Peas & Probabilities (A Simplification of a Process)

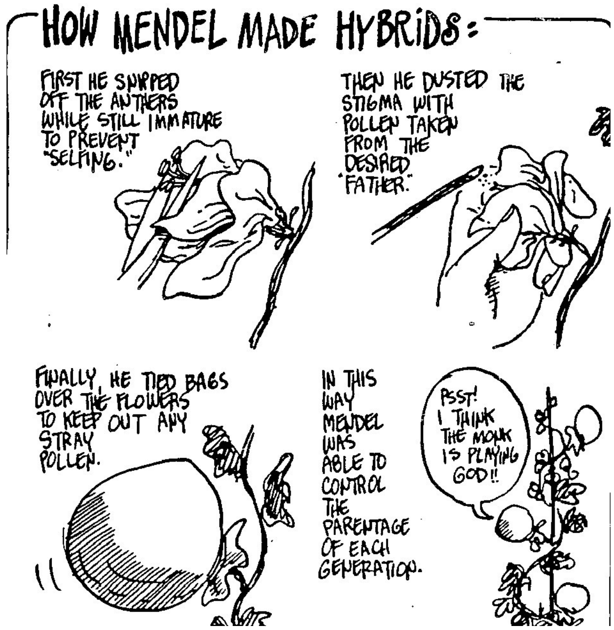
In the mid-19th century, Gregor Mendel, observed that apparently identical peas from plants which appeared identical, or even the same plant, when planted do not produce plants that produce identical peas.

Subsequently Mendel did experiments to try to understand how traits are passed from one generation to the next. Lucky he chose pea plants which had easily distinguishable and independent traits.

Mendel started his studies with simple cases – cross breeding between plants which differed in a singe trait. (flower color, seed color, seed shape, pod color, pod shape, flower position, or plant height)

Such crosses are now called monohybrid crosses.

**Mendel observed the pollen from the anther of a flower would *self-fertilize* by landing on the stigma of the same flower. However, if he removed immature anthers from a plant's flowers, pollen from another plant could be placed on the stigma allowing *cross-fertilization*. He could then make the desired genetic crosses
(*hybrids*).**

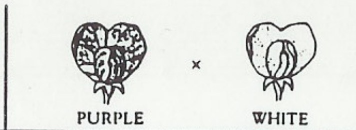
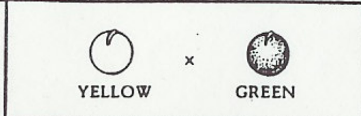
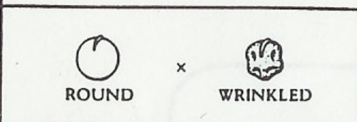
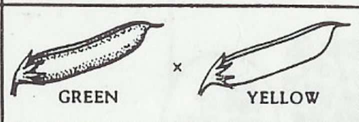
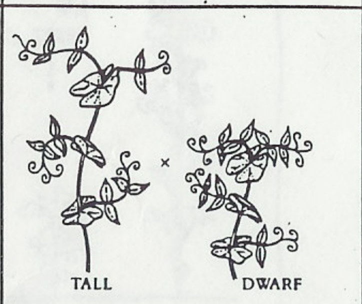


Mendel needed to be sure that his plants were *true-breeding* or of a *pure line*. Such plants produce only plants just like itself when it is allowed to self fertilize.

Once Mendel identified plants of a pure line with respect to a specific trait (or traits), he cross-fertilized them and collected seeds at the end of each generation or growing season. He cataloged each seed keeping track of the characteristics of the parent plant from which they were produced. The next year he planted the seeds and observed the characteristics of the plants produced. In his first generation crosses he observed the following results:

green seeds X yellow seeds → All offspring had yellow seeds
round seeds X wrinkled seeds → All offspring had round seeds
purple flowers X white flowers → All offspring had purple flowers
green pods X yellow pods → All offspring had green pods
round pods X constricted pods → All offspring had round pods
axial flowers X top flowers → All offspring had axial flowers
tall plants X dwarf plants → All offspring were tall

In the second generation Mendel allowed the offspring of the first cross to self-fertilize. The traits found in the second generation are shown in the talble below:

Trait		**Frequency**	
FLOWER COLOR	 PURPLE × WHITE	705	224
SEED COLOR	 YELLOW × GREEN	6022	2001
SEED SHAPE	 ROUND × WRINKLED	5474	1850
POD COLOR	 GREEN × YELLOW	428	152
POD SHAPE	 ROUND × CONSTRICTED	882	299
FLOWER POSITION	 AXIAL × TOP	651	207
PLANT HEIGHT	 TALL × DWARF	787	277

It looks like the alternative forms are demonstrated in about a 3 to 1 ratio.

