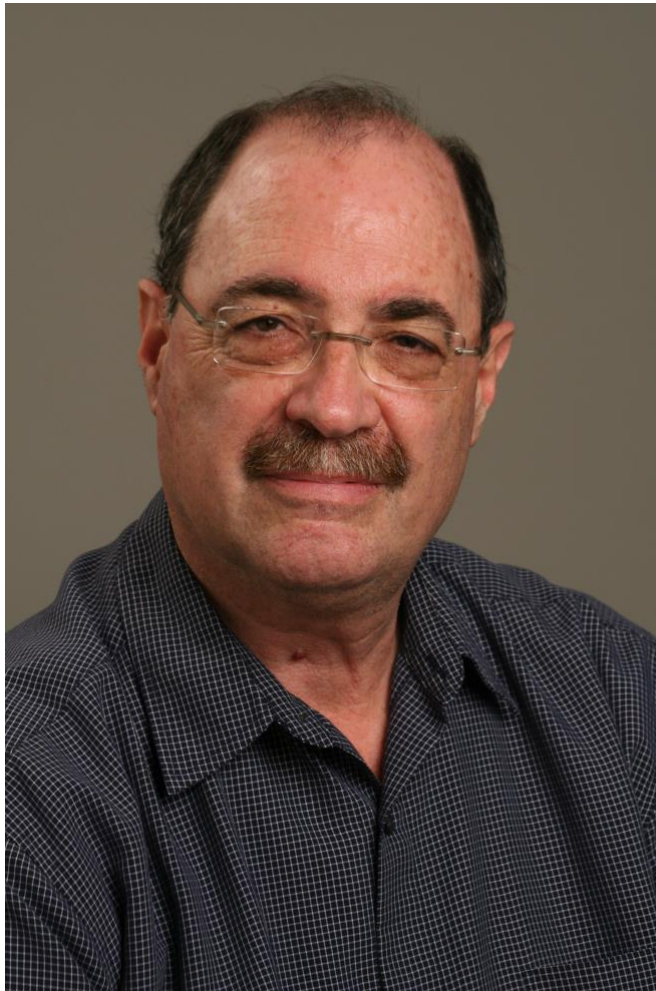**THE ET INTERVIEW:**

**PROFESSOR CHARLES MANSKI**


*Interviewed by Elie Tamer*

*Harvard University*

*Chuck Manski - Professor of Economics, Northwestern University*

*Chuck Manski has made vast contributions to the theory and practice of econometrics in its relation to economics in particular and quantitative social science in general. Chuck's career spans over 45 years of influential and important works. A defining characteristic of this work is its unusual and unchained creativity, thoughtfulness and clarity. His journey in the academy started with his thesis work on the maximum score estimation and discrete choice modelling which helped start the semi parametric literature in econometrics. Other important works include his work on choice sampling and social interactions, his important contributions to the collection and use of expectation data and his recent contributions to statistical decision theory. He is best known for his seminal work on partial identification. This approach to empirical work anchored by the identification problem has had a transformational impact not only in econometrics and economics but in statistics and quantitative social science. In addition, Chuck continues to be a superb advisor and mentor to many graduate students. His legacy as a thinker, researcher, and a teacher is an example in scholarhip that is hard to follow.*

*This interview was conducted in various places during 2017, including coffee and pastry shops, over lattes scones and canolis.*

**Choice Modeling**

**Elie**: This is the session on choice modeling.

The first topic I want to ask about is choice modeling. I want to go back to what I think is a landmark paper, the 1975 multinomial maximum score[1]. What is the intellectual inspiration for the paper and this approach to choice modelling?  And in particular, did you think that there were some concerns with McFadden's conditional <u>logit</u> in terms of its robustness to parametric restrictions which motivated this work?

---

[1] Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice. *Journal of econometrics*, *3*(3), 205-228.

**Chuck**: Okay. More broadly, one has to view the maximum score work in two stages. The 1975 paper is a standalone paper. The second round, in the 1985 paper, embodied an entirely different and more coherent type of thinking.

