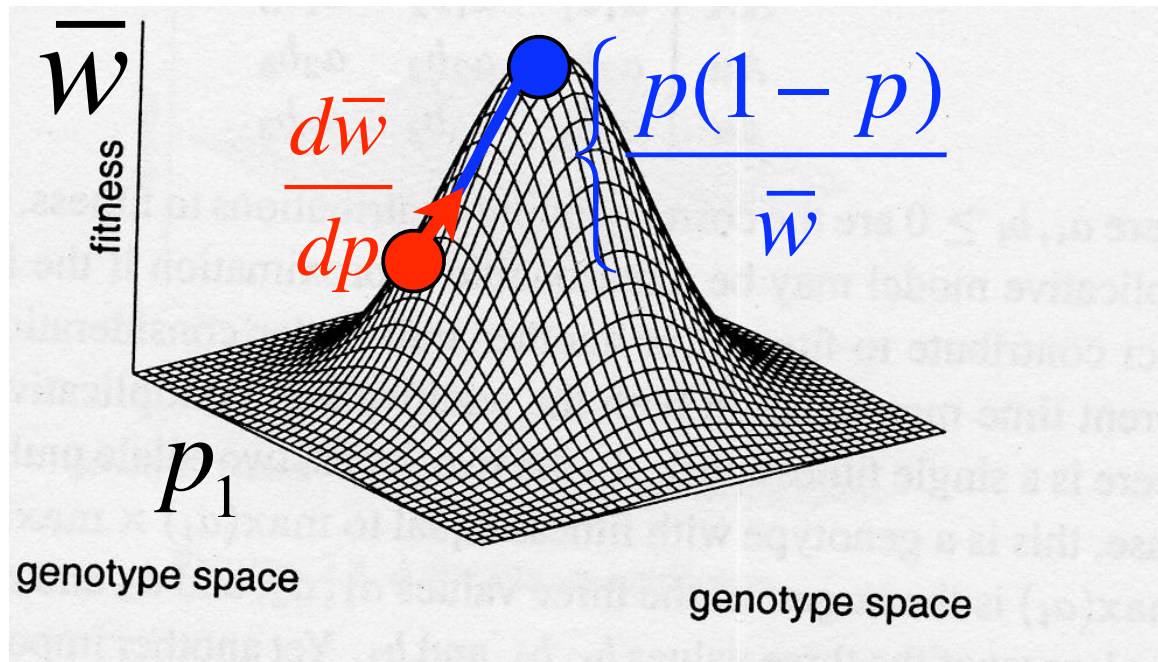The multivariate case: sickle cell anemia example
    & stability
The weak force: mutation
The superstrong force: migration
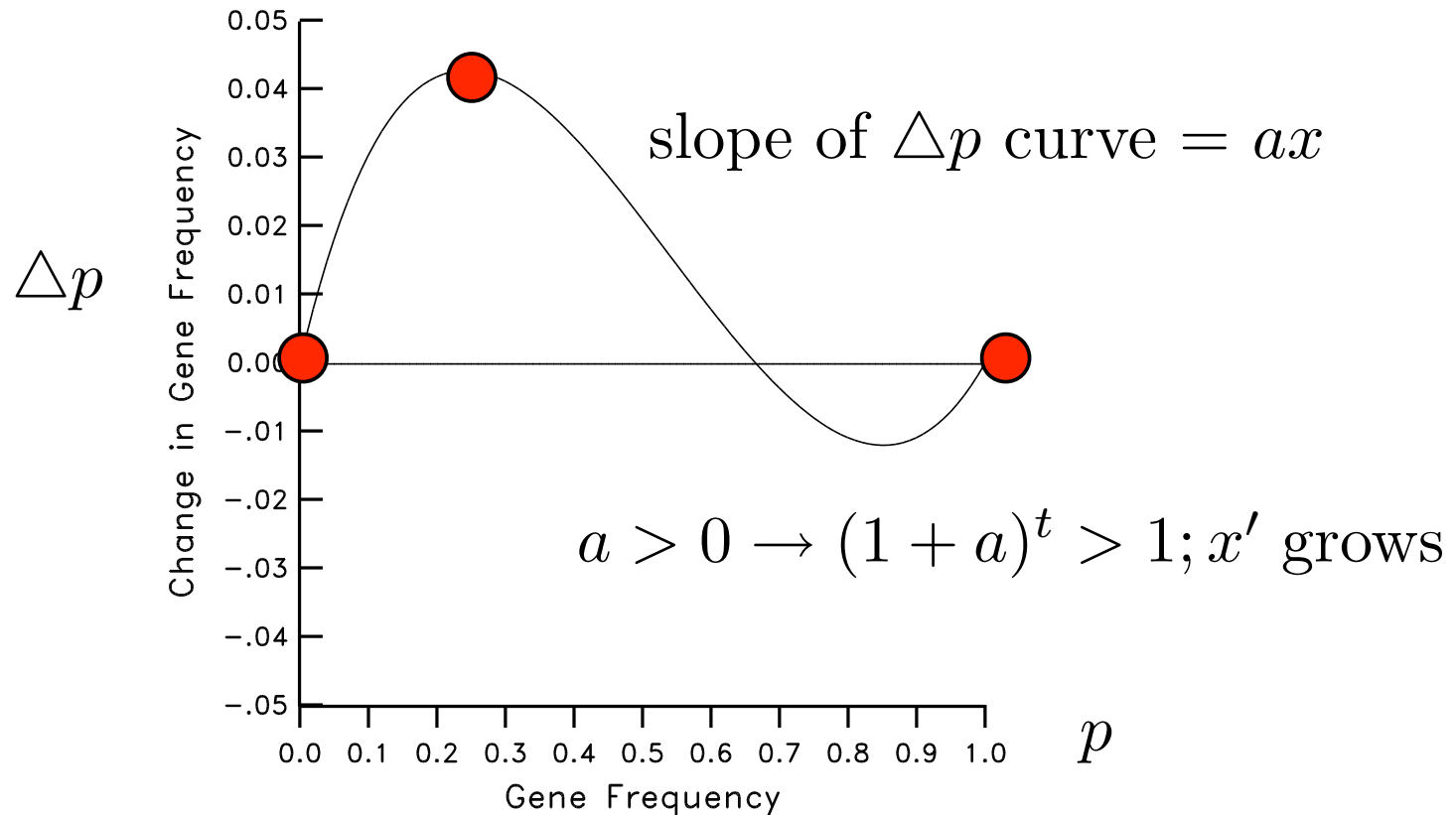The thermodynamic force: drift

