SIMPSON'S PARADOX

To exemplify the phenomenon under discussion, we begin by considering the batting of two baseball players. Suppose that Jackie is at bat 100 times, during which he makes 30 hits for a proportion 0.30. On the other hand, Joe is also at bat 100 times but only collects 27 hits, for a proportion 0.27. Of course a batter's performance may depend on whether the pitcher throws left-handed or right-handed, so we consider how Jackie and Joe fare in these two cases:

	AGAINST LEFTIES			**AGAINST RIGHTIES**		
	At Bats	**Hits**	**Ave**	**At Bats**	**Hits**	**Ave**
Jackie	25	5	.20	75	25	.33
Joe	80	20	.25	20	7	.35

Thus Joe performs better against both left-handed and right-handed pitching, even though Jackie had the better overall performance. Although this might seem surprising at first, the explanation is easy to find. Both batters performed better against right-handed pitching, but Jackie was fortunate to have most of his batting against such pitchers, while Joe faced far more left-handed pitchers. The figures we have used here are ficticious, but Simpson's paradox does arise frequently this way in baseball statistics.

To consider Simpson's Paradox more abstractly, suppose that A and B are two events, and that $F_1, F_2, \ldots, F_n$ is a partitioning of our probability space into pairwise disjoint events. If $\mathbf{P}(A|F_i) > \mathbf{P}(B|F_i)$ for all i then it follows that $\mathbf{P}(A) > \mathbf{P}(B)$, since by conditioning we see that

$$\mathbf{P}(A) = \sum_{i=1}^{n} \mathbf{P}(A|F_i)\mathbf{P}(F_i) \geq \sum_{i=1}^{n} \mathbf{P}(B|F_i)\mathbf{P}(F_i) = \mathbf{P}(B).$$

But this is not Simpson's Paradox. In Simpson's Paradox we are given that $\mathbf{P}(X|AF_i) > \mathbf{P}(X|BF_i)$, but these inequalities do not imply that $\mathbf{P}(X|A) > \mathbf{P}(X|B)$.

Our first example of Simpson's Paradox seemed quite reasonable, but now we consider a second hypothetical example that is a littler harder to live with. Suppose we are treating a life-threatening disease. We have to choose between an old method of treatment and a new one. To assess whether the new treatment is better than the old, we take 80 patients, and treat 40 of them in the new way, and 40 of them in the old. We find that among the patients treated in the new way, 20 are cured and 20 die, while among the patients treated in the old way, 24 are cured and 16 die.

OVER

	cured	**died**
new	20	20
old	24	16

Here the old cure-rate, 60%, was better than the new, 50%. But now we make a finer analysis of the same data by examining separately male and female patients. Suppose that when this is done, the following numbers emerge:

MALES

	cured	**died**
new	8	2
old	21	9

FEMALES

	cured	**died**
new	12	18
old	3	7

Among male patients the new cure rate, 80%, is better than the old cure rate, 70%. Also, among female patients, the new cure rate, 40%, is better than the old, 30%. Thus a man would prefer the new treatment, and so would a woman, but if patients in general are treated in the new way then fewer will survive.

The instances of Simpson's Paradox we have considered thus far have been hypothetical, but the phenomenon has been noted many times in real life data. We mention a few examples.

- Comparison of TB deaths in 1910 in New York City versus Richmond, Virginia reveal that the mortality was lower in New York City. However, the mortality among whites was higher in New York City, and the mortality among blacks was also higher in New York City.

- From January 1979 to February 1979, the renewal rate of subsriptions to *American History Illustrated* rose from 51.2% to 64.1%. However, when the renewals were broken down into disjoint categories (gift, previous renewal, subscription service, catalog order), every category showed a lower renewal rate.

- The US Federal Income Tax rates between 1974 and 1978 decreased in every income category, but rose overall.

Many learned papers have been written in an effort to explain Simpson's Paradox (for a start, see the March 1976 issue of *Scientific American*, or C. R. Blyth, "On Simpson's Paradox and the Sure-Thing Principle", *Jour. Amer. Stat. Assoc.* **67** (1972), 364–381). Perhaps the simplest useful observation is that Bayes' formula expresses a probability $\mathbf{P}(X|A)$ as a weighted average of conditional probabilities $\mathbf{P}(X|AF_i)$,

$$\mathbf{P}(X|A) = \sum_i \mathbf{P}(X|AF_i)\mathbf{P}(F_i|A).$$

This weighted average may lie anywhere between the least or the largest of the conditional probabilities, depending on the weights $\mathbf{P}(F_i|A)$. A similar formula applies to $\mathbf{P}(X|B)$, but a disparity may arise if the weights $\mathbf{P}(F_i|A)$ are large for those i for which $\mathbf{P}(X|AF_i)$ is small, while the weights $\mathbf{P}(F_i|B)$ emphasize those i for which $\mathbf{P}(X|BF_i)$ is large.