With regard to the 1975 paper there were two things on my mind. One is regarding logit. I think from the very beginning everyone recognized that the particular specification with the IID extreme value assumption leading to logit was purely for computational convenience. Dan McFadden was explicit about that. If you read his paper, the seminal paper in the Zarembka book[2], he lays out things very methodically, starting with broad utility theory and then the attribute characterization of utility functions, and then having unobserved variables, and then he gets to the end and he says we need this to be computationally tractable. And logit was computationally tractable. I don't think Dan had a particular fondness for extreme value versus normal. He could have done multinomial probit or something else, but it just led to a simple functional form. I think he saw this as just the beginning and not where the literature should end. Some evidence, and I have correspondence with him that verifies it, is in a letter he sent to me when I was a graduate student, where he made references to random coefficient models from the start. So even from the beginning he wanted to weaken the assumptions of the logit model.

That was partially on my mind when I came up with maximum score, but I think it probably wasn't my dominant concern then. What was going on---the paper was the 2nd chapter of my dissertation and I only began to work on it quite late, after the job market--- was that I was basically mucking around and trying to understand why maximum likelihood works. I wanted to see if maximizing other heuristically reasonable objective functions would work. One that seemed reasonable to me was to find parameter values that maximize the number of correct predictions when the unobserved utility components were ignored. That was the maximum score criterion. At that point in the early 1970s econometricians typically did not understand maximum likelihood and didn't understand estimation principles more generally. The entire literature that I had learned in graduate school was about linear models.

---

[2] McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior.

**Elie**: There is also the work of Amemiya

**Chuck:** This is prior to Amemiya. Amemiya plays a role in this, but everything we had learned was linear models, just linear regression and linear simultaneous equations. The only discussion of maximum likelihood in a Ph.D. econometrics course was that least squares is maximum likelihood in the normal linear model. We did not understand these things in any generality. With regard even to maximum likelihood, when Dan McFadden did his proof of consistency and asymptotic normality for the multinomial logit model, he had to do it from scratch. His proof strategies are not that easy to read. But he basically did an internal proof that this thing works.

You mentioned Amemiya. Amemiya's Tobit paper came out in 1973, which is basically the same time. He had a different way of proving things. He drew a result from the statistics literature, the paper of Jennrich in 1969, on consistency and asymptotic normality of nonlinear least squares. So the Jennrich paper was pretty new. As I recall, Amemiya looked at how Jennrich did that and there were some technical lemmas in the way Jennrich used uniform laws of large numbers and so on. Amemiya extended that work to do Tobit. So the McFadden work and Amemiya paper were basically at the same time.

This was totally new, and I found it amazing that you could estimate any non-linear model. This was a qualitative jump in the literature, to be able to estimate a non-linear parametric model. A non-linear parametric model---one might say big deal, everybody does that. But it's important to understand the history. When I teach econometrics to graduate students, I continually digress on the history and I say: "This may look straightforward to you, and you may think that everybody has known this forever, but this was actually done by certain people at certain times, and before that they just didn't understand these things." That was the situation.

**Elie**: At this stage, the statistics (non-econometrics) literature was further along in its developments of tools that can be helpful with nonlinear models.

**Chuck**: Econometrics was a little bit behind statistics, but not that much because the Jennrich paper was 1969. It was an *Annals* paper, if I recall. So uniform laws of large numbers, which

4

are the basic building blocks of everything today... They didn't exist earlier. We knew basic ones like Kolmogorov-Smirnov, but the uniform laws of large numbers that would allow you to do serious work on non-linear models were only coming into existence in the mathematical statistics literature in the 1960s and early 1970s. What I was doing was on my own, and I had essentially no formal statistics training. Anything I knew I had learned from Frank Fisher, who was wonderful on linear models. Frank was my advisor and was a very serious thinker, but his world was entirely inside linear models. That is what he taught, that is what everybody did.

