Reprinted with permission from the 2001 Yearbook of the National Council of the Teachers of Mathematics, copyright 2001. All rights reserved.

11

Representations of Reversal

An Exploration of Simpson's Paradox

Lawrence Mark Lesser

WHEN updating its *Standards* documents, the National Council of Teachers of Mathematics (NCTM) added a pre-K-12 Standard on representation, urging that students be able to develop a repertoire of mathematical representations that can be used purposefully and flexibly to model and interpret physical, social, and mathematical phenomena (NCTM 2000). This article aims to explore the potential of including multiple representations in one's teaching repertoire through an accessible phenomenon for which full insight is not obvious from using only the single most common representation. The phenomenon chosen, Simpson's paradox, can be concisely defined as the reversal of a comparison when data are grouped. In this particular example, we will see that it is possible for women to be hired at a higher rate than men within each of two departments but at a *lower* rate than men when the data from both departments are pooled together.

THE RELEVANCE OF SIMPSON'S PARADOX

Simpson's paradox was first noted in 1951 by the British statistician E. H. Simpson but was discussed as early as 1903 by the Scottish statistician George Yule (Wagner 1983). Simpson's paradox can involve a comparison of overall rates, ratios, percentages, proportions, probabilities, averages, or measurements that are weighted averages of subgroup counterparts. Students are likely vulnerable to this paradox if they have the related "averaging the averages" misconception, in which they compute the ordinary average in problems requiring the weighted average. In a weighted average, an overall average is computed by weighting the individual averages by the sizes of their corresponding individual groups. For example, if the average final exam

The author expresses appreciation to Ralph Cain, Lindsey Eck, the 2001 Yearbook editorial panel, and his secondary school methods students.

score in a 30-student class is 100 and the average final exam score in a 10-student class is 60, the overall average for all 40 students is not the "unweighted" mean (100 + 60)/2 = 80 but is obtained as a weighted mean: ((100)(30) + (60)(10))/40 = 90. When some courses are worth more credit hours than others, a student's overall grade-point average is a weighted average as well. A geometrical interpretation of the weighted mean is described by Hoehn (1984) and will be discussed later.

Simpson's paradox was chosen for the investigation for several reasons:

1. It is simple enough to encounter (e.g., table 11.1) with mere fraction arithmetic and yet complex enough to model with tools spanning a broad range of high school mathematics content, generating many different representations.

2. A paradox can motivate students (Movshovitz-Hadar and Hadass 1990; Wilensky 1995; Lesser 1998).

3. Its structure relates to common student misconceptions regarding weighted means or even the addition of fractions. Noting that students are taught (correctly) that the statements

$$a/b > e/f \text{ and } c/d > g/h \text{ imply } a/b + c/d > e/f + g/h,$$

Mitchem (1989) suggests that students who (incorrectly) add fractions by adding the numerators and adding the denominators would assume that $(a + c) / (b + d)$ always exceeds $(e + g) / (f + h)$ and thus be vulnerable to the paradox.

4. It provides many opportunities to explore "both the mathematical and developmental advantages and disadvantages in making selections among the various models" (NCTM 1991, p. 151).

5. It allows "a view of a real-world phenomenon ... through an analytic structure imposed on it" (NCTM 2000, p. 70).

This phenomenon is not contrived: it has actually occurred in many natural situations, including university admission rates (male versus female), fertility rates (rural versus urban), death rates (young versus old), death penalty cases (black versus white), categories of federal tax rates, and various baseball statistics (Bassett 1994; Bickel, Hammel, and O'Connell 1975; Cohen 1986; Moore and McCabe 1993; Wagner 1982a). Simpson's paradox underscores the pitfall of basing a conclusion on only a single average and demonstrates a general need for intuition to be checked against mathematical arguments. Exploring Simpson's paradox may also stimulate greater awareness of what one is averaging over, such as the phenomenon that a university's mean class size averaged over students is never smaller than the mean class size averaged over classes (Hemenway 1982; Movshovitz-Hadar and Webb 1998).

REPRESENTATIONS OF SIMPSON'S PARADOX

What are some ways to think about Simpson's paradox? This article presents many ways and invites the reader to grab pencil and paper and explore them. The reader is also encouraged to reflect on the "relative strengths and weaknesses of various representations for different purposes" (NCTM 2000, p. 70). Which ones seem "new" or applicable to other mathematics content?

Numerical or Tabular Representations

This section begins with a numerical or tabular representation because it is concrete and is the most common representation that textbooks use for Simpson's paradox. Table 11.1 is a 2x2x2 table involving the three categorical variables gender (male or female), department (social sciences or physical sciences), and employment application status (hired or denied). The numbers were chosen for ease of computation and to draw attention to the role of where the larger and smaller cell sizes were located. (Later in the article, table 11.2 offers a similarly behaving data set that is more subtle in appearance.) It is routine to verify that within each department, women are hired at a higher rate than men (since 30/80 = .375 > .25 = 5/20 and 15/20 = .75 > .625 = 50/80), yet are hired at a lower rate than men for the overall situation: (30 + 15)/100 = .45 < .55 = (5 + 50)/100.

