May 31, 2025

Here is a piece called **"In my classes, usually the men(women) have more sisters(brothers) than women(men) do: a Mosteller statement that continues to intrigue"** that I submitted to The American Statistician. The two reviewers were of opposite opinions, and the editor recommended I submit it elsewhere.

Below you can read

- the cover letter,
- The (unblinded) submitted version,
- the reviews.

This was one of the most challenging investigations I ever carried out, and I found myself going back and forth several times between believing Mosteller and believing Falk. I also had some very definite and opposite reactions from colleagues and others I consulted. The strong reactions reminded me of the ones I usually heard when people first encountered the Monty Hall problem.

I believe this should be settled by lots of **real-world data** – just as Mosteller himself advocated – and not by beliefs, or by invoking simplistic and untested statistical models. So, I am hoping that teachers who teach large classes will be able to carry out the simple class survey Mosteller suggested, and settle this. Given that the expected difference is close to zero, several teachers might need pool their data in order to have sufficient 'resolving power'.

(I undertook this in 2023, my last year of teaching, amid a somewhat adversarial climate on the uses of sex/gender (pro)nouns, and advice from some colleagues that I not carry out the anonymous class survey)

As we warned in connection with the Poisson model[1], teachers might use this problem as an example where the binomial model may be inappropriate: humans can often subvert Nature's laws. And, today more than ever, as one consultant put it, we need to think hard about how we came to see the data we did see, and what data we did not see! [2]

I would be happy to communicate with any teachers who would to see this 'settled' by data, rather than by the opinions of two referees.

Sincerely,
James Hanley

webpage: https://jhanley.biostat.mcgill.ca

email: `james.hanley@mcgill.ca`

[1] Hanley and Bhatnagar, The American Statistician, 2022

[2] Just as (see p. 198) how someone noticed noticed something 'unusual' in USA Today.

May 22, 2023

Editor
The American Statistician

Dear Editor

I am submitting this manuscript "*In my classes, usually the men(women) have more sisters(brothers) than women(men) do: a Mosteller statement that continues to intrigue*" as a possible item for the **Teacher's Corner** of The American Statistician.

Mosteller's statement, made at the beginning of an article on 'teaching techniques' in TAS in 1980, was challenged by a 'probability-expert' two years later, and from what I have been able to discern, that refutation had not been challenged – nor have similar statements made by authorative 'experts' (and statisticians) 100 years ago.

Trying to understand what Mosteller used (besides data!) to back up his claim makes for what I think is some very nice teaching material. In this note, I tell it somewhat personally, warts and all, to show that it can easily trap teachers and students, and that there are more subtleties than first meet the eye.

[In the end it took a combination of two factors. As my consultants pointed out, I missed the more obvious one, but that one alone did not seem to be enough. I had to go back and combine it with a factor I initially thought was the culprit: but I won't say more here, or I might give away the ending.]

The increasing use of Big (and small) Data means that we need to be even more aware of where these data come from, and what filtering they were subjected to along the way. This note uses Mosteller's "simple (and innocent) situation" to emphasize this additional point.

I look forward to, and will welcome, your and your referees' reactions to this piece.

Sincerely

James Hanley

"In my classes, usually the men(women) have more sisters(brothers) than women(men) do": a Mosteller statement that continues to intrigue

2023.05.22 submission to The American Statistician: Teacher's Corner

Correspondence

Abstract

The title is adapted from a Frederick Mosteller statement in an article in The American Statistician in 1980. In 1982, in *Teaching Statistics*, it was refuted by a psychology professor who has written extensively on students' troubles with probabilistic and statistical reasoning; she argued that if the probabilities of male and female births were equal, differences would be symmetric. Class data I collected in 2001 half-corroborated Mosteller's statement. I recount how I went back and forth between believing Mosteller and believing his critic, and, following consultations, whose expectation I settled on. Along the way, I thought I had figured out what Mosteller was relying on, only to realize that I, like others who have been fooled by another teaser, was using incorrect statistical reasoning. After collecting more data in 2023, I began to doubt Mosteller's statement, but decided to ask some of his collaborators what other forces might be behind the data he had encountered. Their answers, when combined with ideas I had initially discarded, provide one possible explanation. The exercise illustrates the differences between random variables in statistics courses and random variables in the real world. As Mosteller promised, such class data continue to be 'intriguing' and engaging.

KEYWORDS: Class data; conundrum; history; provenance; consultations.

1. INTRODUCTION

Frederick Mosteller's presentation in a session on Statistical Education at the JSM in 1978 was published under the rubric Teaching of Statistics in the 1980 issue of The American Statistician (Mosteller 1980). In it, he discussed "some features of elementary and intermediate courses in applied statistics," and brought out "a few ideas that [he had] found useful in teaching."

In the first section ("Large-scale application") he listed resources[1] that "make it much easier for the teaching statistician to open a lecture with a few remarks about some real-world problem that uses the specific method to be treated in the day's lecture." In the second ("Physical Application"), he began with this principle and this very specific application of it:

> Gathering relevant data in class and then analyzing them with the help of the class gets the students involved because we are analyzing their data and they can contribute to the analysis.
>
> One way to obtain a set of data that can be intriguing is to collect for males and for females, separately, the number of brothers and the number of sisters each student has. One has to specify the degree of relationship, or the matter can get very complicated. I stick to brothers and sisters who have the same biological father and mother as the respondent. *Usually* the men have more sisters than the women have, and the women have more brothers than the men. Sorting out why this is can be instructive. [italics added]

Writing in the journal *Teaching Statistics*, Ruma Falk (1982), a psychology professor who has written extensively on why students have trouble with probabilistic and statistical reasoning, began her article with the second paragraph of Mosteller's 'class demonstration', and then continued

[1] *Statistics: A Guide to the Unknown* (Tanur et al. 1978) and *Statistics by Example* (Mosteller et al. 1973).