**Elie**: You studied Math, Physics as an undergraduate, right?

**Chuck**: Yes, but that math was totally irrelevant. I was a physics major for a year, then a joint economics and physics major for a while, but the kinds of math that you learn in physics were just not relevant. I didn't get tooled up mathematically until later. This was when I was an assistant professor and spent a semester at Berkeley in the spring of 1976. Dan invited me there. By that time I realized I needed more math. I almost killed myself, I sat through, obviously not taking it for credit, but I sat through much of the first year PhD sequence in math at Berkeley.

**Elie**: That must have been fun.

**Chuck**: Not having to take it for credit was a lot easier, so I just tried to pick whatever I thought useful. This was topology and measure theory and a bit of functional analysis. The idea that an econometrician would need measure theory and functional analysis at that point was new… none of us had any background in that. Thinking about Dan McFadden, I don't know how he picked up the math when he was a student. Takeshi, I don't know how he got it. The people of my generation, we didn't learn any of this stuff in graduate school. You were never told you were going to need real analysis or anything like that. I learned some math after writing the first maximum score paper.

What I was doing, going back to the question of maximum score, was just mucking around asking why maximum likelihood works. In an intuitive way. I remember just playing. I said, well okay, the likelihood function has some intuition, but basically all the math was dealing with

the log likelihood, not with the likelihood function itself. I said, well, you look at the log likelihood function, you maximize it, so the log likelihood function is just taking a monotone transformation of the probabilities, the log of the probabilities, and it was heuristically intuitive that you'd want to maximize some monotone function of the probabilities. And I started asking what other monotone functions might work? Instead of maximizing the log of the probabilities, why not maximize the sum of the probabilities? I couldn't make that work, so I started thinking about other ideas, playing around with different functional forms. Looking at the structure of the discrete choice model, I decided to focus on the assumption of IID disturbances. I wanted to get away from the extreme value distribution, but I felt comfortable with IID disturbances at the time. I saw that the IID assumption implied that the higher $X\beta$ is in the systematic part of the utility function, the higher the choice probability is going to be.

At some point I thought---why not use that directly? And that's the score, right? That is, if you have a higher value of $X\beta$, you should have a higher choice probability. And I thought, let's try that. And then I wrote down the basic relationship and asked what objective function $\beta$ is going to maximize. Give yourself a point if you predict correctly. Or don't give yourself a point – that's the score - if you don't predict correctly.

Actually, in the original version of maximum score, it was not just zero-one, that you get one point versus zero. In the 1975 paper, it's any ordinal transformation, an idea that no one has ever used since then as far as I know. The basic reasoning concerned the ordering of the probabilities – as long as you have any function that makes the ordering of the probabilities the same as the ordering of the $X\beta$, it will work.

**Elie**: So the nice insight is that all you needed was that the ordering of the probabilities somehow matched the ordering of the index and that is what makes it work.

**Chuck**: Then it will work.

That was the fundamental thing. And I said, oh that works! I mean you could see the basic thing that was driving it is the IID assumption or conditional IID. I was careful with that. It didn't have to be IID across people, it just had to be IID …

**Elie**: Within the choice set, and this is key.

**Chuck**: Within. I think a lot of people miss this. That was a subtlety that I was proud of. And that you didn't need any particular distributional assumption at all. I wrote it up. That's basically the identification part.

**Elie**: Did anybody understand it at the time, other than you?

**Chuck**: I sent it to the *Journal of Econometrics*. Amemiya was the editor, and he accepted it. I don't have any recollection of the referee reports. I don't remember the process.

**Elie**: I took a look at the paper a few days ago. It is a little difficult to penetrate. The way it is written.

**Chuck**: There are some things on the statistical side that are just wrong.

**Elie**: This is in contrast to the 1985 version[3] which is beautiful. The 1975 paper needs a little thinking to penetrate. Forget the technical part.