TABLE 11.1
Hiring Data (by Gender and Department)

	Social Sciences		Physical Science		Overall	
	Male	Female	Male	Female	Male	Female
Hired	5	30	50	15	55	45
Denied	15	50	30	5	45	55
Total applied	20	80	80	20	100	100

As Wagner (1982b) states, "Because this situation occurs at the level of a purely descriptive data analysis, it can easily bewilder the statistically naïve observer" (p. 47). Indeed, many students have responded with the reaction "I follow the arithmetic, but I still don't believe the result." The numerical representation is undeniably effective in demonstrating *that* Simpson's paradox can happen but limited in offering insight into *how* it can happen. Exploring additional representations can provide insight into the situation that will help resolve this tension. We will keep the underlying numbers the same, however, to keep the primary focus on the representation itself.

The Circle Graph Representation

One alternative representation is the circle graph (Paik 1985) as shown in figure 11.1. Each circle acts as a sort of scatterplot cluster of points representing one of the four gender-department combinations with the y-coordinate of each center at the corresponding hiring rate (.25, .375, .625, or .75). The "weighting" of each hiring rate is reflected in the area of each circle being proportional to the size (20, 80, 80, or 20, respectively) of the applicant pool for that particular gender-department combination. The top right circle and lower left circle are small, each representing a group of 20 people—physical sciences women or social sciences men, respectively. The larger circles represent groups of 80 people—physical sciences men and social sciences women. Since 80 = 4(20), one of the large circles should have quadruple the area (and double the radius) of a smaller circle in figure 11.1. (There is no significance to the absolute length of any particular radius nor to whether the endpoints of the line segment representing the "overall" situation lie inside the circles.)

The highest line segment in the interior of figure 11.1 connects the male and female hiring rates for physical sciences, and the lowest line segment does this for social sciences. Because Females are (arbitrarily) placed to the right of males on the horizontal axis, the positive slopes of these line segments reflect the fact that females have a higher hiring rate than males within either department. The middle line segment connects the male and female hiring rates for the overall university, and we see that now (from its negative slope) it is males who have the higher hiring rate. The descriptions *highest*, *lowest*, and *middle* may not be sufficient to distinguish the three line segments for other data sets, such as if the departmental segments crossed each other (in such a situation, each department would favor a different gender).

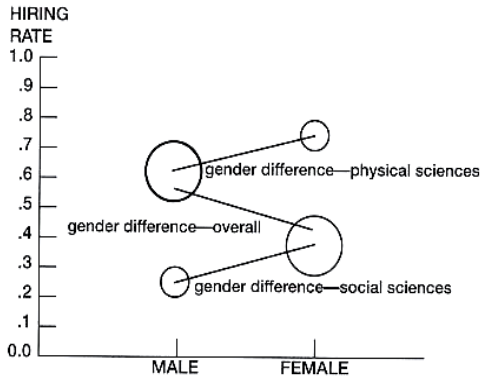

Fig. 11.1. Circle graph representation adapted from Paik (1985)

Although so far we have given an interpretation only to the sign of the slope, we can also interpret its numerical value as the "female minus male" difference in hiring rates if we choose to code the male and female markers on the x-axis as 0 and 1, respectively. With such coding, each department would have a segment with a slope of .125 while the (middle) segment representing all 200 applicants would have a slope of-.l. Coding a qualitative variable in such a "quantitative" manner is mathematically meaningful only

when the qualitative variable is "dichotomous," which means having only two possible values (i.e., male or female). In addition, by similarly coding the dichotomous variable of hiring status (1 = "hired"; 0 = "not hired"), students who know how to calculate a correlation coefficient can verify that the correlation between gender and hiring status within either department is approximately .105 > 0, but for both departments combined is -.1 < 0, which is a connection to statistics showing a further representation of reversal. (The sign of a correlation coefficient is always the same as the sign of the slope of the line of best fit for that same scatterplot.)

When we look at the four circles, most of the "weight" (160 of the 200 applicants) is in the two large circles, whose positions determine a "negative sloping" orientation. The placement of the two smaller circles has a slight effect on this orientation, pulling the middle (overall) line segment somewhat counterclockwise (from the segment that would be determined by the centers of the two large circles alone), somewhat toward the orientation of the top and bottom lines, but not enough to attain a nonnegative slope. Students might try making circle graph sketches that vary the sizes or positions (i.e., heights) of the circles to suggest how all permutations (with replacement) of "positive slope," "negative slope," or "zero slope" could be possible for these three line segments for a new data set.

For example, if we change table 11.1 so that 60 of the 80 male physical sciences applicants and 20 of the 80 female social science applicants were hired, each department would produce a segment with zero slope (meaning males and females were hired at exactly equal rates within each department), but a disparity would still appear in the aggregate. This demonstrates that different "cell sizes" alone can cause the effect.