> The question of what to expect of data collected as above [i.e., as Mosteller specified] can be posed to beginners in an introductory probability course. Intuitively, it may seem that, because "on the average" families have an equal number of sons and daughters, picking male students for questioning will result in an excess of sisters over brothers, since each respondent does not consider himself. The converse would seem to be true for female subjects. This, indeed, was the opinion of 25 out of 45 students in my class of "Introduction to Probability" during our first session at the Hebrew University of Jerusalem. Only 12 students expected the men and the women to have equal numbers of siblings of the two sexes.

Basing her reasoning on the "independence of the sexes in different births, whether those of siblings or others," Falk then indicated that the 12 students were correct, and appeared to imply that Mosteller was not:

> Thus, the expected outcome of the experiment suggested by Mosteller would be men and women having equal numbers of brothers and sisters. Obviously, the men may turn up with more sisters than the women, and the women may have more brothers than the men, in some classes. However, deviations from equality should be symmetrical with respect to the sex of the siblings and the sex of the respondent.

I first encountered Falk's article in early 2023, and then read her other writings on probability traps. Not finding any rebuttals of her statement,[2] I was inclined to side with her, and was embarrassed that I had fallen into the same reasoning trap that the majority of her students fell into. I wondered how many of today's teachers and how many of their students would also fall for her clever way of enticing her students into it.

But I could not imagine that Mosteller, of all people, could be deceived by faulty intuition. I had seen him up close for three years as my department chair; I knew that he had lectured live on television to lay audiences in the Continental Classroom in the spring of 1961; among the many books he (co)authored[3] was one called "50 challenging problems in probability"; even

[2] Her final paragraph (where she suggested readers might like to consider the effect of using the fact that the overall P(boy) = 0.51 and the possibility that it may vary) may have been in response to reviewers.

[3] On the occasion of Fred's 70th birthday in 1986, Donald Berwick celebrated Fred and

today he is still referred to as the esteemed "Dean" of American statisticians. I could not imagine Fred would promise a "*Usually*" without having empirical evidence to back it up.

Close collaborators of his attest to this. Back in the early 1980s, a colleague who worked extensively with Mosteller on the possible harms of Red Dye 40 used to refer to him euphemistically as "fast Fred" for the extensive feedback he solicited, the many drafts, and the care he took with every publication/report he produced. Another (see below) recently told me "I spent a great deal of time with Fred in the 1980s and 1990s, and a good portion of it was one-on-one. I remember Fred as being an exceedingly careful thinker. Unlike many of us, he would not make a claim without having examined the issue with great care and understanding it thoroughly. He never rushed to a conclusion or position. He would often ruminate on a question for a very long time..." Yet another wrote me that "Before engaging the class in that part of the exercise, I cannot imagine that Fred would not have worked out an answer."

In the remainder of this note, I recount how I went back and forth between believing the statement of Mosteller and that of Falk. Along the way, I *thought* I had figured out what Mosteller may have been relying on. However, I collected more class data, and they supported Falk. At one point, I was

his collaborative proclivities in a poem that begins: "On a high and secret mountain on a South Pacific isle; Lived a hermit in a mud house in a most reclusive style; He had not clothes nor money, neither dishes nor a bed; And he had never even written one short monograph with Fred."

so unsure that I wrote to several people who knew Mosteller well, and asked them what insights he might have had relied on. I show a slice of a large US dataset – assembled to address a related problem – that partially supports their insights. But it took of combination of their suggestion and another (pre-consultation) idea I had discarded, to arrive at a possible resolution.

2. IN GOD WE TRUST; ALL OTHERS MUST BRING DATA

Curiously, Falk only presented class opinions and probability theory, but does not appear to have heeded Deming or asked the students how many brothers and sisters they actually had. On reading her article, I remembered one class dataset I had collected 20 years ago; I was also able to examine and to simulate datasets from an era when families were much larger.

2.1 Class data from 2001

When teaching large introductory statistics courses to graduate students in the health sciences, I often asked the students to provide data about themselves. Although questions 5, 11 and 12 in this 2001 version were not directed at this particular topic, we can re-arrange the answers to them into the type of dataset that Mosteller suggested. The frequencies are shown in Table 1.

				M S,B **F**				
				1 0,0 **6**				
			3 1,0 **7**		5 0,1 **14**			
		1 2,0 **1**		1 1,1 **3**		1 0,2 **5**		
	- 3,0 **-**		4 2,1 **-**		- 1,2 **4**		2 0,3 **-**	
- 4,0 -		- 3,1 -		2 2,2 -		1 1,3 **2**		- 0,4 -

*Table 1: Numbers of Sisters (S) and Brothers (B) reported by 42 **F**emale (smaller, bold font) and 22 Male (smaller, plain font) students first year students in graduate programs in *** in 2001.*

In accord with Mosteller's promise, the men did have more sisters (mean 20/22 = 0.9) than the women did (mean 18/42 = 0.4). But the women did not have more brothers than the men did (means of 41/42 = 1.0 and 29/22 = 1.3 respectively). Clearly, more data would be required to settle the issue.

