STEVE'S INFLUENCE ON MY WORK

VISHESH KARWA

Steve has shaped the lives and careers of many people within and outside the statistics community, over the course of his long career. I am very fortunate to be one of those people. Steve was unique in his ability to forge connections between seemingly unrelated fields. He was the "match maker" — connecting people and research areas as varied as privacy and statistical inference, algebra and statistics, copulas and log-linear models, sampling and design of experiments, networks and contingency tables and so on.

My approach towards research, in general, and in the area of privacy and statistical inference in particular, has been greatly influenced by Steve. I want to give a few examples to honor his memory. I first met Steve when I was a graduate student at Penn State University working with my PhD advisor Aleksandra Slavkovic. I gave a talk at a grant meeting explaining an idea to incorporate the additional randomness introduced due to privacy in the likelihood function for statistical modeling. I was just a graduate student, so my thoughts were not very clear. Steve immediately knew what I was trying to say, because he had been advocating a similar approach! The key goal of statistical inference is to make statements about population parameters, hence one needs to design privacy procedures with an eye towards this goal. This philosophy, which surrounds my work on privacy, is straight out of Steve's book! Another direction that I work on, that Steve often advocated, was the focus on finite sample inference as opposed to asymptotic inference. This is evident from his pioneering contributions to the work on sparse contingency tables where asymptotic tests don't always make sense. One of his pet peeves was the problem of analyzing sparse contingency tables under privacy, for which one necessarily has to take a finite sample viewpoint. Another point that always stays with me was Steve's suggestion to explore the possibility of a "Bayesian" version of privacy. While there are many such notions out there, I don't think they would have answered Steve's question. Quoting from my

(cont.)

not-so-good memory, Steve said something to the extent that "Differential privacy requires one to reason about all the datasets that one would have collected, and not the one we actually have at hand - this is similar in spirit to what is done in a frequentist setting, where one reasons about other datasets one could have seen, as opposed to a Bayesian setting where one conditions only on the observed dataset."

There are many other instances where Steve influenced my thinking, including other problems that I work on in network modeling, causal inference and algebraic statistics, all of which were Steve's favorites. In fact, I still channel Steve in my work and my talks - What would have Steve said, or done for this problem? One thing that used to happen a lot with me when talking to Steve was the following - he would say something to me about a research problem during our meetings, and I would come out of the meeting assuming that I understood what Steve was saying. But six months later, out of nowhere, I would suddenly realize what Steve was actually trying to say! This happens even now, and in fact I frequently go back to re-read his papers and my email conversations with him to get advice and insights. I truly miss him!

Vishesh Karwa