The Platform Scale Representation

Perhaps the most concrete representation besides the numerical table is the platform scale (see fig. 11.2), as described by Falk and Bar-Hillel(1980,p.l07):

> Suppose a set of uniform blocks arranged in stacks of varying heights is located on a weightless platform, which is balanced on a pivot located at the center of gravity.... One can ... shift the entire construction to the *right*, while simultaneously moving individual blocks to other stacks on their *left*. If done

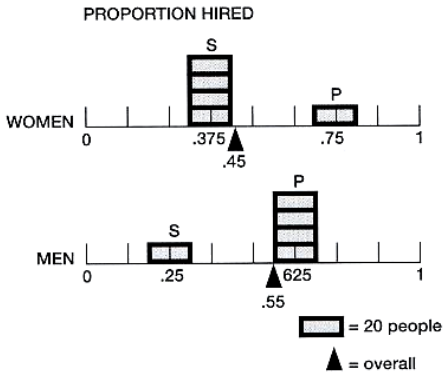

Fig. 11.2. Platform scale representation
Adapted from Falk and Bar-Hillel (1980)

> appropriately, the net result could then be a new center of gravity which is to the left of the old one.

This helps students identify the weights and, we hope, recognize that weighted averages depend on the weights as well as on the values being averaged. The position of each stack represents an average, and the weight of the stack is the weight for that average (in computing the overall weighted average of averages). Students can certainly see with this representation that, for example, the weighted average and unweighted average of two stacks will be the same (i.e., have the same balance point) only if the sizes of the stacks are equal. In other words, if x and y are the stack position values and m and n are the stack weights, then setting $(nx + my)/(n + m)$ equal to $(x + y)/2$ yields $x(n-m) = y(n-m)$, which forces m and n to equal each other for $x \neq y$.

This representation builds naturally on intuition already provided by various textbooks, such as Billstein, Libeskind, and Lott (1993): "We can think of the mean as a balance point, where the total distance on one side of the mean (fulcrum) is the same as the total distance on the other side" (p. 459). Freedman et al. (1991) illustrate "histograms made out of wooden blocks attached to a stiff, weightless board. The histograms balance when supported at the average" (p. 59). This representation could be readily extended to more than two stacks (departments, in this instance) and can be built with physical materials readily available in a typical classroom (e.g., using a ruler or meterstick for the platform). The platform scale representation may be limited, however, to numerical examples in which the weighting numbers (20, 80, 80, and 20) have a convenient greatest common divisor. Nevertheless, the intuition it provides would, it is hoped, give students intuition that could transfer to situations that cannot be as neatly modeled in this particular representation.

The Trapezoidal Representation

Tan (1986) provides a trapezoidal representation of Simpson's paradox (see fig. 11.3) that is built only on the observation (Hoehn 1984) that "[t]he length of any line segment which is parallel to the two bases and has its endpoints on the nonparallel sides of a trapezoid is the weighted mean of the lengths of the two bases" (p. 135). Specifically,

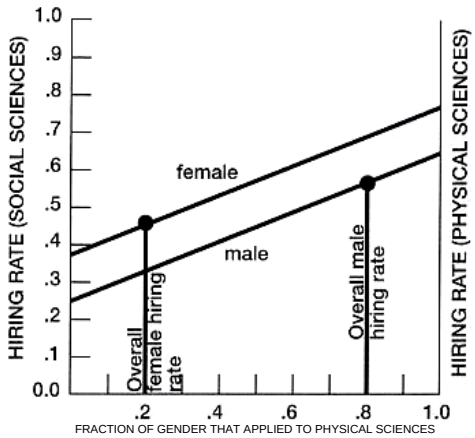

Fig. 11.3. Trapezoidal representation adapted from Tan (1986)

each base is weighted by the proportion of the trapezoid's height traveled toward that base to reach the weighted mean segment. For example, the "female" trapezoid is determined by the points (0, 0), (0, .375), (1.0, 0), and (1.0, .75) and has bases of length .375 and .75, which have corresponding weights .8 and .2, respectively. We may conjecture that when the weights are each .5, the weighted average segment and the unweighted average segment coincide (in what geometry students would call the median of the trapezoid). This relationship can be verified algebraically by setting the usual formula for the area of the overall trapezoid equal to the sum of the areas of the two smaller trapezoids formed by the new segment, and then recognizing that we now have exactly the same equation we encountered when discussing the weighted and unweighted averages in the context of the platform scale representation

Applying this to our university employment example, we find that each gender would have a trapezoid in which the two vertical bases represent that gender's hiring rates in the two departments. (This orientation is sideways from the more common depiction of a trapezoid, and the reason for this choice will soon be clear.) The trapezoids have one leg in common—the segment of the horizontal axis, which allows tracking the department "weights." The fact that females are hired at a higher rate than males in each department is clear by noting the vertical heights of the endpoints of the top male and female trapezoid legs (legs that happen to be parallel to each other but that do not have to be in general). However, because the genders had different proportions of applicants applying to the physical sciences department, it turned out that the overall male hiring rate (the large dot formed by extending the male departmental application proportion out to the top leg of the "male trapezoid") was higher (which is easier to see with the "sideways" orientation) than the large dot representing the overall female hiring rate.