2.2: Male-respondent-only data from 1921

Interestingly, Falk was not the first psychology researcher to question the statistical reasoning of a then eminent (but nowadays less respected) statistician on this very topic of sisters and brothers. One hundred years ago, the psychologist Cattell (1921) examined family information collected from those listed in "The Biographical Directory of American Men[4] of Science." It consisted of the thousand leading scientific men of the United States who had earlier been selected and arranged in the order of the merit of their work. As is seen in Figure 1, Table XVI listed, for each size of family, the numbers of sisters and brothers reported by 832 of these men.

The data in this table do not allow us to compute either of the two 'men-respondents vs. women respondents' contrasts that Mosteller suggested. But they do fit the "picking male students for questioning" scenario that Falk lured her students into. Ironically, and contrary to the correction Falk supplied to her 24 students, these century-old data show a substantial "excess of sisters over brothers" (1705 over 1527).

[4] The first woman was elected to the US National Academy of Science in 1925. In 1974, the year Mosteller was elected, 2 of the 107 newly elected members were women; in 2022, 60 of the 149 were.

Size of Family	No. of Families	Family of Parents Sibs	Sisters	Brothers
(1)				
2	126	126	69	57
3	128	256	123	133
4	152	456	244	212
5	125	500	266	234
6	123	615	327	288
7	71	426	235	191
8	48	336	172	164
9	31	248	121	127
10	16	144	83	61
11	7	70	34	36
12	5	55	.31	24
Total	832	3232	1705	1527

Fig 1: Numbers of sisters and brothers of 832 American Men of Science (Cattell 1921)

While Cattell remarked that this “disparity may at first strike the reader as inexplicable,” he cited a similar statement/reasoning in Francis Galton’s study of “Hereditary Genius,” based on judges of England between 1660 and 1865, “a group peculiarly well adapted to” Galton’s topic. Galton had noted “I also found the (adult) families to consist on an average of not less than 2.5 sons and 2.5 daughters each. Consequently each judge has on an average of 1.5 brothers and 2.5 sisters.”

But then, just like Falk, Cattell disagrees with the majority opinion

> Nearly all those whom I have questioned about this statement think that it is correct. It seems to most people obvious that if there are equal numbers of boys and girls, a boy must on the average have one more sister than brother. However, a boy has as many brothers as sisters, owing to the sex composition of families. Thus in families of two, one fourth of the families will consist of two boys, one fourth of two girls, and one half of a boy and a girl. On the average, four boys will have among them two brothers and two sisters, and there is a similar equality for large families.

Writing in *Science* a few years later, mathematical statistician Henry Rietz of the State University of Iowa, who himself appears in that Directory, and was a member of the American Statistical Association and the Royal Statistical Society, followed up on Cattell's remarks (Rietz 1924). He re-iterated that "a scientist with the keen statistical intuition of Francis Galton seems to have drawn an incorrect inference[5] where this question was involved". He then went on to "demonstrate the fact stated by Cattell for families of any size, say families of n children each, in which boys and girls occur on the average in equal numbers." Just like Falk, his demonstration consisted of algebraically deriving the expected value of a random variable with a (symmetric) $Binomial(n-1, 1/2)$ distribution.

2.3 Simulated datasets using population data from 1911, and crowdsourced data from 1884

While waiting to approach new classes with sufficient numbers of male and female respondents, I used data from the digitized household returns from a 1911 census to anticipate what I might find in these classes. No matter whether or not I restricted the selection to make family sizes closer to those today, I was unable to find either of the two patterns that Mosteller promised. In the family data crowdsourced in 1884 by Francis Galton (and unpacked by Hanley (2004)) the Mosteller promise did hold for the mean number of brothers (2.65 among the women, and 2.52 among the men), but not for the mean number of sisters (2.49 among the men, but 2.70 among the women).

[5] If Rietz were writing this today, he might also criticize Galton's racial views.

2.4 Class data from 2023

I gathered new data from a class in February 2023. To have a larger sample size, I asked students not only about their own siblings, but also about their parents' siblings. Table 2 tallies of the numbers of same- and opposite-sex siblings the 64 students they themselves have, and how many their parents have/had.

		Number of			No. Paternal			No. Maternal	
Respondents		Brothers	Sisters		Uncles	Aunts		Uncles	Aunts
10 Males	.	7	6	.	23	18	.	22	24
54 Females	.	30	31	.	79	91		75	69
Their Parents									
64 Fathers		102	109						
64 Mothers		97	93						

Table 2: Numbers of Sisters and Brothers, and Paternal and Maternal Uncles and Aunts reported by 64 first year students in McGill's Master's programs in Public Health and in Epidemiology, 2023. The data on uncles and aunts are aggregated in the last two rows

Whether one considers tables 1 and 2 separately, or combines them, it is difficult to find much support for, or understanding of the pattern Mosteller promised.

3. TIME TO CONSULT OTHERS – AND HELP FROM A RELATED PROBLEM

Despite the new data, I *still* trusted Mosteller, and continued to wonder what might have been producing the patterns he had seen. So, I emailed a draft of the manuscript to several statisticians who knew Fred very well, and

I asked these consultants what it could be.[6]

The first comment arrived next morning, from a person "awaiting (in just a few hours!) the birth of their 7th grandchild," and focused on the importance of considering the *provenance* of the data

> I do remember reading the Mosteller article and being both intrigued and humbled by the complexity of this "simple" situation. Your draft article puts a laser focus on the importance of the sampling plan more generally. Regarding that, I suggest highlighting at the end of the article, guidance to ask, "How did I get to see the data?" It underlies everything from length-biased sampling to (almost) all of causal analysis.

