AP Statistics: Passion, Paradox, and Pressure

(PART I)

Xiao-Li Meng

Same Passion But Different Populations

Flowing in Rossman, Peck, Franklin, Hartlaub, and Scheaffer's letter (November, 2009, hereafter RPFHS) is passion and persuasion: passion for AP statistics and persuasion for its effectiveness. I greatly applaud their effort as passionate leaders and devoted promoters of statistical education, e.g., Professor Rossman is the current President of International Association for Statistical Education, Professor Scheaffer is a past President of ASA, and Professor Franklin's textbook (with Professor Agresti) is actually the reference book for the Happy Course described in my op-ed (which is the piece that RPFHS commented on). The demand for statistics is such that we now need significantly more passionate and effective statistical educators, especially for introductory courses, for reasons discussed in Brown and Kass (2009) and Meng (2009).

The same passion for the future of statistics has led to recent efforts at Harvard Statistics; the Happy Course is just one of them (see Meng (2009) for others). Strong persuasion was also intended in my op-ed for *The Harvard Undergraduate Research Journal (THURJ)*, persuading undergraduates to take at least one course in statistics, especially those who had been turned off by their AP statistics experiences. Coincidently, a colleague just forwarded an email from a Harvard undergraduate who read my op-ed and whose reaction is the type that prompted me to acknowledge their frustration (evidently this student made an effort to have himself "turned on" —Stat 104 is the course taught by the protagonist of the "Jesus" quotation in my op-ed):

> "That article's section #2 was actually one of my largest concerns last year because my AP Stat class was dreadfully unexciting and just pure regurgitation (and probably why I have to listen to Stat 104 lectures that a friend who took it last year downloaded for me to refresh myself). The article was a really fun read, unlike popular opinion about statistics being "boring." "

I was therefore puzzled by the statement in RPFHS that my "criticism" of AP statistics is "misdirected". My op-ed was not assessing—much less criticizing—the *overall* quality of AP statistics, but rather addressing a situation on cases where it had a negative impact, as reported by students. In contrast, what RPFHS cited are cases where students have been turned on by their AP statistics experiences, also reported by students.

There were two sentences in my op-ed that mentioned AP statistics. One of them is (all emphases in the current article are added)

> "And for nearly every one of you (i.e., undergraduates) I have spoken with, the number one reason that you *did not even consider majoring (or concentrating, to be true to the Harvard spirit!) in statistics* is because the AP statistics you took convinced you that statistics is the most boring subject."

The other is a literal quotation from one of those students: "AP Statistics was the most boring course I took in high school!" I was therefore quoting from students who chose to *stay away from statistics.* Let's denote that population as A. By contrast, RPFHS was quoting from students who are already *in statistics major/minor/courses:*

> "The number of undergraduate students at Cal Poly and the University of Georgia *choosing statistics as a major or minor* has steadily increased since the advent of AP Statistics, with many students attesting that their choice was based largely on their positive experience in AP Statistics. Similarly, the number of students *enrolling in statistics* courses at Kenyon College has increased substantially over this time as well, with many students citing AP Statistics as the reason for their interest."

Putting aside the issue that different institutions are involved, the students quoted in RPFHS belong to the *complement* of A, A^C. We statisticians understand well that conditioning on A and conditioning on A^C are different matters, just as association (over time) and causation are not the same thing. Therefore, the intended or unintended uses of such mixed arguments in RPFHS puzzled me.

There is of course no puzzle whatsoever in the fact that an education program can have both significant positive and negative impact, depending on how it is actually implemented. When an AP statistics course or any other introductory course is done well, such as those described in RPFHS, it does a great service to our profession. When it is done poorly, as experienced by some Harvard undergraduates, it has a strong effect but in the opposite direction. Therefore, the evidence from these different observations actually reinforce the same point: the first courses in statistics, regardless whether at the high school level or the college level, are absolutely critical and we need substantially more passionate and skilled educators in order to maximize their positive impact. In other words, the worry discussed below is not about the AP program per se, which of course has increased awareness of statistics at the high school level by putting statistics on an equal footing with many other subjects. The worry is simply due to the severe shortage of qualified statistical teachers who can teach introductory courses in such a way to arouse students' interests in statistics, or at least not to turn them away from statistics.

The Off-Diagonal Paradox: Do We Turn On More than We Turn Off?

With the evidence of the non-emptiness of both A and A^C, a scientific assessment of the overall effectiveness of any program such as AP statistics then must ask, minimally, *has the program attracted more students to statistics than if it were not in place?* This is squarely a causal inference question, one that is arguably as hard as—and therefore needs to be addressed as carefully as—"does smoking cause lung cancer?" Counterfactual causal questions as such are often impossible to answer definitely, but nevertheless they are essential for formulating relevant comparisons, designing meaningful studies, and guarding against what I called incentive bias (Meng, 2009), that is, humans' tendency, however subconsciously, of selectively collecting and presenting evidence that support one's causes. Minimally, it reminds us that given that the general demand for statistics has been increasing rather dramatically, especially in recent years, any type of increase in enrollments over the years itself cannot be taken as scientific evidence of the effectiveness of a particular program intended to attract students to statistics.