Furthermore, it is straightforward to see with this representation when this reversal does and does not happen. For example, if the hiring rates for the four gender-department combinations were unchanged but the proportion of females who applied to the physical sciences became .8 (matching the proportion for males), then the dot representing the overall female hiring rate would slide along its line and clearly be higher than the dot representing the overall male hiring rate. Actually, we can solve the linear inequality $.75x + (.375)(1 - x) > .55$ to see that greater than 7/15 would suffice. This approach could also make it clear that when each gender has a 50-50 split of applications between departments, a reversal cannot occur. Related applications of this representation are made by Witmer (1992).

On the positive side, this model can be constructed physically (with a clip or bead sliding along a string strung between two vertical poles) or computationally (in interactive geometry software). These models would not be as limited to the data being "nice" as was required by the platform scale model. On the negative side, some students may be initially overwhelmed with the number of features of the graph, especially its having three axes.

The Unit Square Representation

Another type of geometric representation is adapted from the unit square model of Bea and Scholz (1994), who originally used it to represent conditional probabilities. Comparing shaded proportions of side-by-side rectangles (of equal length within each square), figure 11.4 shows at a glance that in each individual department the gender with the greater fraction hired is women and yet overall it is men. To understand better the representation, let's explain how the physical science unit square was constructed. Beginning with a square, draw a vertical line segment that partitions the square into rectan-

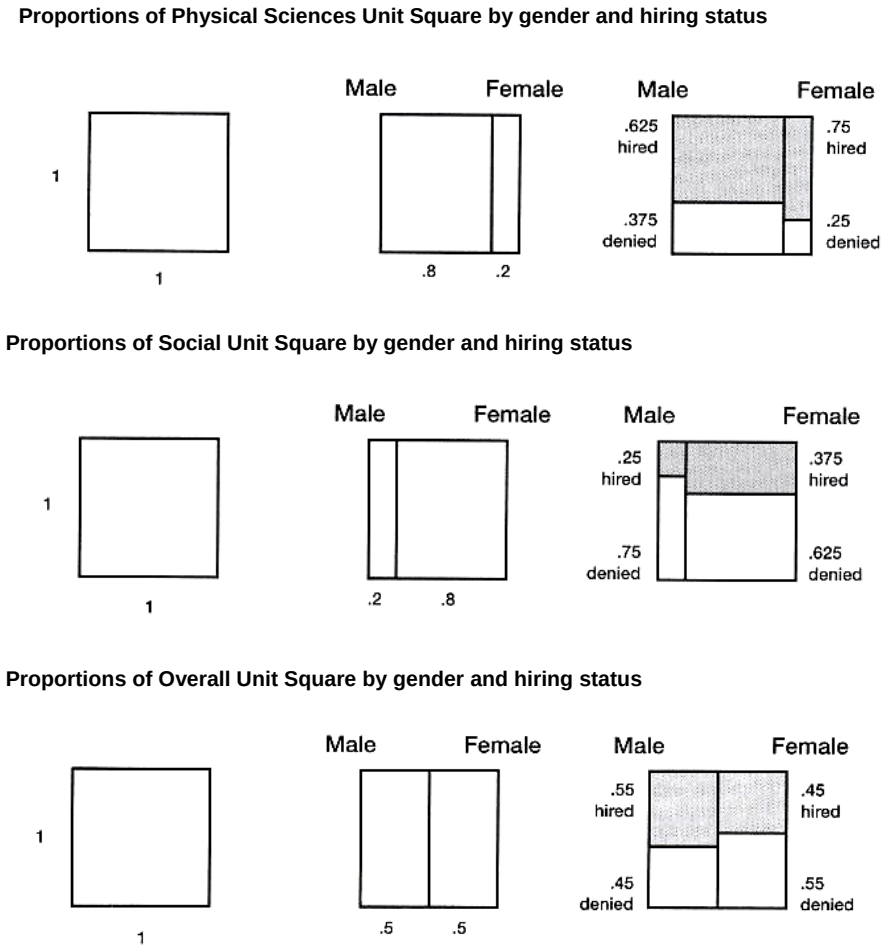

Fig. 11.4. Unit square representation of table 11.1 data adapted from Bea and Scholz(1994)

gles whose areas (and shorter sides) are proportional to the respective numbers of men and women (i.e., the vertical line would be 80/(20 + 80) of the distance from the left side to the right side). Now shade the fraction of each rectangle that represents the fraction of the corresponding gender that was hired (i.e., we shade 5/8 of the men's rectangle and 6/8 of the women's). Because the rectangles have the same height and because shading is done in the same direction, we see at a glance that the fraction of women hired was greater because its shaded region extends further down. Students who focus on the absolute amount of shaded area of each gender rather than the proportion of each rectangle that is shaded are simply noticing that within physical sciences, males had a greater number hired than females, but the *rate* at which men were hired is still less than the rate women were hired. This potential confusion (which would not occur in the other two unit square diarams) offers a good opportunity to distinguish between amounts and rates.