A second answer arrived that afternoon, and it was even more humbling! [7]

> Jim , you reach me at the airport, soon to depart for London for 2 weeks. I skimmed your draft.
> 1. Well written.
> 2. You mention Fred's 50 Problems. Have you looked closely at it?
> 3. Fred would have known prob M not = prob F. Also he would have known that the answer hinges on the distribution of family balance. Which depends on stopping rule and family planning. If M-F is closer on average to zero than binomial you get one answer. If further you get the other answer.

Of course! How easy to forget what we preach: that the Binomial must have a pre-specified fixed number of (identical Bernoulli) trials. More expansive answers from two other consultants arrived within days

> Suppose that we limit our universe to families with two or more children. So, a randomly chosen child will always have at least one sibling. At some fixed point in time, the second child was born. At that point, we have (B,B), (B,G), (G,B), or (G,G). Does the probability that the couple has additional children vary depending on which of the four situations we are in? For example, might those with (B,B) and (G,G) be MORE likely to elect to have a larger family? This is an example of a statistical (and empirical) question, more than a question about probability model specification.

[6] Since I could not think of any reasons, I had even wondered (as some of Martin Gardner's readers had) whether Mosteller might have deliberately blundered in order to test his readers.

[7] As per Fred's legendary advice on how to provide feedback, it started out with praise!

> Isn't there an issue of preference for sons? So if a first child is a girl there may be an impetus to have more children in order to have a boy? Can't figure out how that would affect your reasoning, but I think it would. In addition, that preference for boys might especially affect the all-male data set of judges discussed above. Also, it may be that the preference for boys is influenced by social class, with a stronger preference among higher classes (judge data and college sample) than for the Irish sample. And finally, my understanding is that population data show that the conditional probability of a third child being of the same sex as the two previous children in families with two same-sex children is greater than one-half. Wish I knew a reference for this, but I don't.

3.1 Evidence of these selective forces

The theoretical and empirical literature on family planning, and its consequences for the sex ratio, the numbers of sibs, birth orders etc., is extensive. There is also the added complexity that the probability of a male birth may vary across and within families ('Lexian', 'Poisson'' and 'Markovian' variation – see, for example James (1987)). I began by simulating various scenarios that just involved planning, but also looked for more modern-day data, preferably on completed families. A part of one such dataset came to light while I followed up on variants of the so-called other-child problem, popularized by Gardner (1959) and vos Savant (1997).[8] Mistakenly, I had initially believed this was involved in the Mosteller promise. It is the focus of the articles by Carlton and Stansfield (2005), Stansfield and Carlton (2009) and the response by Garenne (2009), and their data on the sex-composition of 2-children families. Carlton and Stansfield's initial purpose was to use family composition data collected within the US National Health Interview Survey (NHIS) using scientific sampling – rather than from volunteers – to settle the (close to 1/3? 1/2?) question magazine columnist Marilyn vos Savant (1997)

[8] It continues to be of interest – see Senn [2022] and Paindaveine [2023].

had passed on to her readers

> A woman and a man (unrelated) each have two children. At least one of the woman's children is a boy, and the man's older child is a boy. Do the chances that the woman has two boys equal the chances that the man has two boys?

But, along the way, they noticed that the numbers of families with 2 girls (2G), 2 boys (2B), and 1 of each (1G1B) did not fit a binomial distribution: as is evident in the first row of Table 3, there were clear deficits in the numbers of same-sex sibships. These deficits suggest – as my consultants had – that, compared with those who already had "gentleman's family",[9] the families with two same-sex children were more likely to continue on to have more children, and thus *select themselves out* of the dataset of 2-children families.

Garenne, who fitted differential continuation probabilities (0.28 when it is already 1B1G, 0.31 when it is 2B, and 0.36 when it is 2G), was more emphatic:

> This shows that U.S. families seem to have a net preference for stopping after two children when they have a balanced family (one boy and one girl) and are more likely to have a third child if they already have two girls than if they already have two boys.

He also carried out similar calculations on African data, where family limitation is not yet an issue, and where the continuation probability after 2 children is 0.94, and does not differ according to family composition. "The distribution of the first two births and the distribution of two-children families among women age 40 and older are not different from the expected values computed from the binomial distribution with heterogeneity. This is because few African women use contraception after two children, and women who have only two children by age 40 are primarily women with secondary

[9] A two-child family, one of each sex.

sterility." This may explain why I did not find evidence of selection in the more 'natural fertility' settings such as the 1911 census data from Dublin, Galton's crowd-sourced data from 1884, and the British Peers 1603-1959 dataset (Reid 2021).

The last four rows of Table 3 *do* show show the pattern that Mosteller referred to, but they are limited to the subset of NHIS families with *exactly* 2 children.

Sex composition	2G	1G1B	2B		
Number of ...				Total	
Families	5,844	13,079	6,545	25,468	
Men	-	13,079	13,090	26,169	
Women	11,688	13,079	-	24,767	
				Mean	
Sisters (S) of Men	-	13,079	-	$\frac{13,079}{26,169} = 0.50$	$\bar{S}_{men}$
Sisters (S) of Women	11,688	-	-	$\frac{11,688}{24,767} = 0.47$	$\bar{S}_{women}$
Brothers (B) of Women	-	13,079	-	$\frac{13,079}{24,767} = 0.53$	$\bar{B}_{women}$
Brothers (B) of Men	-	-	13,090	$\frac{13,090}{26,169} = 0.50$	$\bar{B}_{men}$

Table 3: Sibship statistics: 2-children families in 1998–2002 National Health Information Surveys (Carlton and Stansfield, 2005).