**Chuck**: I think that's right. My ability to write has gone through stages. It was not very good then.

**Elie**: But it was not your job market paper?

---

[3] Manski, C. F. (1985). Semiparametric analysis of discrete response: Asymptotic properties of the maximum score estimator. *Journal of econometrics*, *27*(3), 313-333.

**Chuck**: No. The job market paper was the college choice paper. The Kohn-Manski-Mundel paper.

**Elie**: That gets me to the next question. Were you bothered at the time when you were writing it that this maximum score was only able to get at the index parameters, the betas?

**Chuck**: No. I know that in the more recent literature people say--who cares about the betas? You care about prediction. But within the model I was working with, the ordering of $X\beta$ gave the ordering of the choice probabilities. Having the ordering of the choice probabilities would allow you to do partial predictions of choice behavior. This was quite important to me later on in my paper "Identification of Binary Response Models" in *JASA* in 1988.[4] If you read that paper, you see that the interpretation of $\beta$ was important to me, particularly what you can get with $\beta$ in terms of being able to extrapolate and do counterfactual prediction. It depended upon which framework you're working in. In the case of binary response I was working with quantile independence. With median independence, you get a very simple ordering. If $X\beta$ is greater than zero, the choice probability is greater than one half. That allows you to make a real prediction.

In contrast to that, what was happening in the mid-1980s and getting lots of attention were the linear index models. Powell, Stock, and Stoker, and all the literature on that. I thought that was really interesting as a piece of mathematics, but not so useful for prediction. The linear index structure is just a local derivative property, so you couldn't use it for prediction. The interpretation of $\beta$ within that setup is very different than the one I was working with. That was important to me from the beginning.

**Elie**: But still, you're not going to be able to get average effects by just getting this ordering.

**Chuck**: Well, we're going to talk later about partial identification. Maybe it wasn't entirely clear at the time, but there is partial identification of average effects embedded in maximum score. What you learn about $\beta$ suffices to get an ordering of the choice probabilities, but not to pin them

---

[4] Manski, C.F. (1988). Identification of binary response models. Journal of the American Statistical Association, 83(403), 729-738.

down. So there is partial identification of the choice probabilities. I think to this day that it is a useful form of partial identification, including for the multinomial case.

**Elie**: Certainly, there is a connection. In modern day prediction or classification in this case, the maximum score is related to support vector machines. They predict even if you have hundreds and hundreds of alternatives. The best predictor is the one that has the highest choice probability. That is what you can actually do in practice.  You get the choice probabilities and you see which one is highest.

**Chuck**: That motivated maximum score from the beginning.  It was not only $\beta$ per se.

**Elie**: But contrast that then with the next paper that I would like you to talk about, your joint paper with Kohn and Mundel[5].

**Chuck**: It started with David Mundel. I was a senior as an undergraduate at MIT. He was a political science PhD student at MIT. I had been working as an undergraduate for an organization of independent colleges in Massachusetts. I was doing analysis for them on issues having to do with college going. I worked there one summer and during the academic year when I was a senior. David Mundel was studying education.  He was a political scientist but he was close to economics. Somehow we came into contact and we had this idea of studying the effect of federal financial aid policy on college going, or in general what determines college going. Then I became friendly with Meir Kohn, who was a Ph.D. student with me at MIT starting in the same year, the fall of 1970.

**Elie**: So, you were all students together?

**Chuck**: Yes, we were all students. We were trying to model demand for higher education. That was the term we were using at that time: "demand for higher education."  We started out trying

---

[5] Kohn, M. G., Manski, C. F., & Mundel, D. S. (1976). An empirical investigation of factors which influence college-going behavior. In *Annals of Economic and Social Measurement, Volume 5, number 4* (pp. 391-419). NBER.

to use neoclassical demand theory.  It didn't make sense, but it was what we knew. We were uncomfortable with neoclassical demand because the issue wasn't how many units of schooling you are going to buy. It was rather what kind of school you choose, and whether you go to college or not. That is, do you go to Harvard or MIT or Northwestern or wherever?  It was a qualitative rather than quantitative choice. But there was no literature at the time.