Other Representations

There is also a probability representation (e.g., Movshovitz-Hadar and Webb 1998; Mitchem 1989) of this paradox that can be physically represented (and empirically simulated) in the classroom with four boxes and two colors of balls. Basically, the distribution of objects in the boxes is chosen so that the probabilities of drawing (with replacement) a certain color from box A or from box C are less than the probabilities of drawing that color from box B or from box D, respectively, but the inequality direction is reversed when the probability for the combined contents of boxes A and C is compared to the probability for the combined contents of B and D. Using the data from table 11.1, this would mean box A (labeled "social science males") would have 5 red chips and 15 green chips, box B ("social science females") would have 30 red chips and 50 green chips, box C ("physical science males") with 50 red chips and 30 green chips, and box D ("physical science females") having 15 red chips and 5 green chips. The phenomenon of combining data becomes very literal with this representation. Mitchem (1989) uses smaller numbers and a picture to shed further light on the dynamic.

Lord (1990) offers a representation involving determinants. Ignoring row or column totals in table 11.1, we can break the table into three 2 x 2 matrices, each of the following form:

$$\begin{bmatrix} \text{males hired} & \text{females hired} \\ \text{males denied} & \text{females denied} \end{bmatrix}$$

Recall that the determinant of the 2x2 matrix

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is *ad* - *bc*. Students may find it interesting to verify and investigate the fact that the determinant (55)(55) - (45)(45) = 1000 of the overall university matrix has a different sign than the determinant (-200) of either of the matrices representing an individual department. The reversal of determinant sign is not a coincidence. Students can verify that the male hiring rate being higher than the female hiring rate can be expressed as the inequality

$$\frac{a}{a+c} > \frac{b}{b+d}$$

Algebraic transformations produce an equivalent inequality of $ad - bc > 0$, which corresponds precisely to a positive determinant for a 2 x 2 matrix! These manipulations also show that a determinant *ad* - *bc* equal to 0 corresponds to the female hiring rate and male hiring rate being equal. Intuitively, then, the sign of the determinant tells us which gender has it better, and the larger the absolute value of the determinant, the greater the evidence of a statistical relationship or interaction between gender and employment status.

A more advanced representation uses vector geometry of the plane. Lord (1990, p. 55) demonstrates that "the following at-first-sight-plausible statement about complex numbers is, in fact, false: If arg (z_2) > arg (z_1) > 0 and arg (z_2') > arg(z_1') > 0, then arg $(z_2 + z_2')$ > arg $(z_1 + z_1')$." This translates into a corresponding statement about slopes of vectors. Trying to make a diagram (such as fig. 11.5) that shows a specific counterexample is an interesting

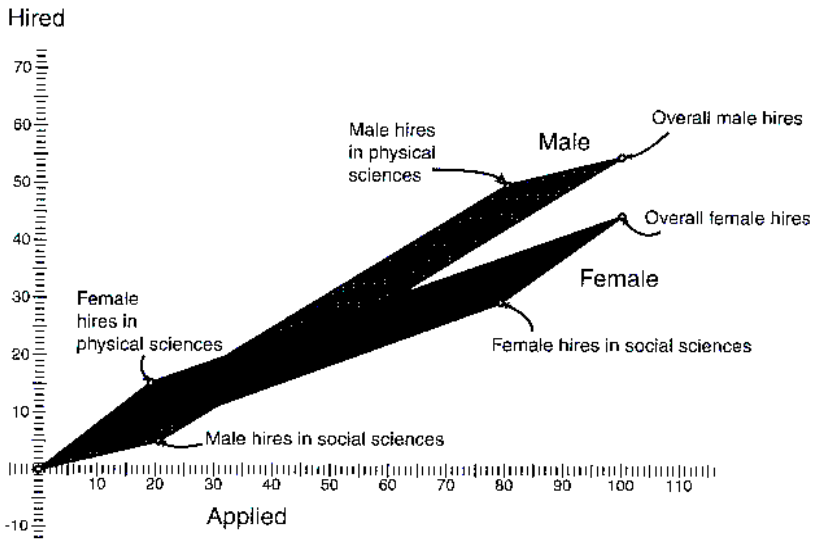

Fig. 11.5. A complex numbers representation of table 11.1

exercise. To facilitate the interpretation of our figure 11.5, let us agree that the ordered pair (*a, b*) now represents (number applied, number hired), so that slope corresponds to hiring rate. The physical science male vector goes from the origin to (80, 50). whereas the physical science female vector goes from the origin to (20, 15). The slope of the line segment or the angle relative to the horizontal axis made by the female vector is larger, thus showing that in the physical sciences, females have a better hiring rate. Similarly, the social science female vector has a slope larger than the corresponding male vector. But the overall male vector has a larger slope than the overall female vector.