Natural selection works by local gradient ascent

Stability Considerations

$$x' = p + \Delta p - p_e = p_e + x + \Delta p - p_e$$

$$= x + \Delta p \quad \simeq \quad x + ax \quad = \quad x(1+a)$$

slope of $\triangle p$ curve $= ax$

$\triangle p$

$a > 0 \rightarrow (1+a)^t > 1; x'$ grows

The change in gene frequency (Δp) plotted against the gene frequency in a case of overdominance where fitnesses of $AA : Aa : aa$ are 0.85 : 1 : 0.7.

Stability Analysis

$$x' \;\; = \;\; p + \Delta p - p_e \;\; = \;\; p_e + x + \Delta p - p_e$$

$$= \;\; x + \Delta p \qquad \simeq \;\; x + ax \qquad\qquad = \;\; x(1 + a)$$

Locally stable if:

$$-2 \;\; < \;\; \left[\frac{d(\Delta p)}{dp} \right]_{p \,=\, p_e} \;\; < \;\; 0,$$

Stability Considerations

Cases:

(1) $a > 0, (1 + a)^t > 1$, so $x' = x(1 + a)^t$ unbounded growth

(2) $-1 < a < 0, 1 + a$ lies between 0 and 1, so $(1 + a)^t \to 0$

(3) $-2 < a < -1, 1 + a$ lies between -1 and 0.
then multiplying x by $1+a$ changes its sign, reduces its magnitude, but converges

(4) $a < -2, 1 + a < -1$, then deviation changes sign and grows

Summary conditions for local stability:

$$-2 < [\frac{d(\triangle p)}{dp}]_{p=p_e} < 0$$

Human variation at genetic code level (genotype) to variation in protein to variation in...

ATG GTG CAC CTG ACT CCT **GA** GAG AAG TCT GCC GTT ACT
ATG GTG CAC CTG ACT CCT **GT** GAG AAG TCT GCC GTT ACT

MVHLTP**E**EKSAVT (E is the single letter abbreviation for glutamic acid)
MVHLTP**V**EKSAVT (V is the single letter abbreviation for valine)

Glutamic acid is a hydrophilic amino acid. Valine is a hydrophobic amino acid.