And then, some things are total serendipity. Dan McFadden was visiting MIT when I was a first-year student. This was the spring of 1971. There was nothing published on multinomial logit yet. Dan gave a short course and I sat in on it. First year students don't normally sit in on short courses given by faculty, but I did. And I said, "Oh! This is the way to do it." And so David, Meir, and I gave up the old stuff we were doing and we got excited.



*Chuck as head of IRP in 1992 participating at a conference in Washington DC.*

**Elie**: Back to the paper. It is striking in style relative to the maximum score.

**Chuck**: It's all parametric. Of course.

**Elie**: You would think it was written in the mid-1980s. It reminds me of Taber's job market paper on college choice with a two-stage choice problem with explicit modeling of choice sets.

**Chuck**: I suffered grievously. I could not get a job on the market. Period! I mean, I got nothing!

**Elie**: This was your thesis. And this is your job market paper.

**Chuck**: This was my job market paper. I became familiar with what Dan McFadden was doing and I eventually got comfortable with parametric maximum likelihood and discrete choice. But other economists found the conceptualization of discrete choice difficult, not just the technical aspects.

The idea of thinking of goods as attribute bundles would seem obvious. It wasn't entirely original to Dan. I talked with him about this and he told me he was influenced by the work of Zvi Griliches on hedonic price indices and perhaps some other ideas. And then maximum likelihood. I'd go on the market and I'd give seminars and people would say, "What's the dependent variable?" I said, "Well, a choice." But unless you can write it as $y = X\beta + \varepsilon$, people just didn't understand. I must have given pretty bad seminars.

**Elie**: But the paper went even beyond that. You were worried about modeling the choice set which people don't even give that much thought to today.

**Chuck**: Yes, the choice set generation process. That turned into the third chapter of my thesis, on "The Structure of Random Utility Models[6]," which laid out the idea of choice set generation. That paper is extremely hard to read too.

**Elie**: But it has a lot of citations.

---

[6] Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, *8*(3), 229-254.

**Chuck**: That paper is published in *Theory and Decision*. I think it may be the most cited paper in the history of that journal. To this day! So, the notion of choice set generation has mattered. That came out of the original work with Kohn and Mundel in the particular context of college choice. Then I thought about it abstractly in the third thesis chapter. But the Kohn-Manski-Mundel paper was rejected everywhere.

**Elie**: Is that because people did not understand it or were not comfortable with the modeling style?

**Chuck**: Again I'd have to go back and look at the referee reports.

**Elie**: Even at JPE?

**Chuck**: I remember being rejected at *AER* and at the *Journal of Human Resources*. I think we tried *JPE* too, but I'm not sure. We tried four or five places. The only reason it ever got published was there was this special issue of *Annals of Economics and Social Measurement* that Dan co-edited with Jim Heckman.

**Elie**: It's actually quite hard to get.

**Chuck**: It was a special issue, it was an obvious pick for that special issue, and it got published there. The paper might never have been published otherwise.

**Elie**: Can you talk about the 1985 version of the Maximum score paper?

**Chuck**: Okay, so now we have to move forward. After I wrote the first maximum score paper, I didn't think about it, and I don't think anyone else thought about it, until the mid-1980s.

**Elie**: The Wisconsin days?

**Chuck**: Until the Wisconsin days. Maximum score in 1975 was a standalone paper. I did not try to extend it afterwards. I didn't really think much about it. My attention turned to other things. The work that had more success was on choice-based sampling, which was soon after 1975. That was accepted quickly in E*conometrica*.[7] People understood it. I was also doing applied work in transportation---automobile choice modeling and other things like that. But maximum score…