To see this clearly, let us use a generic binary "in" and "out" variable; here "in" can mean to take a statistical course or to major in statistics, or some other outcome. Regardless of its actual meaning, Table 1 is applicable to any program designed to attract membership.

	In	**Out**	
Would	P_{ALWAYS}	P_{OFF}	P_{WOULD}
Would Not	P_{ON}	P_{NEVER}	P_{NOT}
	P_{IN}	P_{OUT}	

Table 1—The row and column variables represent before and after "the treatment"

Here P_{ALWAYS} and P_{NEVER} are respectively the percentages of students (at time t) who will enter and not enter statistics regardless of whether our program is in place or not; P_{OFF} is the percentage of students who would enter statistics but got turned off by the program, and P_{ON} is the percentage of students who would not enter statistics but got turned on by the program. From this setup, conceptually it is clear that observing $P_{IN} = P_{ALWAYS} + P_{ON}$, or even directly observing P_{ON}, large or even increase over time, says little about the *overall effectiveness* of the program because P_{OFF} can also be large and even increase over time, which can offset the gain by P_{ON} whenever $P_{OFF} > P_{ON}$.

This, of course, is trivial arithmetic. But just as Simpson's paradox can be explained by trivial arithmetic, yet has led to numerous erroneous conclusions throughout the history of quantitative investigations, it is easy for us to focus on the "In" column because it is the population most easy to identify and sample from (as in RPFHS), and arrive at assertions of the benefit of the program while it actually

might be doing harm in reality. To raise awareness of this phenomenon, I suggest it be recognized as the "*Off-Diagonal Paradox*", invoking a similar connotation of "paradox" as in Simpson's paradox to urge investigators to always keep in mind the need for comparing the two cells along the off-diagonal in Table 1. The word "off" should also serve as a reminder that the effectiveness of the overall program for recruitment cannot be assessed by only asking those who are already *in*: we need also to ask those who are *off or out*.

For AP statistics, the inequality $P_{OFF} > P_{ON}$ can hold even when P_{ON} increases as long as P_{OUT} remains large (which is true, for example, for majoring in statistics), and when there are more poorly taught AP courses than well taught ones; the latter is one of the issues needing to be examined in any assessment of the overall impact of the AP program. Not having enough well-qualified teachers is a well-known problem, even at the college level, a situation summarized so vividly by a college professor who wrote to me:

> "While I am not losing sleep over the three "puzzles" you posed in your fine article in the recent issue of the *Amstat News*, I am losing sleep over the very serious problem of not having competent classroom instruction in lower level undergraduate courses. In particular, the first and second introductory courses are extremely problematic, as you have pointed out. Unfortunately, our mathematics colleagues think that "anyone" can teach an introductory statistics course or, that having had one or two courses with emphasis in probability rather than statistical methodologies and thought, more than meets/surpasses any possible qualifying criteria to teach introductory stats courses. As we know, this problem is not new, but it is reaching critical mass."

An Even Harder Question: What Kind of Students Are We Turning Off?

An astute reader may have been wondering why I have not addressed an obvious question, that is, could it be that the "turn-off" experience is unique to Harvard undergraduates? At the time of writing my op-ed, Harvard undergraduates were my *targeted population*, so this question was of little interest. However, now that the "Harvard observation" is brought to the national level, I am indeed inspired by RPFHS to consider its general applicability. Coming with the inspiration, however, is perspiration.

I perspired not because of the realization that I committed an elementary error of extrapolation, but rather because of the realization that if the "turn-off" phenomenon is indeed more likely for students at Harvard or Harvard-like institutions, then our profession has an even deeper and perhaps more disturbing problem to worry about. Harvard undergraduates are undoubtedly a highly selected group. But they are not so different from undergraduates at many of Harvard's peer institutions, nor did they all come from one school district or one state. It is, however, quite possible that academically strong students have a higher likelihood to walk away from a poorly taught subject, statistics or not, than academically less able ones. Other than the fact that the former will have more choices at their disposal, it is not hard to imagine that the former are also more likely to be turned off by mechanical teaching emphasizing memorization for testing, as alluded to in the aforementioned student's email. Indeed, the latter may even prefer such mechanical teaching because it is less challenging than inspirational teaching which requires high-level creativity and independent thinking, the very traits we all look for when we recruit students (and faculty). Therefore, if Harvard undergraduates can be viewed as a sample of high achieving high school students and if such students report more "turned-off" experiences, then we must ask ourselves not only what percentage of students are turned off by poorly taught AP classes, but also what kind of students are more likely to be turned away.