Analysis of Hb-a, HB-s, and HB-c data
(from Cavalli-Sforza, 1977)

	AA	SS	CC	AS	AC	SC
Observed	25374	67	108	5482	1737	130
Expected	25616	307	75	4967	1769	165
Obs/Exp	0.99	0.22	1.45	1.10	0.98	0.79
Relative fitness	0.89	0.20	1.31	1	0.89	0.70

Suppose just A, S alleles

$$\hat{p}_S = \frac{w_{22} - w_{12}}{(w_{11} - w_{12}) + (w_{22} - w_{12})} = \frac{0.2 - 1.0}{0.89 - 2.0 + 0.2} = 0.1209$$

$p_A = 0.8791$

$\bar{w} = 0.90$

Suppose just a <u>few</u> C
  alleles introduced

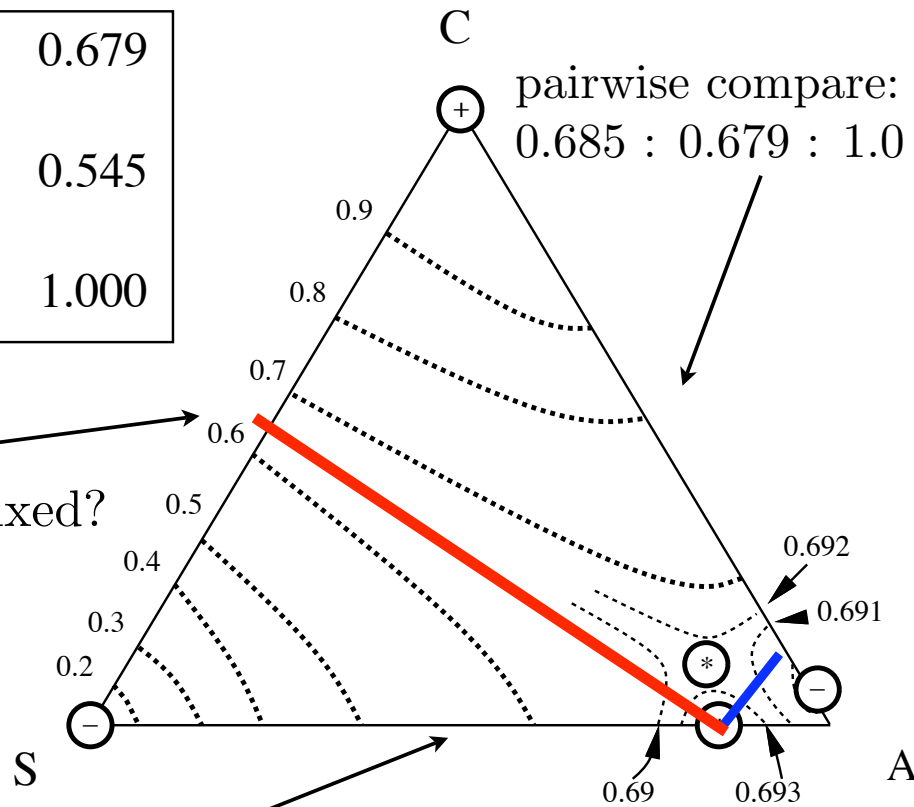$$w_C = p_A w_{AC} + p_S w_{SC} + p_C w_{CC}, \text{ when } p_c \approx 0,$$

$$w_C = p_A w_{AC} + p_S w_{SC} = 0.8670$$

C cannot invade when rare, even
though this yields global fitness!

= frequency of S (sickle)
= frequency of A (normal)

	A	S	C
A	0.685	0.763	0.679
S	0.763	0.151	0.545
C	0.679	0.545	1.000

pairwise compare: C–A
$0.685 : 0.679 : 1.0 \rightarrow$ underdomin

pairwise compare: S–C
$0.151 : 0.545 : 1.0 \rightarrow C$ fixed?

pairwise compare: S–A
$0.685 : 0.763 : 0.151 \rightarrow$ overdominance; $S - A$ mix

Does sex make you fitter?

Why Sex Hurts

Mixing due to diploid mating actualy *reduces* fitness!
In other words: an asexually reproducing (haploid)
population could actually do better!
Let's see how much − this is called *segregational load*
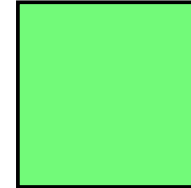Puzzle: why does sex survive?

$$
\begin{array}{ccc}
\text{AA} & \text{Aa} & \text{aa} \\
0.4 & 1 & 0.8
\end{array}
$$

$p(A) = 0.2;$
$p(AA) = 0.04; p(aa) = 0.64; p(Aa) = 2p(1 - p) = 0.32$
$\text{H} - \text{W}$ ratios $:0.04 : 0.32 : 0.04$

What is mean fitness?