Towards a Real Model

In what follows we will restrict our studies to the consideration of two traits – color and texture. Each of those two traits has two attributes. The color can be green or yellow; the texture can be smooth or wrinkled.

We assume that each pea can be assigned to one of the following types.

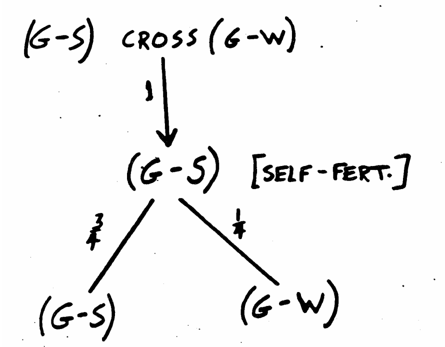
green-smooth (G-S)	**green-wrinkled (G-W)**
yellow-smooth (Y-S)	**yellow-wrinkled (Y-W**

Once pure lines have been identified, first we study and classify the results of cross fertilization between pure lines. Next we allow the first generation plants to reproduce by self fertilization, and we consider the characteristics of the second generation plants.

Observations:

- **All the first generation peas are of the same type.**
- **When the first generation peas are allowed to reproduce by self fertilization they do not reproduce in a consistent manner.**

Example:



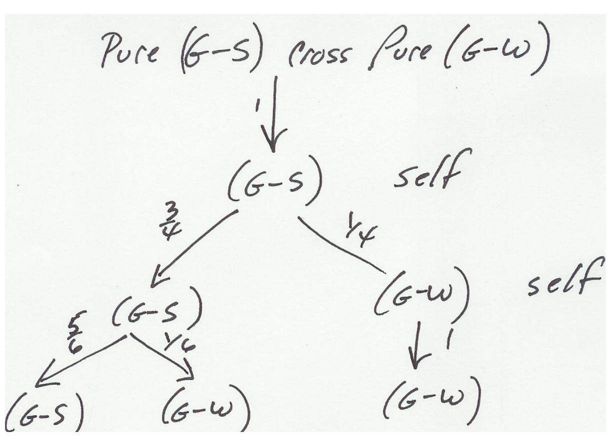
(G-S) seed can have either (G-S) or (G-W) descendents. Superficial appearance does not completely determine the results of reproduction by self fertilization.

Recognizing that type based on appearance was inadequate for his purposes, Mendel used the term *phenotype* to refer to type based on appearance, and moved on to consider further notions of type.

Mendel conjectured that seeds must carry some undetectable units which enter into the reproductive process which could account for the 3 to 1 distribution observed in his experiments.

Mendel continued his experiments by allowing his seeds in successive generations to reproduce by self fertilization.

Example:



For now we will focus on a single characteristic of the pea plant – texture. As noted, Mendel conjectured that each cell of a plant other than those involved in reproduction carries a fundamental unit that determines the texture of its descendants. This fundamental unit consists of a pair of *genes*. (Cells actually involved in reproduction carry only one gene.) Since were two alternative forms for texture, Mendel assumed there were two forms for the associated genes. During reproduction the pair of genes would be split and one of them would be contributed to each offspring. That is, according to Mendel, reproductive cells, *gametes*, each carry one gene associated with texture and that gene is selected at random from the pair of the respective parent. The cell created during reproduction receives two gametes, one selected at random from each parent.

On the basis of observation a plant will bear smooth peas if its seeds acquired a gene associated with smoothness from either the male or female gamete, and it will bear wrinkled peas only if its seed acquired a gene associated with wrinkledness from both gametes. The phenomenon of an individual demonstrating the form of a characteristic associated with one form of a gene even though both forms are present is known as *dominance*. In this case we say smoothness is the *dominant* form of texture. Wrinkledness will be called the *recessive* form of texture.

Each cell may be classified according to its genetic composition and so can the plant. The genetic nature of a plant is called its *genotype*.

Question: What is the connection between a plants phenotype and genotype?

Question: Knowledge of the genotype of an individual and the dominance relations is sufficient to determine its phenotype, but is the converse true?