This may not necessarily be seen across *all* the families in the NHIS. Indeed, so far, I been unable to find any family planning scenario where ***both*** of Mosteller's contrasts (i.e., $\bar{S}_{men} - \bar{S}_{women}$ ***and*** $\bar{B}_{women} - \bar{B}_{men}$ have a positive expectation. The reasons became clearer after I asked some alumni of Harvard's Statistics Department if any of them had been a teaching fellow in Mosteller's classes, and received the following reply

> Jim - I was not one of Mosteller's TA's but I have a theory about why this may be true. Suppose a family wants at least one boy and and least one girl. If the first k children are girls, the family will continue to have kids until they have a boy. So the boy has k sisters but each girl only has $k-1$ sisters. If the stopping rule is independent of the sex composition, then the expectation is even. So on average what Mosteller says is likely correct.

This scenario was more extreme than I had considered, and initially seemed to offer an explanation. However, since family sizes cannot be extended indefinitely, some families will stop before reaching their goal. When, as in Table 4, one includes these unisex families, this *seeming deficit of sisters of women* – and its counterpart when we focus on brothers of men – *disappears*.

Family no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	total	TOTAL
1st born	G	G	G	G	G	G	G	G	B	B	B	B	B	B	B	**B**		
	sss	ss	s	s	–	–	–	–	s	s	s	s	s	s	s	–		
2nd born	G	G	G	G	B	B	B	B	G	G	G	G	B	B	B	B		
	sss	ss	s	s	s	s	s	s	–	–	–	–	s	s	s	–		
3rd born	G	G	B	B									G	G	B	B		
	sss	ss	ss	ss									–	–	s	–		
4th born	G	B													G	B		
	sss	sss													–	–		
No. Men	–	1	1	1	1	1	1	1	1	1	1	1	2	2	3	[4]	18	22
No. Men's Sisters	–	3	2	2	1	1	1	1	1	1	1	1	2	2	3	[–]	22	22
No. Women	[4]	3	2	2	1	1	1	1	1	1	1	1	1	1	1	[–]	18	22
No. Women's sisters	[12]	6	2	2	–	–	–	–	–	–	–	–	–	–	–	[–]	10	22

B = Boy
G = Girl

s = Boy's sister
s = Girls' sister

Table 4: Expected distribution of numbers of sisters (s) if families (shown in separate columns) continue to have children until they have at least one boy and and least one girl, or until the have four children, whichever comes first. So as to have integral frequencies, the probability of a male birth is taken as 1/2, and the number of families as 16. The resulting expected 22 Boys and 22 Girls are color-coded as B or G; The lowercase textstring underneath each B or G, indicates the number of sisters each child will have at the time the family is completed. In the rightmost end of each row at the bottom, and assuming that all 44 reach adulthood, the numbers in the (a) 'total' and (b) 'TOTAL' columns are respectively aggregated over the (a) 18 persons from the 14 families that reached their goal, and (b) over all 22 persons in the 16 families

Likewise, I was unsuccessful when I introduced Lexian variation (e.g., setting the probability of a male birth at say 0.615 in half the families and 0.415

in the other half).

3.2 Changing the focus: opposite- vs. same-sex siblings, and counting respondents rather than comparing means

Some of those to whom I mentioned Mosteller's statement processed it differently. Later, they remembered Cattell's and Galton's datasets, where *men had more sisters than brothers* and wondered whether *women would have more brothers than sisters* as well. This prompted the question: what would happen if, in class, we simply asked for a *show of hands* as to whether students had more opposite-sex or same-sex siblings? A reworking of the data in Table 4 shows that whether we ask the 22 men or the 22 women, we can expect 11 to have more opposite-sex than same-sex siblings, 7 to have more same-sex than opposite-sex siblings, and 4 to have the same number of same-sex and opposite-sex siblings.

3.3 Further selectivity: who exactly these data were collected from

Since stopping rules that were a strong function of the sex-composition of the family did not produce the pattern Mosteller described, I went back to another initial (pre-consultation) idea, namely that the probability of being a *student* in Harvard or Radcliffe in that era was also a strong decreasing function of family size. I was not able to find instances where this scenario *alone* produced the patterns he saw. But, when I *combined* it with differential stopping probabilities, it did.

The reason can be understood by stratifying the individuals in Table 4 by the size of their family:

	Number of		.	Number of sisters(s)		
Family Size	Men	Women	.	of Men	of Women	$s_W - s_M$
2	8	8	.	8	0	- 8
3	6	6	.	8	4	- 4
4	8	8	.	6	18	+12

Note that if the fraction of individuals that attend college is *lower* in those from *larger* families – and if these family sizes/compositions were governed by the wish for a gentleman's family – then the excess is not fully offset. Thus, if ***both*** of these selective forces are opeating, the Mosteller expectations, even if not very pronounced, should prevail.

This example strongly reinforces my first consultant's advice to understand the *provevance* of the data. It points up how easily we can go astray if we do not pay attention to the specific (here, the *student*) context, and how easily the message can be lost in the translation and re-telling, such as in the omission of the word 'student' in the title of Falk's article.[10]

Interestingly, the Stack Exchange (2023) and Quora (2023) online groups contain long discussions of the question raised by the title of Falk's article, but neither considered a specific data context: their theoretical calculations and simulations seem directed at a hypothetical population, where neither the family-composition nor the education-based selectivity is operating.