AN EXPLORATION WITH PRESERVICE TEACHERS

In the fall 1998 semester at a midsized, state-supported university, a class period of the secondary school mathematics methods course was devoted to exploring how multiple representations of a particular phenomenon served as tools for preservice secondary school teachers' thinking. Connections between mathematics content and pedagogy (e.g., using multiple representations) were one of several ways the author reformed this course as part of the National Science Foundation-funded Rocky Mountain Teacher Education Collaborative (Lesser 1999). All preservice teachers enrolled in the class participated, but the size of the class (seven) and time available for the exploration yielded anecdotal observations rather than definitive inferences.

Although all the students had had upper division coursework in probability and statistics, three of the seven initially answered that it was not possible "women could be hired at a higher rate than men within each of the two divisions, but still be hired at a lower rate than men for the university as a whole." To make sure that all students realized that, yes, it was possible, they were then shown table 11.1 and asked to verify by themselves the hiring rates for males and females within each department as well as for the overall university.

During the period, they had time to be exposed to seven of the representations previously mentioned (each with a written explanation of "how to read it" somewhat less detailed than is given in this article). Their comments made it clear that they had never seen most of the representations before in a textbook or class used to analyze any phenomenon, let alone Simpson's paradox in particular. There was a strong tendency among the students to say that although they might try the unit square or platform scale representations, they would most likely rely on the numerical representation in discussing the paradox with any future students. There are many possible explanations for students' reliance on the tabular (numerical) representation, including (1) it was the first one they saw, (2) it is the one most familiar to them from their mathematics classes, (3) it is the most concrete, (4) it is

the one they could feel most confident teaching with, (5) it makes the most clear demonstration that the paradox occurs, and (6) they lacked the time to absorb others fully. The author believed the last explanation was certainly an issue. In fact, the next time he taught a secondary school methods course, a class period happened to be 2.5 times as long, and the students expressed a consensus that some of the physical and geometric representations were clearly superior to the table.

Biff, a preservice secondary school teacher in the author's methods course, addressed the tradeoffs of using multiple representations in general. On the one hand, he states: "[Multiple representations] can reach different kids with different approaches, and reinforce the learning—each representation is in fact teaching something new." The last part of this statement seems to suggest the recognition that a new representation is not just passively delivering the same piece of content but giving a new angle that may itself contain content (or even a transferable tool). Piez and Voxman (1997) believe that activities using multiple representations lead to more thorough understanding "[b]ecause each representation emphasizes and suppresses various aspects of a concept" (p. 164) and expect that "students will gain the flexibility necessary to work with a wide range of problems using an appropriate representation. In our work with students with weak mathematical skills, we have seen definite improvement" (p. 165).

However, Biff also expressed a pitfall: "Very time-consuming in the end. More sophisticated solutions will leave the slower in the dust." Going through all possible representations for each piece of content certainly would make it hard to stay on pace with a packed curriculum, but knowing that there may be far more than three representations possible makes it more likely to access the most useful one for the situation.

ISSUES AND IDEAS FOR THE CLASSROOM

The specific numbers chosen could be tailored to the audience. For example, there is more subtlety and "realistic appearance" in the data in table 11.2 (as opposed to table 11.1, whose entries are all multiples of 5 and included "swapped numbers").

TABLE 11.2

Slightly Revised Set of Hiring Data (by Gender and Department)

	Social Science		Physical Science	
	Male	Female	Male	Female
Hired	4	24	48	14
Denied	16	56	32	6

Constructing a data set with smaller numbers may make Simpson's paradox accessible even to students in the upper elementary grades and could even lend itself to being kinesthetically modeled by the typical number of students in a classroom by having students stand in marked-off regions of two 2x2 tables on the floor and then physically combine into a single 2x2 table. For example, a 19-student class could arrange themselves into two 2 x 2 tables as follows (1/3 < 3/8, 3/5 < 2/3, but 4/8 > 5/11):

1	3	3	2
2	5	2	1

If we admit the possibility of cells being 0, then Simpson's paradox can actually be physically modeled with as few as 9 students (0/1 < 1/4, 2/3 < 1/1, but 2/4 > 2/5):

0	1	2	1
1	3	1	0

The numerical representation is certainly the easiest way to introduce the phenomenon and can be presented in a very accessible manner through a structured sequence of questions (e.g., Smith 1996, p. 188) or as a story problem (e.g., Movshovitz-Hadar and Webb 1998, p. 113). In general, the way a particular representation is introduced may affect a student's ability to use or apply flexibly that representation. For example, in a project by McFarlane et al. (1995, pp. 476-77),

> children in the experimental classes were introduced to line graphs not as a Cartesian plot, where the ability to correctly identify positions on a grid was the objective; rather, they were introduced to graphs as a representation of the relationship between two variables.... Their ability to read and interpret temperature/time graphs was greatly enhanced as a result and it is particularly significant that their ability to sketch temperature time curves to predict the behavior of a novel system also improved.

Robust examples such as Simpson's paradox that have the potential to expand the repertoire of representations available may be especially valuable in the early part of the year to get students (and teacher!) primed to look for multiple ways of representing all future phenomena encountered, and to get them shaken out of habits they may have to fixate on nothing but one feature of some familiar graphical representation (see Berenson, Friel, and Bright 1993). More research in this area would be useful.