What is mean fitness *without* sex?

$$0.04 \times 0.4 + 0.32 \times 1 + 0.8 \times 0.64 \;=\; \boxed{}$$
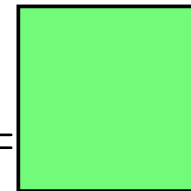
After selection (viability, roughly):

$$0.04 \times 0.4/0.848 : 0.32 \times 1/0.848 : 0.8 \times 0.64/0.848$$

$$0.0189 : 0.3774 : 0.6038$$

Suppose just apply fitness to this in next round
(asexual reproduction), so just multiply
by fitness without meiosis rejuggling:

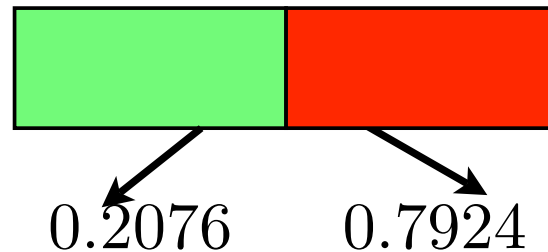$$0.0188 \times 0.4 + 0.3774 \times 1 + 0.6038 \times 0.8 = \boxed{} > 0.848$$
(*by* 0.02)

What is mean fitness *with* sex?

Before meiosis (AA, Aa, aa frequencies *not* in
H-W proportions):

$$0.0189 : 0.3774 : 0.6038$$

H-W (i.e., sex) spreads some (half) of the A's from Aa to A,
noindent and some of the a's from Aa to aa:

$$0.2076 \qquad 0.7924$$

So after meiosis, AA, Aa, aa ratios are p^2, $2pq$, q^2, or:
$$0.0431 : 0.3290 : 0.6279$$

mean fitness =
$$0.0431 \times 0.4 + 0.3290 \times 1 + 0.6279 \times 0.8 = 0.84856$$
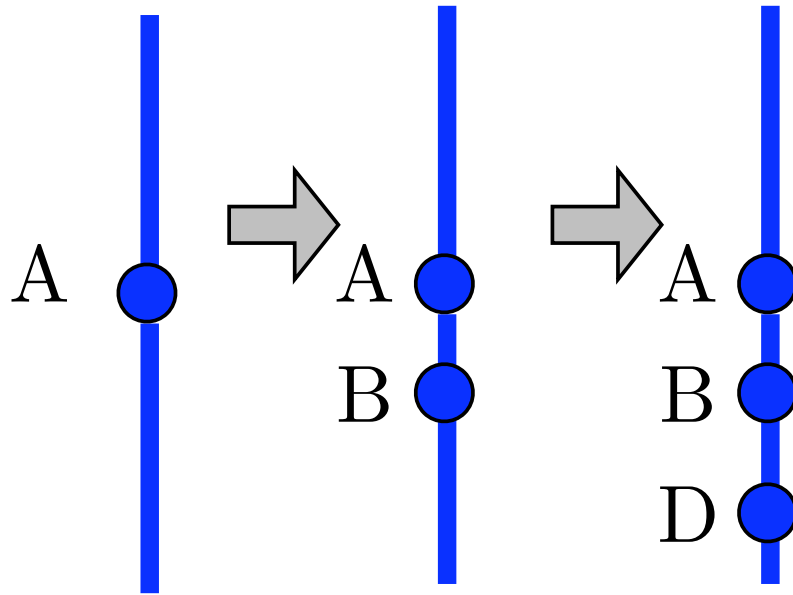
Sex hurts