[10]The title of this current article includes a similar qualifier, but I fear it might give away the punchline. If reviewers think it does, I can remove it.

4. USING SIBLINGS DATA IN TEACHING

"Gathering data in class and then analyzing them with the help of the class" is probably more common – and certainly easier – today than it was it was in the Mosteller era. The topics that are relevant to students have changed considerably still then – as have family planning (Wood 1977; Sloane 1983; Pollard 2002; Tian 2015) and access to higher education. But family sizes continue to be of pedagogic value, such as to highlight the different answers obtained when sampling by family versus by child (Gelman and Nolan, 2017).

The sex-composition of students' families can add several interesting twists, once students have been introduced to the models underlying Bernoulli, geometric, binomial and negative binomial random variables. One important aspect, missed by Falk, is the difference between hypothetical random variables in the classroom, and the real-world random variables induced by modern family planning and by the socio-economic forces that determine who goes to college, and which college. Compared with earlier eras, *completed* families of size n are no longer as close to a seemingly random subset of the families that had *reached* size $n-1$. Nor are the family sizes seen in surveys in college classrooms necessarily representative of those seen in the NHIS.

A second twist is the value of recasting the family composition data: instead of depending on the traditional presentation (typically as numbers of *families*), Mosteller's questions instead focus on individuals (students) themselves, on their sibs, and on specific differences in means, or (if one just asks for shows of hands) in proportions.

A third use of sibs data is to address the 'framing' of the question and to stress the importance of considering the provenance of data (how did the data come to be, and how did we come to see them?) Falk and Cattell made effective use of narratives to seduce readers. It was interesting to watch how students and colleagues differed in their opinions about Mosteller's 'usually' statement, depending on what else I told them. I am embarrassed that I initially invoked the 'other-child' problem, and its conditional structure, after my 2001 data seemed to support Mosteller's statement. It is possible to extend the Gardner and vos Savant problems to families of any size, but in the classroom survey one does not know *in advance* how large a respondent's family is before asking about his/her sibs. Moreover, Mosteller's statements address $\bar{S}_{men} - \bar{S}_{women}$ and $\bar{B}_{men} - \bar{B}_{women}$, whereas I initially split up the 2001 class data into men and women, and instead focused on $\bar{S}_{men} - \bar{B}_{men}$ and $\bar{B}_{women} - \bar{S}_{women}$. I had a sample size of 42 for the second of these, and blamed the non-significant difference within the men on the very small sample size of 22.

Fourth, despite the apparent simplicity of Mosteller's question, there are many exits in the path between the cradle and the classroom. And so it is important to spell out all of these assumptions and pathways, just as Senn (2022) did when addressing the vos Savant question. And, in keeping with the "How did I get to see the data?" comment from one of my consultants, I am reminded of the 'filtering' that led to a seeming coincidence involving draws in lotteries in 2 states that are adjacent, both geographically and in the USA Today listings (Hanley 1992, Lottery Case 1). Performing simulations can help to trace the data and bring out the assumptions made at each step.

How likely one is to observe realizations of both of Mosteller's *usually*'s is a function of class size, its sex-composition, the extent and nature of family planning, and of differential college attendance. I am hoping that this journal will serve as a forum where teachers can share tabulations of the class data they collect on their students – and maybe their student's parents – and tell us whether his *usually*'s[11] still hold. I leave the very last words on this to another of my consultants

> Unfortunately I was a TF for Fred only once, in spring semester of 1970, when I knew almost nothing about anything. My main memory from that experience was giving my very first class to a class of about 120 on the birthday problem. To my utter astonishment, I required over 40 student birthdays before getting a match. In the post-mortem Fred gave on my class, I really learned about variability, rare events and coincidences, so at least I learned something useful from that class.
>
> I don't recall ever hearing about the sib count problem, so I can't add much from 50 years ago. I did have an aha moment about knowing how data came to be from a talk by Dennis Lindley here. I forget the context, but using a binomial ignoring the data generating process that was geometric led to, shall we say, wrong answers. Reading the problem in your article I automatically doubted a binomial model for siblings. Since family structure and decisions on family size have likely changed over time, Mosteller's answer might have been true when he wrote it and false now.

[11] If we needed one more example of the range and relevance of Mosteller's work, it would be the one by Mosteller and Youtz (1990). They found ten studies of the quantitative meaning of the word '*usually*', yielding probability values from 70% to 85%, and unweighted and weighed averages of 77% and 79% respectively. Only one study reported the probability meaning of '*unusually*' – 19%.

Acknowledgments

I thank my consultants: Allan Donner, John Emerson, William Fairley, Katherine Halvorsen, David Hoaglin, Tom Louis, Milica Miočević, Joe Sedransk, Judith Singer, Stephen Stigler, Judith Tanur, Sanford Weisberg, and Janet Wittes I apologize to my McGill biostatistics colleagues who suffered through my early theories based on the 'other-child' problem. And I thank my epidemiology colleague Jay Kaufman for key references from the "rich literature on this (very complex!) issue in demography, sociology and fertility journals." He was surprised that I "spent so much time talking to biostatisticians and so little time talking to the people who actually study this: demographers!"

Funding: None.

The author reports there are **no competing interests** to declare.

REFERENCES

Bar-Hillel, M., Falk, R. (1982), "Some teasers concerning conditional probabilities," *Cognition* 11 (2): 109–122. doi:10.1016/0010-0277(82)90021-X. PMID 7198956. S2CID 44509163.