Some representations are certainly more suited toward technology representations than others. The numbers in the tables (as well as commands for the functions involved in any of the nonpictorial representations) could certainly be entered into a spreadsheet, for example, and the effect of changing various numbers instantly apparent. Shaughnessy, Garfield, and Greer (1996) list having a choice of dynamic representations for interpreting and displaying a data set to be an important attribute of a technological environment to facilitate the learning of data handling.

Simpson's paradox is rarely experienced by students in any of their courses, and if it were, it would likely be in numerical form only. Even a popular introductory college statistics textbook (Moore and McCabe 1993) that is quite "Simpson's paradox-oriented" (by involving the paradox in the only three-way table example in the text as well as in every three-way table exercise following that section) does not offer a nonnumerical representation of it. In a case study of a preservice secondary school teacher in a course integrating content and pedagogy, Wilson (1994) found that being able to translate between multiple representations was deeply related to conceptual understanding, a finding that supports Held (1988) and NCTM (2000). Therefore, a deep understanding of how Simpson's paradox can occur seems difficult without the aid of representations beyond only a numerical one.

Perhaps exposing students and even teachers to rich representations of reversal representations will create a "reversal" of some of their attitudes, such as the common perception that representations are limited to context-free discussions of functions and are limited to a "rule of three" that they have already seen before! And maybe it will also reverse a perception of representations as some checklist of unrelated items to go through rather than as a dynamic source of new insights, connections, and tools for thinking whose roles should even further expand throughout the new century.

EXTENSIONS

As a follow-up assessment, students can be given the following data (table 11.3) and asked for a quick "gut" answer to the question "Is it possible that overall mean female salary is less than overall mean male salary even though mean female salaries are higher within each category?" This example may be a more subtle manifestation of the paradox in that the largest cell sizes for men and women are in the same category (i.e., support staff) this time, unlike table 11.1 (where 80 and 20 noticeably "swap roles" between departments). A "weighted average" computation shows the overall male and female salaries here are approximately $41000 and $37000, respectively.

TABLE 11.3
Annual Salary Data

	Men	Women
Support staff employees	70 males (their mean salary is $20000)	90 females (their mean salary is $30000)
Executive-level employees	30 males (their mean salary is $90 000)	10 females (their mean salary is $100000)

Although Simpson's paradox may be new to students, they should be reminded that they have certainly seen inequality reversals earlier in their mathematical career, and they might now look for a representation that yields insight into those examples beyond "rules to memorize." For example, $2 < 4$ and yet $1/2 > 1/4$. An elementary school class might represent this with fair divisions of pizza, whereas a high school algebra class might consider the decreasing property of (either half of) the graph of $f(x) = 1/x$. Another "simple" example of reversal that students might look for a way to illuminate is why $-2 > -4$, since $2 < 4$.

For a final challenge, classes may look for a representation that indicates if it is possible to have the "double Simpson's paradox" posed by Friedlander and Wagon (1993, p. 268):

> It is possible for there to be two batters, Veteran and Youngster, and two pitchers, Righty and Lefty, such that Veteran's batting average against Righty is better than Youngster's average against Righty, and Veteran's batting average against Lefty is better than Youngster's average against Lefty, but yet Youngster's combined batting average against the two pitchers is better than Veteran's. ... [I] s it possible to have the situation just described [which would indeed be a feasible "single" Simpson's paradox] and, at the same time, have it be the case that Righty is a better pitcher than Lefty against either batter, but Lefty is a better pitcher than Righty against both batters combined?

If we adapted this to the employment context we have been working with throughout the article, the possibility of a "second" Simpson's paradox added to our scenario would correspond to asking, "Is it also possible that the social sciences department has a lower hiring rate than the physical sciences department for either gender and yet the physical sciences department has a lower hiring rate for both genders combined?"

REFERENCES

Bassett, Gilbert W. "Learning about Simpson's Paradox." Stats 12 (Fall 1994): 13-17.

Bea, Wolfgang, and Roland W. Scholz. "The Success of Graphic Models to Visualize Conditional Probabilities." Paper presented at International Conference on Teaching Statistics 4, Marrakesh, Morocco, July 1994.

Berenson, Sarah B., Susan Friel, and George Bright. "The Development of Elementary Teachers' Statistical Concepts in Relation to Graphical Representations." In *Proceedings of the 15th Annual Meeting of the International Group for the Psychology of Mathematics Education, North American Chapter,* Vol. 1, edited by Joanne Rossi Becker and Barbara I. Pence, pp. 285-91. Pacific Grove, Calif.: International Group for the Psychology of Mathematics Education, 1993. ERIC Document ED372917.

Bickel, Peter]., Eugene A. Hammel, and]. W. O'Connell. "Sex Bias in Graduate Admissions: Data from Berkeley." *Science* 187 (February 1975): 398-404.