mean fitness *without* sex =
$0.01881 \times 0.4 + 0.3774 \times 1 + 0.6038 \times 0.8 = 0.862$
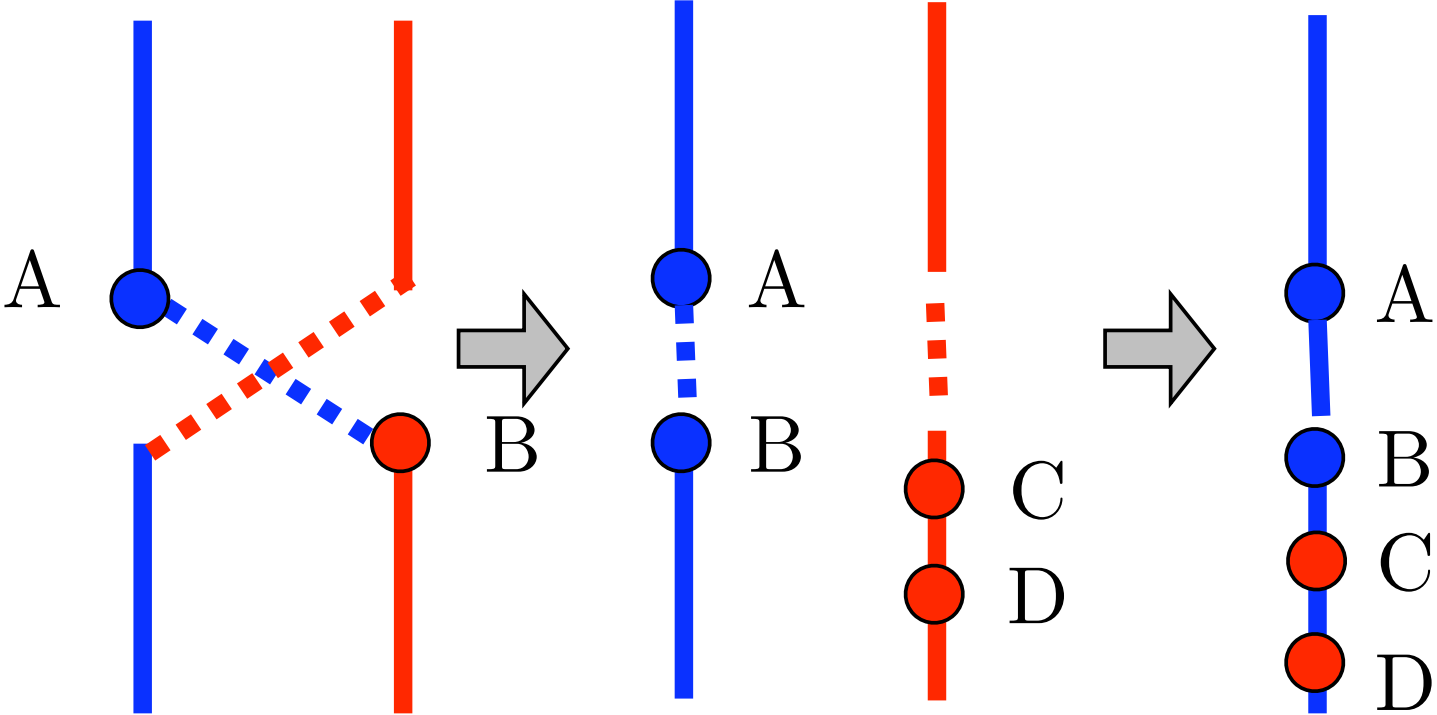mean fitness *with* sex =
$0.0431 \times 0.4 + 0.3290 \times 1 + 0.6279 \times 0.8 = 0.84856$

Question: why then does sex exist?

Sex helps!

Sex helps!

When does fitness maximization fail?
The case of frequency dependent selection

1. *Specialization on different limiting resources.* If two genotypes eat different foods, then an individual of the rare genotype will have a more abundant source of food, by virtue of the rareness of other individuals who eat that food. The same argument will hold for many other limiting nonfood resources, such as breeding sites.

2. *Different diseases or parasites for different genotypes.* If each genotype has its own diseases and parasites, then whichever type is rarer will be less likely to come into contact with carriers of its own particular pests.

3. *Specialization of different predators on different genotypes.* When each genotype has its own predators, then the genotype which is rare will presumably sustain a lower population density of predators, and hence might suffer a lower mortaility rate from predation.

4. *Predator search images: apostatic selection.* Many intelligent visual predators form "search images" of the desired appearance of their prey. They tend to reject potential prey which do not fit this image. The search image depends on the last few prey eaten. Thus the predators may tend to avoid taking the rare genotypes, which they have not encountered recently.

When does fitness maximization fail?

5. *Rare male advantage.* In some species, notably *Drosophila melanogaster*, males of a rare genotype seem to have an advantage in mating simply because they are rare. This pattern of female choice may be an adaptation to avoid inbreeding.

6. *Social Interactions.* In a social species, if the genotypes differ in their social behavior, the fitness of a genotype may depend on the frequencies of the genotypes among the individuals it encounters in the population.