Carlton, M. A., and Stansfield, W.D. (2005), "Making Babies by the Flip of a Coin?," *The American Statistician*, 59(2), 180-182. DOI: 10.1198/000313005X42813

Cattell, J.M., and Brimhall, D. R. (1921), "American Men of Science, a Bibliographical Directory," 3rd Edition, Garrison, New York The Science Press.

Census of Ireland, 1911. http://www.census.nationalarchives.ie/pages/1911/Dublin/Stillorgan/Woodpark__Part_of/91800/

Edwards, A. W. F. (1958), "An analysis of Geissler's data on the human sex ratio," *Annals of Human Genetics*, 23(1):6-15. doi: 10.1111/j.1469-1809.1958.tb01437.x.

Edwards, A. W. F, (1966), "Sex-ratio data analysed independently of family limitation," *Ann. Hum. Genetics* , 29, 337-347.

Falk, R. (1982), "Do men have more sisters than women?," *Teaching Statistics,* 4, 60-62.

Falk, R. (2011). "When truisms clash: Coping with a counterintuitive problem concerning the notorious two-child family," *Thinking & Reasoning*, 17 (4): 353-366. doi:10.1080/13546783.2011.613690. S2CID 145428896.

Gardner M. (1959) "Mathematical Games: The Two Children Problem". *Scientific American.*

Gardner, M (1961). *The Second Scientific American Book of Mathematical Puzzles & Diversions.* New York : Simon and Schuster, 1961. University of Chicago Press 1987.

Garenne, M. (2009), "The Sex Composition of Two-Children Families: Heterogeneity and Selection for the Third Child: Comment on Stansfield and Carlton (2009)," *Human Biology*, 81(1): 97-100.

Gelman, A., and Nolan, D. (2017), "How large is your family? Section 6.1.6 Data collection," *Teaching Statistics: A Bag of Tricks* (2nd edn). Published: 2017 Online ISBN: 9780191827518 Print ISBN: 9780198785699

Hanley, J. A. (1992), "Jumping to coincidences: defying odds in the realm of the preposterous," *The American Statistician*, 46(3) 197-202.

Hanley, J. A. (2004) "'Transmuting' Women into Men: Galton's Family Data on Human Stature" *The American Statistician*, 58(3), 237-243.

James, W. H. (1987), "The Human Sex Ratio. Part 1: A Review of the Literature," *Human Biology* 59, 721-752.

Khovanova T. (2012). Martin Gardner's Mistake. The College Mathematics Journal , 43(1), 20-24.

Monakhov, M. (2015), "Association between sexes of successive siblings in data from Demographic and Health Survey program," *bioRxiv preprint* doi: https://doi.org/10.1101/031344. List several large datasets, starting with Geissler.

Mosteller, F. (1980), "Classroom and platform performance," *The American Statistician*, 34 (1), 11-17.

Mosteller, F., Youtz, C. (1990), "Quantifying probabilistic expressions," *Statistical Science*, 5 (1): 2-34, JSTOR 2245869, MR 1054855

Paindaveine, D. and Spindel, P, (2023), "Revisiting the Name Variant of the Two-Children Problem" *The American Statistician.*

Reid, A., and Newton, G. (2021), "British Peers, 1603-1959. [data collection]," *UK Data Service.* SN: 8698, DOI: http://doi.org/10.5255/UKDA-SN-8698-1

Senn, S. (2023), *Dicing with Death: Living by Data.* Second Edition. Cambridge, United Kingdom ; New York, NY : Cambridge University Press.

Stansfield, W. D., and Carlton, M.A. (2009), "The Most Widely Publicized Gender Problem in Human Genetics," *Human Biology*, 81(1): 3-11.

Mathematics Stack Exchange (2023), "Do men or women have more brothers?," https://math.stackexchange.com/questions/1794123/do-men-or-women-have-more-brothers

Pollard, M. S., and Morgan, S. P. (2002), "Emerging parental gender indifference? Sex composition of children and the third birth," *Am Sociol Rev.* 67(4): 600-613.

Quora (2023), "Do men or women have more brothers?," https://www.quora.com/Do-men-or-women-have-more-brothers

Rietz, H. L. (1924), "Note on the average numbers of brothers and sisters of the boys in families of n children", *Science*, Vol 60, No. 1541, 46-47.

Seidl, C. (1995), "The Desire for a Son Is the Father of Many Daughters: A Sex Ratio Paradox," *Journal of Population Economics*, 8(2), 185-203.

Sloane, D. M. (1983), "Sex of previous children and intentions for further births in the United States, 1965-1976," *Demography*, 20(3) 363-367.

Tian, F. F., and Morgan, S. P. (2015), "Gender Composition of Children and the Third Birth in the United States," *Journal of Marriage and Family* 77: 1157-1165.

vos Savant, M. (1991), "Ask Marilyn". Parade Magazine. October 13, 1991 [January 5, 1992; May 26, 1996; December 1, 1996; March 30, 1997; July 27, 1997; October 19, 1997].

Wikipedia. Boy or Girl paradox

Wood, C. W., and Bean, F. D. (1977), "Offspring Gender and Family Size: Implications from a Comparison of Mexican Americans and Anglo Americans," *Journal of Marriage and Family*, 39(1), 129-139.