Billstein, Rick, Shiomo Libeskind, and Johnny W. Lott. *A Problem-Solving Approach to Mathematics for Elementary School Teachers.* Reading, Mass.: Addison-Wesley, 1993.

Cohen, Joel E. "An Uncertainty Principle in Demography and the Unisex Issue." *American Statistician* 40 (February 1986): 32-39.

Falk, Ruma, and Maya Bar-Hillel. "Magic Possibilities of the Weighted Average." *Mathematics Magazine* 53 (March 1980): 106-7.

Freedman, David, Robert Pisani, Roger Purves, and Ani Adhikari. *Statistics*. 2nd ed. New York: W. W. Norton, 1991.

Friedlander, Richard, and Stan Wagon. "Double Simpson's Paradox." *Mathematics Magazine* 66 (October 1993): 268.

Heid, M. Kathleen. "Resequencing Skills and Concepts in Applied Calculus Using the Computer as a Tool." *Journal for Research in Mathematics Education* 19 (January 1988): 3-25.

Hemenway, David. "Why Your Classes Are Larger than 'Average.'" *Mathematics Magazine* 55 (May 1982): 162-64.

Hoehn, Larry. "A Geometrical Interpretation of the Weighted Mean" *College Mathematics Journal* 15 (March 1984): 135-39.

Lesser, Larry. "Countering Indifference Using Counterintuitive Examples." *Teaching Statistics* 20 (February 1998): 10-12.

Lesser, Lawrence Mark. "Investigating the Role of Standards-Based Education in a Pre-Service Secondary Math Methods Course." *In Promoting Excellence in Teacher Preparation: Undergraduate Reforms in Mathematics and Science*, edited by Myra L. Powers and Nancy K. Hartley, pp. 53-64. Fort Collins. Colo.: Colorado State University, 1999.

Lord, Nick. "From Vectors to Reversal Paradoxes." *Mathematical Gazette* 74 (March 1990): 55-58.

McFarlane, Angela, Yael Friedler, Paul Warwick, and Roland Chaplain. "Developing an Understanding of the Meaning of Line Graphs in Primary Science Investigations, Using Portable Computers and Data Logging Software." *Journal of Computers in Mathematics and Science Teaching* 14 (Winter 1995): 461-80.

Mitchem, John. "Paradoxes in Averages." *Mathematics Teacher* 82 (April 1989): 250-53.

Moore, David S., and George P. McCabe. *Introduction to the Practice of Statistics.* 2nd ed. New York: W. H. Freeman, 1993.

Movshovitz-Hadar, Nitsa, and Rina Hadass. "Preservice Education of Math Teachers Using Paradoxes." *Educational Studies in Mathematics* 21 (June 1990): 265-87.

Movshovitz-Hadar, Nitsa, and John Webb. *One Equals Zero and Other Mathematical Surprises*. Berkeley, Calif.: Key Curriculum Press, 1998.

National Council of Teachers of Mathematics (NCTM). *Professional Standards for Teaching Mathematics*. Reston, Va.: NCTM, 1991.

————. *Principles and Standards for School Mathematics.* Reston, Va.: NCTM, 2000.

Paik, Minja. "A Graphic Representation ofaThree-Way Contingency Table: Simpson's Paradox and Correlation." *American Statistician* 39 (February 1985): 53-54.

Piez, Cynthia M., and Mary H. Voxman. "Multiple Representations: Using Different Perspectives to Form a Clearer Picture." *Mathematics Teacher* 90 (February 1997): 164-66.

Shaughnessy, J. Michael, Joan Garfield, and Brian Greer. "Data Handling." In International Handbook of *Mathematics Education,* Part I, edited by Alan J. Bishop et. al., pp. 205-38. Dordrecht, Netherlands: Kluwer, 1996.

Smith, Sanderson. *Agnesi to Zeno*. Berkeley, Calif.: Key Curriculum Press, 1996.

Tan, Arjun. "A Geometric Interpretation ofSimpson's Paradox." *College Mathematics Journal* 17 (September 1986): 340-41.

Wagner, Clifford H. "Simpson's Paradox." *Journal of Undergraduate Mathematics and Its Applicatio*ns 3 (1982a): 185-98.

————. "Simpson's Paradox in Real Life." *American Statistician* 36 (February 1982b): 46-48.

————. "Simpson's Paradox." UMAP: Modules and Monographs in Undergraduate Mathematics and Its Applications Project, Unit 587. Newton, Mass.: Education Development Center. UMAP, 1983.

Wilensky, Uri. "Paradox, Programming and Learning Probability: A Case Study in a Connected Mathematics Framework." *Journal of Mathematical Behavior* 14 (1995): 253-80.

Wilson, Melvin. "One Preservice Secondary Teacher's Understanding of Function: The Impact of a Course Integrating Mathematical Content and Pedagogy." *Journal for Research in Mathematics Education* 25 (July 1994): 346-70.

Witmer, Jeffrey A. *Data Analysis*: An Introduction. Englewood Cliffs, N.J.: Prentice Hall, 1992.