Simplest case of frequency dependence: haploid A, a

$$w_A = 1 + t - sp$$

$$w_a = 1.$$

$$\frac{p'}{1 - p'} = (1 + t - sp)\frac{p}{1 - p}$$

$$p' = \frac{p(1 + t - sp)}{1 + (t - sp)p}.$$

$$\Delta p = p' - p = (p(1 + t - sp) - [1 + (t - sp)p])/(1 + (t - sp)p)$$

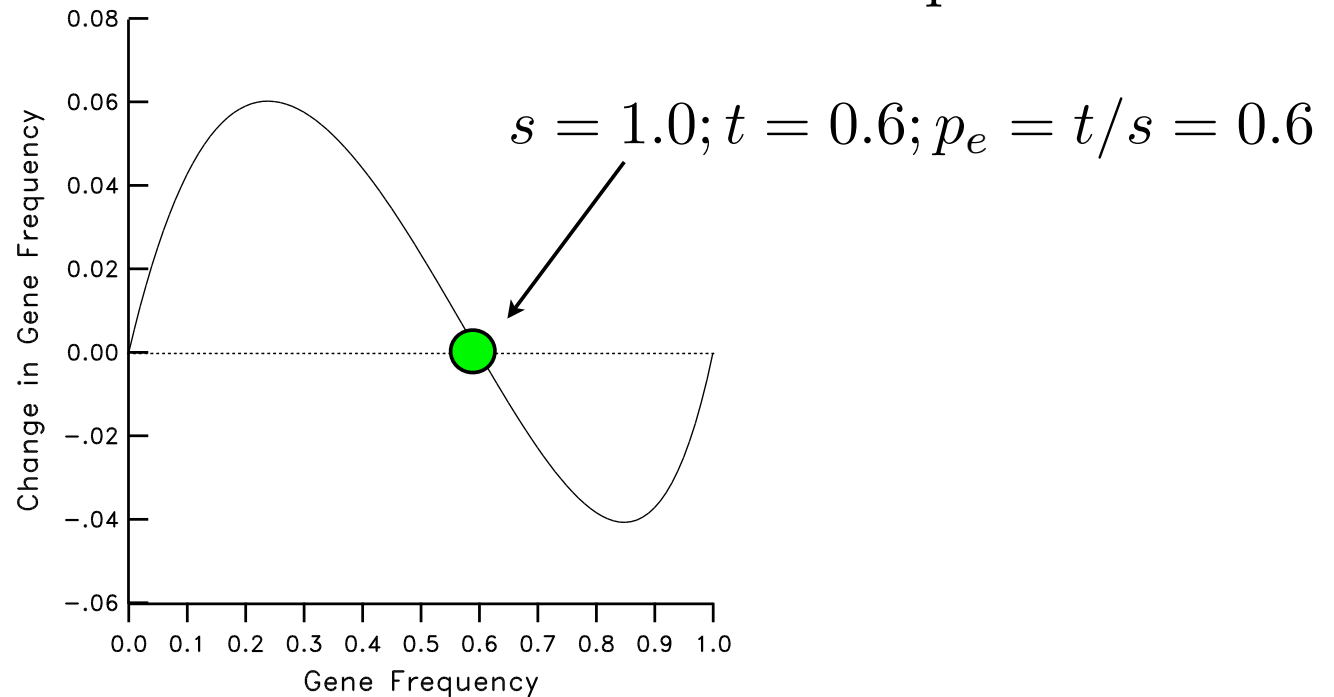$$= p(1 - p)(t - sp)/(1 + (t - sp)p).$$

$$\Delta p \quad = p(1-p)(t-sp)/(1+(t-sp)p).$$

Equilibrium p when $\triangle p = 0$
i.e., numerator $=0$, so $p = 0$ or $p = 1$ or $p = t/s$
Last in $[0, 1]$ if $s > t > 0$ or if $0 > t > s$ (o.w., $w_A > w_a$)

Fitness is *not* maximized at this stable point!

$$s = 1.0; t = 0.6; p_e = t/s = 0.6$$

Δp as a function of p for a case of frequency-dependent selection. The relative fitness of genotype A is $1.6 - p$.

Fitness is *not* maximized at this stable point!