I perspired more when this was connected with the following anecdote, first heard from a Harvard undergraduate and then independently from a high school student in California (the son of a friend). In both cases, the student reported that his high school guidance counselor advised students that they should consider taking AP stat courses only if they cannot survive AP calculus courses. This suggests that at least in some high schools AP statistics is perceived as a "softer alternative" for students who cannot yet handle calculus.

Before we all get enraged by such a condescending perception, let us collectively keep a cool head and ask a deeper question. Should we then consider the impact of the AP program on the overall quality of students it helps to attract in addition to how many it attracts? Consider two versions of Table 1, one for strong students and one for weak students (the dichotomy, of course, is for simplicity of illustration), labeled correspondingly with a subscript S for strong and W for weak. Then arithmetically it is possible that we have

$$P_{OFF,S} + P_{OFF,W} < P_{ON,S} + P_{ON,W},$$

but $P_{OFF,S} > P_{ON,S}$

That is, even if the program succeeds in attracting more students in total, we can still end up with a population with lower quality compared to the population we would have attracted if the program were not in place.

I have no data on this, and sincerely hope this is NOT true! What I have is a hunch that something could go wrong and we cannot detect it if we only look at the "In" column. But historically, hunches (or more formally "case studies") have led to both happy and miserable discoveries, when followed up by well-designed studies; two cases I used in my Happy/Misery Course are the (Happy) Viagra Trial and the (Misery) Fen-Phen Study. It is in this spirit that I suggest that we consider the issue of quality when we design further studies on assessing the effectiveness of the AP program. Such studies are not easy at all, but if there is any profession which has the most, and best, experts for designing and conducting them, it must be us. Indeed, other professions are paying high consulting fees to us for conducting these types of assessment studies for them. It would be ironic if we could not engage our own profession, with the same rigor, to address a problem that is directly about our own future!

Mis-take of Passion?

My initial explanation for the lack of acknowledging the "turn-off" issue in RPFHS is that they were misled by the non-trivially edited version (without my knowledge) as printed in *Amstat News* (September, 2009), where the quotation marks were removed from the aforementioned literal quotation from a student on AP courses being boring. It therefore could be perceived, particularly as a subtitle, as a depiction of the entire AP statistics program. Consequently, RPFHS serves, justifiably, as a defense of the AP program by providing positive examples. But after I raised such a possibility, the editor assured me that RPFHS was written as a reaction to my original op-ed (which contains five puzzles; see *http://www.amstat.org/publications/amsn/2009/september.cfm*).

That leaves me wondering if the authors of RPFHS and I have shared a similar type of "mistake of passion", that is, when driven by our passion for wanting to see a particular outcome, we forego careful consideration or critical thinking that otherwise are quite customary to us. The phrase "mistake" here is less a criticism of "mistake" but more an admonishment to ourselves that the human tendency of "mis-taking" passion in our effort to persuade is almost innate. Indeed, in a sense, it is my mis-take (or even mistake) of passion that has led to the current discussion, though I am hoping for forgiveness if the discussion ultimately leads to positive outcomes.

When THURJ invited me to write an op-ed, I said yes immediately even though I was completely overwhelmed by four courses and many other tasks. My passion obviously took over – where else could I find such an opportunity to potentially reach all current Harvard undergraduates, especially those who have already decided to stay away from statistics?

The article was initially circulated within Harvard, but the enthusiasm from readers soon encouraged me to make it available more generally. I was well aware of the selection bias in what one tends to hear, but when I read reactions such as the following, I told myself, well, why not let more students see it?

> "Just got a chance to read this… it's fantastic!!! I wish we had a way to circulate it to incoming college freshman all over the country—I feel like this piece alone would substantially increase the number of people taking statistics classes!!!"

Retrospectively, my passion for attracting as many students as possible blinded me to a completely obvious fact, that once an article is printed and especially circulated in general, it will not be read by students only! It simply escaped me that my intended dialogue with some Harvard students, that is, "I understand you had a bad AP stat class, but please give statistics another try" could be taken as an assertion about the entire AP program. I of course could have said "Sorry you had a miserable experience, but AP stat classes worked well for many others" – one can imagine how persuasive that would be. But I should have added a footnote when circulating it to emphasize the intended purpose of mentioning AP statistics in my op-ed. I therefore thank RPFHS for providing this opportunity for me to prevent any future inappropriate quotations of my two sentences on AP stat courses. In the same vein, the discussion that will appear in the next issue of *Amstat News* examines the issue of potentially mis-citing the study of the AP program by ASA and College Board (Patterson, 2009; hereafter ASA/CB) in order to prevent us from making more mistakes of passion, as surely we would all love, and are eager to see, programs such as AP statistics be a complete success! ■

Editor's Note: Due to space limitations, Part II will appear in the January issue of *Amstat News*. The full piece, however, is available now at *http://www.amstat.org/publications/amsn/2009/december.cfm.*