Occam’s Razor

The Nature of Scientific Theories

WILLIAM D. STANSFIELD

The early 14th century philosopher William of Occam (Sahakian 1968, p. 116) is credited by historians for proposing that, in interpreting scientific data, the simplest explanation consistent with all the known facts is preferred as a working hypothesis over more complicated ones. In other words, scientific explanations should be pared down (as if using a knife to cut away unnecessary “fat” from a carcass) to the simplest hypothesis as a preliminary orientation for further study of any scientific problem. This guiding principle of scientific methodology has come to be known as “Occam’s razor.” Genetics is a scientific discipline in which Occam’s razor (and other facets of scientific methodology) can be easily explained and grasped by most middle- and high-school students. The following genetics problems can be used to illustrate this point.

In nature, mice usually have a nondescript "mousy" color pattern called “agouti” (named after a South American rodent with the same pattern). The agouti pattern involves a subapical (near the tip) band of yellow pigment in the hair. If the rest of the hair is black, the phenotype is known as “wild-type” (the normal phenotype found in natural populations). Thus, the wild-type mouse is actually a black-agouti phenotype. If the yellow band appears on an otherwise brown hair (a mutant or rare character in nature), the phenotype is called brown-agouti or cinnamon. Two other mutant types are nonagouti patterns (no yellow band): solid brown (chocolate) and solid black. Students should have, as background information, an understanding of dominant and recessive alleles, multiple alleles, independent assortment, and the statistical inferences that are normally drawn from 3:1 and 9:3:3:1 phenotypic ratios.

Five sets of hypothetical matings will be presented. The parental generation (P) is assumed to consist of genetically pure (homozygous) breeding individuals. Large numbers of progeny are produced in the F1 and F2 generations. The F1 generation is produced by allowing the F1 population to mate at random. As the data from each set of matings is displayed (e.g., on an overhead projector), the instructor can ask the class to propose suitable explanations. The fifth set of matings should not be revealed until the first four sets have been analyzed.

Set #1

P: wild-type x cinnamon
F1: all wild-type
F2: 3/4 wild-type: 1/4 cinnamon

WILLIAM D. STANSFIELD is an Emeritus Professor in the Biological Sciences Department at California Polytechnic State University, San Luis Obispo, CA 93407. He can be reached via e-mail: wstansh@calpoly.edu.
Inference

A single pair of alleles is segregating in the production of the F<sub>2</sub> generation, with a dominant gene (say D) coding for wild-type and its recessive allele (say d') for cinnamon; i.e., D > d'.

Set #2

P: wild-type x black
F<sub>1</sub>: all wild-type
F<sub>2</sub>: 3/4 wild-type: 1/4 black

Inference

Another pair of alleles of the same gene is segregating in the production of the F<sub>2</sub> generation, with dominant gene (D) for wild type, as proposed previously, and a different recessive allele (say d'') for black. Here we are using Occam’s razor for the first time. We are trying to explain both sets of data (Sets #1 and #2) by proposing that three alleles exist at the D-locus in this population of mice. The dominance hierarchy is proposed as follows: D > (d'' and d'). We do not know, at this point, the relationship (Allchin 2000) among the alleles in parentheses.

Set #3

P: cinnamon x brown
F<sub>1</sub>: all cinnamon
F<sub>2</sub>: 3/4 cinnamon: 1/4 brown

Inference

Another pair of alleles is segregating in the production of the F<sub>2</sub> generation, with the cinnamon gene (d'') dominant to the recessive brown allele (say d'). A proposed dominance hierarchy is D > d' > d''. At this point, we infer that allele d'' is lower in the dominance hierarchy than allele D, but its exact relationship to alleles d' and d'' has not been established.

Set #4

P: black x brown
F<sub>1</sub>: all black
F<sub>2</sub>: 3/4 black: 1/4 brown

Inference

The black allele (d'') is dominant to the brown allele (d'). A proposed dominance hierarchy is: D > (d' and d'') > d'. Now, the only unknown relationship is that between the alleles for cinnamon and black.

Set #5

P: cinnamon x black
F<sub>1</sub>: all wild-type
F<sub>2</sub>: 9/16 wild-type: 3/16 cinnamon: 3/16 black: 1/16 brown

The results of Set #5 will “knock the socks” off of some students. Even the phenotype of the F<sub>1</sub> is obviously inconsistent with the original hypothesis of multiple alleles at a single genetic locus. The most dominant phenotype (wild-type) cannot be produced by matings involving only alleles lower in the dominance hierarchy. Can a more complex hypothesis accommodate all the data? What inference would normally be drawn from a 9:3:3:1 F<sub>2</sub> phenotypic ratio? Answer: Let us assume that the F<sub>1</sub> mice in Set #5 were dihybrids (heterozygous at two independently assorting loci, as on different pairs of chromosomes). We could then define a pair of alleles (A and a) at one locus that governs the agouti (yellow band) and nonagouti (solid color) patterns, respectively; and at an independently assorting locus, another pair of alleles (B and b) that
governs black and brown, respectively. Set #5 can then be genetically defined as follows:

P:  $AAbb$ (cinnamon) $\times aabb$ (black)

F₁:  all $AaBb$ (wild-type)


(Note: Each dash in the above genotypes indicates an unknown and irrelevant allele to the production of the phenotype.)

Similarly, the other four sets of matings can be explained in like fashion; it would be good experience for students to do this on their own for homework or perhaps work on it as teams in the classroom. The instructor can take this opportunity to point out that the procedure we have used is precisely the way that scientific knowledge grows. Using Occam’s razor (proposing a single genetic locus rather than two loci) as our initial hypothesis was good scientific methodology and worked well in explaining the data through the first four sets of matings. But then came the fifth set of data that (according to Thomas Henry Huxley) exemplifies “The great tragedy of science – the slaying of a beautiful hypothesis by an ugly fact” (Mackay 1977). So, it was back to square one (perhaps a Punnett square?) in attempts to formulate a more complicated hypothesis that could explain all of the data. All scientific knowledge is tentative in this sense. We may never have all of the data that we want or need to formulate a robust hypothesis. Progress is not made by waiting for more data. Oftimes we must act on the data and hypotheses that we have, realizing that either may be faulty. Any good scientific hypothesis, however, should point the way to new experiments that will test its falsifiability (Popper 1980, p. 22). A hypothesis that has “stood the test of time” and resisted all attempts at falsification will gradually become accepted as a valid scientific theory.

References