At equilibrium, $p = t/s,$ $w_A = 1 = w_a$
so mean relative fitness is 1:

$$w_A = 1 + t - sp$$

$$w_a = 1.$$

$$\bar{w} = pw_A + (1-p)w_a = p + (1-p) = 1.$$

$$\bar{w} = pw_A + (1-p)w_a$$
$$= p(1 + t - sp) + (1-p)$$
$$= 1 + p(t - sp).$$

$$\frac{d\bar{w}}{dp} = t - 2sp = 0, \qquad p = \frac{t}{2s} \qquad \bar{w} = 1 + \frac{t^2}{4s}.$$

Frequency dependent fitness wrecks method of
using slope of w to direct local search

$$p_e = \frac{w_{Aa}(p_e) - w_{aa}(p_e)}{[w_{Aa}(p_e) - w_{aa}(p_e)] + [w_{Aa}(p_e) - w_{AA}(p_e)]}$$

This equation has many roots, depending on w wrt p
we divide w's by $w_{Aa}(p)$ to get fitnesses:

$$
\begin{array}{ccc}
AA & Aa & aa \\
1 - s(p_e) & 1 & 1 - t(p_e)
\end{array}
$$

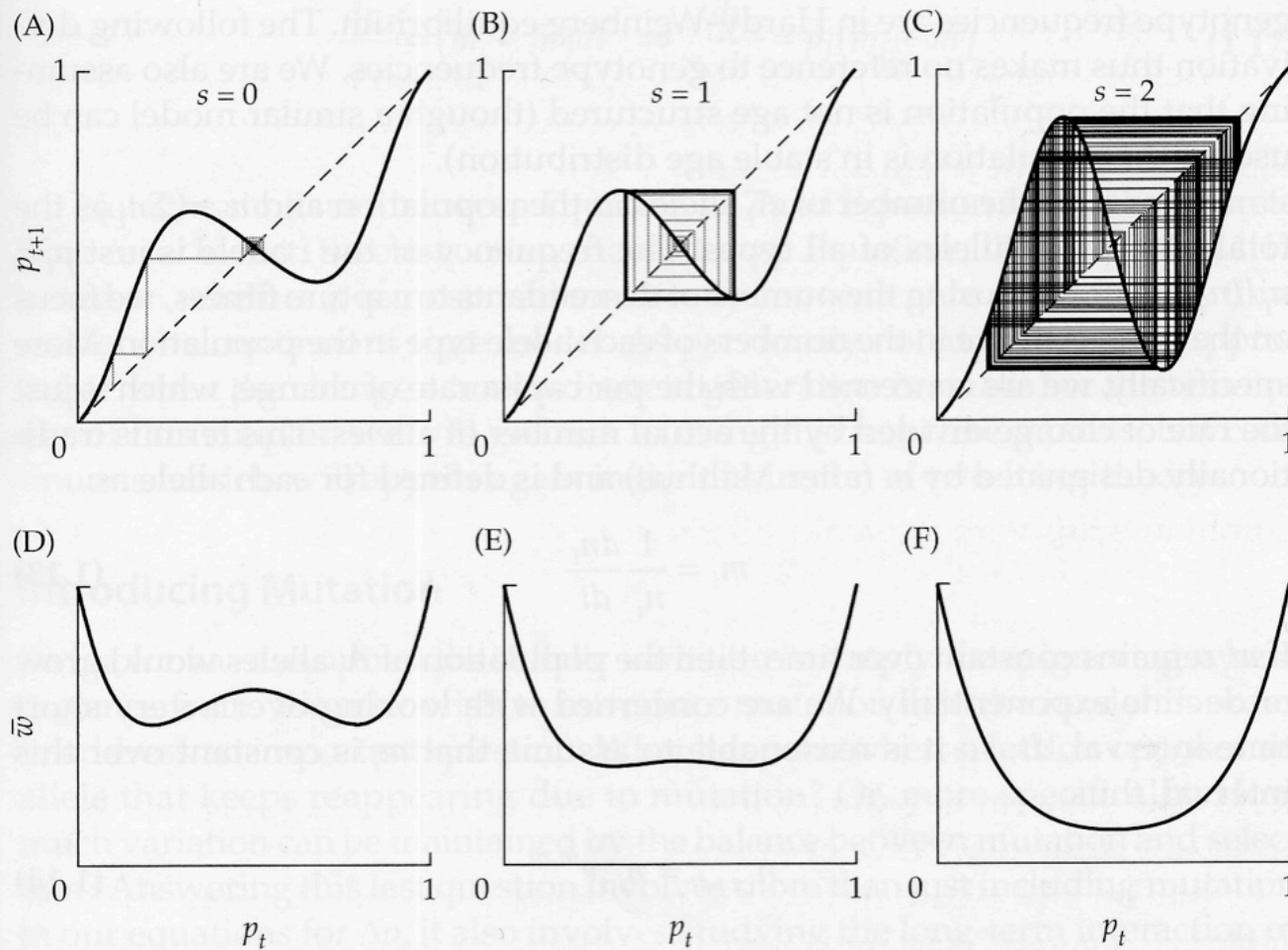$$p_e = \frac{t(p_e)}{s(p_e) + t(p_e)}.$$