The American Statistician: Decision on TAS-23-134

The American Statistician <onbehalfof@manuscriptcentral.com>
Fri 2023-09-01 10:24 AM
To:James Hanley, Dr. <james.hanley@McGill.Ca>

1 attachments (48 KB)
Attached standard file: revTAS.pdf;

01-Sep-2023

RE: TAS-23-134, "In my classes, usually the men(women) have more sisters(brothers) than women(men) do":
a Mosteller statement that continues to intrigue"

Dear Professor Hanley:

Thank you for submitting your paper to The American Statistician (TAS). Your paper has been reviewed by an Associate Editor (AE) and two referees. Based on the reports and my own assessment, I have decided to reject your paper. Comments from the AE and referees are at the bottom of this email and in the attachment.

I think everyone enjoyed reading your paper, including me. The reviews, however, were quite mixed, and both the AE and I did not believe the level of the paper was suitable for TAS. Upon my second reading, I would recommend submitting your paper to Significance or Chance. I would recommend shortening your title, which I found to be confusing and a little off-putting.

I am sorry to write disappointingly, especially because your paper describes an interesting classical problem. However, please remember TAS receives many more worthwhile submissions than we can possibly publish. I hope you are able to find a suitable outlet for your paper.

Thank you for giving us the opportunity to consider your work.

With best regards,

Joshua M. Tebbs
Editor, The American Statistician

Reviewer 1:
Comments to the Author
The paper is well written, and emphasizes how important it is to make sure you are asking and answering the right question. The arguments are correct, and a good discussion is given of the various perspectives.

Reviewer 2:
Comments to the Author:
See attached file.

Associate Editor:
Comments to the Author:
Often the best solution to a tricky-looking problem is the simplest, and this seems to be indeed the case here. Consider a family with n siblings and let the number of boys be X. Then $X \sim bin(n,1/2)$ and $EX=n/2$. If a boy is selected, the number of sisters he is expected to have is $n/2$ whilst the number of brothers he is expected to have is $n/2-1<n/2$. A similar argument applies if a girl is selected. Falk (1982) does mention self-exclusion as the solution to the problem, but then goes to give an incorrect probabilistic analysis that seems to contradict Mosteller's claim. Based on Falk's wrong analysis, the current author unfortunately meanders around the real solution without actually hitting it.

Comments to authors of TAS21341

This paper revisits a quote from Frederick Mosteller in the context of a suggested project to engage students in a statistics class. Mosteller alleges that the women in the class tend to have more brothers than sisters, while the men in the class tend to have more sisters than brothers. Mosteller does not offer an explanation, but psychology Professor Ruma Falk assumed that the basis of his error was that the numbers of boys and girls in a randomly selected family of a given size should balance out (under the assumption that sexes at birth are iid Bernoulli with parameter (1/2)). Because a woman from the class is one of the girl children in her family, she tends to have one more brother than sister to make the family balance out. Similarly, a man from the class tends to have one more sister than brother to make his family balance out. The current paper, both interesting and provocative, puts the best possible light on Mosteller's statement by proposing mechanisms by which his statement may be correct.

1. I would like to see an unequivocal statement that, under the above assumption implicit in Ruma Falk's argument (i.e., conditioned on number of children=n, the sexes are n iid Bernoullis with parameter 1/2, Mosteller's contention is false. The fallacy in the above argument that there should be one more brother than sister if the respondant is a female, and one more sister than brother if the respondant is male, is as follows. The numbers of boy and girl children from a **randomly** selected family tend to be balanced, but the family of a girl in the class is NOT randomly selected from the set of all families. The fact that this family has at least one daughter tells us that the number of girl children from this family is stochastically larger than the number of girl children from a randomly selected family. Thus, from the family of a girl from the class, we should not expect balance among all children. Instead, we should expect more daughters than sons, on average. The fact that we should no longer expect total balance among the girl from the class and her siblings destroys the cornerstone upon which the above specious reasoning is based. Therefore, it is quite clear that, under the assumption that Ruma Falk implicitly makes, Mosteller's statement is false. If Mosteller had said that, when the class respondant is female, the **total** number of girl children (including the class respondant) tends to be greater than the total number of boy children, that would have been correct. Similarly, when the class respondant is male, the total number of boy children (including the class respondant) tends to be greater than the total number of girl children.

2. The alternative explanations justifying Mosteller's statement are interesting and may be true. It is quite possible that families want at least one son and at least one daughter, so there may be implicit stopping rules not accounted for by Ruma Falk's assumption. The fact that the data don't consistently support Mosteller's statement makes it harder to make that conclusion. Of course, there are other, less plausible explanations. For example, the probability of taking the class might be related to family dynamics such as the number of brothers and sisters. Maybe females are taking the class to learn why they have fewer sisters than brothers! Also, I am no expert in genetics, but it seems reasonable to me that each family could have their own random probability of producing a boy, so a mixed binomial may be a more reasonable assumption. The mixing distribution might have mean 0 and a small standard deviation. Under this model, the answer would change. I feel that there are many possible explanations, but the real data you provided don't convincingly support Mosteller's statement.

3. The most interesting question to me is whether Mosteller made a straightforward probability error, or he was thinking of some other explanation such as one or more of those you propose or the ones in my previous comment. Unfortunately, he offers no explanation for his statement. I understand the thinking that Mosteller was a luminary in statistics, and would not have made such a probability error, but if that is true, why did he not write a response to Ruma Falk's 1982 paper explaining why his thinking was not erroneous after all? I certainly would have done so if I had been accused of making a probability error that I did not make. Even the brightest people make mistakes. Given that (1) the data are inconsistent, (2) Mosteller gave no expanation for his thinking, and (3) Mosteller remained mute after being accused of making a probability error, I lean more toward the conclusion that he simply made a probability error.

4. Minor issue: I count only 63 families in Table 1, whereas the caption implies there are 64.