A Mathematical Model

We begin by considering a single characteristic (texture) which has two alternative forms (smooth and wrinkled). The *gene* associated with texture can take on two alternative forms called *alleles*. We denote the alleles associated with texture by the symbols *A* and *a*. In the real model above genes occur in pairs so we consider the set of unordered pairs V= { [A,A], [A,a], [a, a]} = {AA, Aa, aa}. If we agree that alleles denoted by A and a correspond to smooth and wrinkled forms respectively, then as a consequence of the dominance relation, peas with genes AA or Aa will be smooth and those with genes aa will be wrinkled.

Our First Mathematical Model

Undefined terms: gene, reproduction

Axiom 1: Each gene occurs in two forms (alleles) denoted by A and a respectively.

Definition: V = {AA, Aa, aa} is the set of genotypes.

Axiom 2: Reproduction is a function from V X V into R^3.

We denote the reproduction function by *r*; so

***r:* V X V $\rightarrow$ *R*^{3}.**

Let r_1, r_2, **and** r_3 **denote the coordinate functions; so for** (u,v) $\in$ **V** X **V,**

$$r: (u,v) \rightarrow [r_1(u,v) \quad r_2(u,v) \quad r_3(u,v)]$$

Definition: For each u $\in$ **V and** $\boldsymbol{\alpha} \in$ **{A, a}, let** $p(\boldsymbol{\alpha} \mid u)$ **be the conditional probability that** $\boldsymbol{\alpha}$ **is selected when a random choice is made between the two letters that make up** u, **each choice being equally likely.**

$p(\boldsymbol{\alpha} \mid u)$		u **AA**	**Aa**	**aa**
$\boldsymbol{\alpha}$	**A**	**1**	**0.5**	**0**
	a	**0**	**0.5**	**0.5**

Axiom 3: The reproduction function *r* satisfies

$$\begin{aligned} r_1(u,v) &= p(\mathbf{A} \mid u)\, p(\mathbf{A} \mid v) \\ r_2(u,v) &= p(\mathbf{A} \mid u)\, p(\mathbf{a} \mid v) + p(\mathbf{a} \mid u)\, p(\mathbf{A} \mid v) \\ r_3(u,v) &= p(\mathbf{a} \mid u)\, p(\mathbf{a} \mid v) \end{aligned}$$

A *probability vector* is a vector whose coordinates are nonnegative numbers that sum to 1.

Theorem: The range of the function *r* is a set of probability vectors in R^3

Example: If u **= AA and** v **= Aa, then *r*(**u,v**) = [½ ½ 0].**

***r*(AA, Aa) =**

[{*p*(A | AA) *p*(A | Aa)} {*p*(A | AA) *p*(a | Aa)+ *p*(a | AA)*p*(A | Aa)} {*p*(a | AA) *p*(a | Aa)}]

= [(1)(½) (1)(½) + (0)(½) (0)(½)] = [½ ½ 0]

r(u,v)		v	
	AA	Aa	aa
AA	[1 0 0]	[½ ½ 0]	[0 1 0]
u Aa	[½ ½ 0]	[¼ ½ ¼]	[0 ½ ½]
aa	[0 1 0]	[0 ½ ½]	[0 0 1]

Consider a mating between two pure lines.

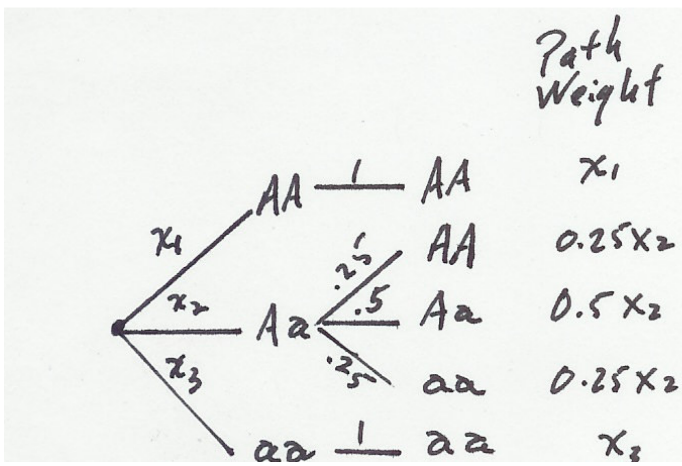
***r*(AA,aa) = [0 1 0]**

What occurs when if a pea of genotype Aa, called a *hybrid*, mates with itself or a pea of the same genotype? (What does that result signify for the phenotypic distributions noted by Mendel?)