Figure 1.2 Evolution under frequency-dependent selection. Denoting the frequency of genotype A_iA_j by γ_{ij}, the fitness functions shown are all of the form: $w_{11} = 1 - 3\gamma_{12} + 3\gamma_{22}$, $w_{12} = 1 - s\gamma_{12}$, $w_{22} = 1 - 3\gamma_{12} + 3\gamma_{11}$, for different values of the parameter s. In (A) and (B) the gray lines show the first 200 iterations from the starting point, with the following 200 iterations shown in black. The black lines thus show the asymptotic behavior of the system. (A) For $s = 0$, there is a stable point; (B) for $s = 1$, the system settles into a limit cycle; (C) for $s = 2$, the system shows chaotic fluctuations, which are aperiodic and would eventually fill in a region of the graph. (D), (E), and (F) show mean population fitness as a function of p for the systems shown in (A), (B), and (C), respectively.

Why this causes problems:
Current fitness is no longer a good guide
to *future* fitness!

Multiple allele case

$$\bar{w} = \sum_i \sum_j p_i p_j w_{ij} = \text{mean fitness}$$

$$p_i' = \frac{\frac{1}{2} \sum\limits_{j=1}^{n} p_i p_j w_{ij} + \frac{1}{2} \sum\limits_{j=1}^{n} p_j p_i w_{ij}}{2 \times \frac{1}{2} \sum \sum p_i p_j w_{ij}}$$

$$= \frac{p_i \left(\sum\limits_j p_j w_{ij} \right)}{\bar{w}},$$

$= \text{mean fitness of organisms in } A_i \text{ alleles find themselves weighted by the number of } A_i \text{ alleles they contain}$

$$p_i' = \frac{p_i \bar{w}_i}{\bar{w}}$$

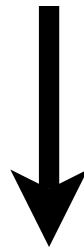$$\Delta p_i = p_i' - p_i = p_i \frac{\bar{w}_i}{\bar{w}} - p_i$$

$$= p_i \frac{(\bar{w}_i - \bar{w})}{\bar{w}}.$$

Equllibrium conditions

At equillibrium, for allele i, either
$p_i{=}0$ or $\bar{w}{=}0$

The shape of things to come

$$\Delta p = \frac{p(1-p)}{\overline{w}} \frac{d\overline{w}}{dp}$$

$$\downarrow$$

$$\Delta \vec{p} = \frac{G}{\overline{w}} \nabla \overline{w}$$

$$\bar{w} = \sum_i \sum_j p_i p_j w_{ij}$$

$$\frac{\partial \bar{w}}{\partial p_i} = 2\sum_j p_j w_{ij} = 2\bar{w}_i \qquad p_i' = p_i \frac{\bar{w}_i}{\bar{w}}$$

$$p_i' = \frac{p\bar{w}_i}{\bar{w}} = \frac{2}{\bar{w}} \frac{\partial \bar{w}}{\partial p_i}$$

$$\Delta p_i = \frac{p_i}{2\bar{w}} \frac{\partial \bar{w}}{\partial p_i} - p_i \frac{\bar{w}}{\bar{w}} \text{ substitute for } \bar{w} = \sum_j \frac{p_j}{2} \frac{\partial \bar{w}}{\partial p_j}$$

$$\Delta p_i = \frac{p_i}{2\bar{w}} \left(\frac{\partial \bar{w}}{\partial p_i} - \sum_j p_j \frac{\partial \bar{w}}{\partial p_j} \right)$$

The multiple allele jet-fuel formula

$$\Delta \vec{p}_i = \Delta \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \frac{1}{2\overline{w}} \begin{bmatrix} p_1(1-p_1) & -p_1 p_2 & -p_1 p_3 \\ -p_2 p_1 & p_2(1-p_2) & -p_2 p3 \\ -p_3 p_1 & -p_3 p_2 & p_3(1-p_3) \end{bmatrix} \begin{bmatrix} \dfrac{\partial \overline{w}}{\partial p_1} \\[2mm] \dfrac{\partial \overline{w}}{\partial p_2} \\[2mm] \dfrac{\partial \overline{w}}{\partial p_3} \end{bmatrix}$$

variance

$$\Delta \vec{p} = \frac{G}{\overline{w}} \nabla \overline{w}$$

grad mean fitness
wrt p_i