Theorem. **If the genotypic distribution at one generation is given by x = [** $x_1 \quad x_2 \quad x_3$ **], where** x_1, x_2, x_3 **are the proportions of the population that are of types AA, Aa, and aa respectively and if reproduction is by selfing, then the genotypic distribution at the next generation is given by the matrix product xM, where M is the matrix**

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0.25 & 0.50 & 0.25 \\ 0 & 0 & 1 \end{bmatrix}.$$

Consider the following decision tree.



Proof of Theorem. **If the genotypic distribution at one generation is given by x = [** $x_1 \quad x_2 \quad x_3$ **], by the above discussion at the next generation the genotypic distribution is**

$[x_1 + 0.25x_2 \quad 0.50x_2 \quad 0.25x + x_3] =$

$x_1[1 \;\; 0 \;\; 0] + x_2[\tfrac{1}{4} \;\; \tfrac{1}{2} \;\; \tfrac{1}{4}] + x_3[0 \;\; 0 \;\; 1] =$ **xM.**

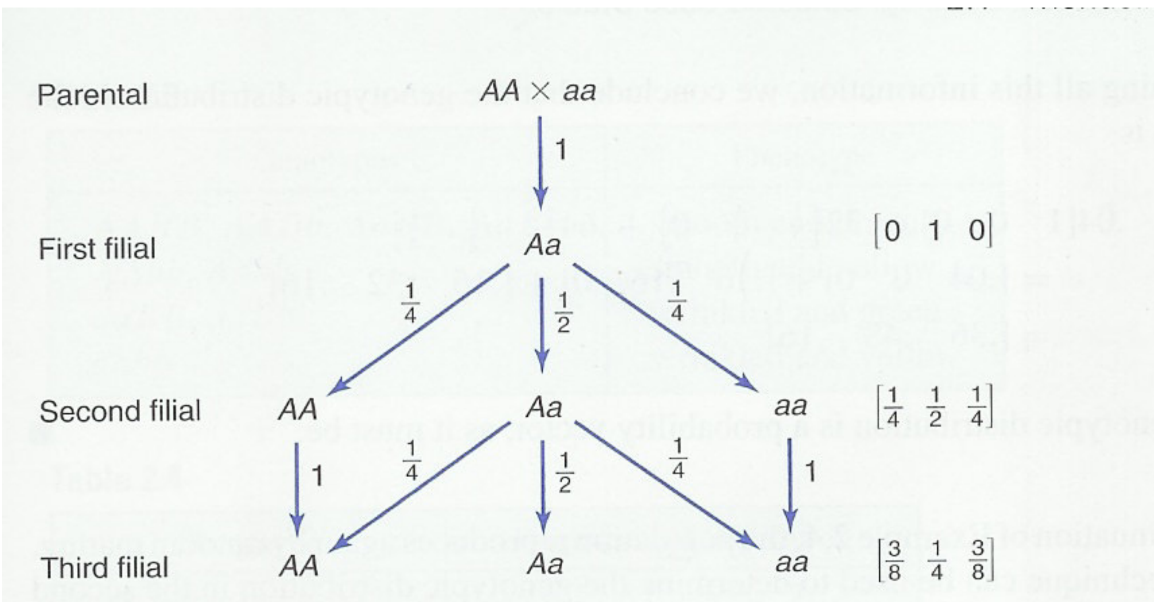
Question: If a population reproduces by selfing, what happens to the percentage of hybrids in each succeeding generation?

Hint: **If the genotypic distribution in the initial generation is $\mathbf{x_0} = [x_1\ \ x_2\ \ x_3]$, then by the above theorem, if the population continues to reproduce by selfing, the genotypic distributions in succeeding generations will be given by**

$$\mathbf{x_1} = \mathbf{x_0}\mathbf{M},\quad \mathbf{x_2} = \mathbf{x_1}\mathbf{M} = (\mathbf{x_0}\mathbf{M})\mathbf{M} = \mathbf{x_0}\mathbf{M}^2,\quad \mathbf{x_3} = \mathbf{x_2}\mathbf{M} = (\mathbf{x_0}\mathbf{M}^2)\mathbf{M} = \mathbf{x_0}\mathbf{M}^3, \ldots$$

Example: Suppose a pure-line dominant pea is crossed with a pure line recessive pea. Find the genotypic distributions of the first three *filial* generations, assuming reproduction is by selfing after the initial cross.

We illustrate this situation with a *reproduction diagram*.



Example: Suppose a population has genotypic distribution [0.2 0.8 0.0], and suppose this population reproduces by random mating. Find the genotypic distributions for the next two generations.