**Elie**: Essentially you were doing what we would call structural choice modelling.

**Chuck**: Yes, I was doing structural parametric work. I know exactly when that changed. It began to change when I was out of the country, in Jerusalem. That was the very beginning of the semiparametric literature. I called the second maximum score paper "Semiparametric Analysis of Binary Choice." I used the word "semiparametric" because in 1982 or 1983, there had been a paper using that term, I think by John Wellner. It was the first time I'd seen the word semiparametric. And I thought, "Ah! That's what maximum score was doing." It was semiparametric. I decided to de-emphasize the maximum score criterion and emphasize the semiparametric aspect. I explicitly titled my paper "semiparametric," drawing the word semiparametric from the statistics literature, which was just coming into existence in the early 1980s. By that point, I was much more keyed in. I knew some of leading statisticians by that point. I met Wellner a few times. I was close with Peter Bickel, who had spent a semester in Jerusalem when I was there. He was doing adaptive estimation, which was also semiparametric. Following Peter, I wrote a paper on adaptive estimation of nonlinear models. So I was getting into that frame of thinking--- of trying, in a coherent way, to get away from parametric assumptions.

That was what motivated me in the second round of maximum score. I decided to go back and to look at the whole problem again, having in mind that at this point I knew more math than I did before. And I was older and a little more mature about it. But what really influenced me heavily was Jim Powell's censored LAD paper[8].

---

[7] Manski, C.F. and S.R. Lerman (1977). The estimation of choice probabilities from choice-based samples. Econometrica, 45(8), 205-228.

[8] Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, *25*(3), 303-325.

**Elie**: Yes, this is Jim's work in the early 80's on nonlinear models with quantile restrictions.

**Chuck**: I knew Roger Koenker's paper on regression quantiles. I didn't make connections there, because that was linear regression quantiles. But seeing what Jim could do for the Tobit model…

**Elie**: That's interesting.

**Chuck**: Right.

**Elie**: Especially that one is able to do that without parametric assumptions on the unobservables.

**Chuck**: Yes. And I remembered that in my 1975 paper, there is a small section specific to binary response that shows maximum score works with a median independence assumption. It is in the 1975 paper, but I had not drawn any conclusion from it… It was just a side comment as a special case. There was no open identification issue anymore when I began the 1985 paper. The open issue was to be able to say something formally…

**Elie**: Now moving towards consistency.

**Chuck**: Yes, consistency.   By that time, there were uniform laws of large numbers that I could draw on. And remember that this was with a step function, which is not trivial.

**Elie**: Even today.

**Chuck**: By that time, I could search the *Annals* type statistics literature, and I found a uniform law of large numbers that I could use, by Ranga Rao in 1962. So my motivation was to go back to maximum score and see if I could pin this down, because I knew I hadn't proved consistency in the first paper …

**Elie**: As a big picture comment, it is safe to say that the maximum score work launched this field of semiparametrics in econometrics. So, in the 1980s and 1990s, we see ingenious papers that provided insightful results on approaches to combining stochastic restrictions, support conditions and some functional form assumptions to get point identification in a wide variety (of mostly nonlinear) models. Also, the estimation approaches were non-trivial. But, on the other hand, it is also safe to say that this literature has not had as much impact on empirical work directly. Do you think it is a problem?

**Chuck**: I very much view it as a problem. In fact, to jump to what we will talk about later, I see a somewhat similar problem with partial identification work. In both cases, the literature has had much more influence on econometric theory than on empirical work. This bothers me quite deeply. One of the things I began to realize in the 1980s was that the motivation for doing econometric research differs across people. My motivation, Dan McFadden's motivation, Jim Heckman's motivation, and I would say Gary Chamberlain's to some extent, was always to do research that would be useful for empirical work. I would put Jerry Hausman in that category also. We all viewed ourselves as economists first, with econometrics meant to be in the service of economics. And so, if you're doing econometrics that isn't going to be useful to economics, then who the hell cares?

In the 1980s, a different breed of econometrician started developing who were much more like mathematical statisticians, who were more intrinsically interested in the theory. So the semiparametric field developed mainly with internal motivations rather than aiming to influence applications. One could get papers published in *Econometrica* whether or not they had empirical applications, provided that the econometric theory was good. This became self-perpetuating. So how much has the semiparametric field influenced empirical practice? Very little.

**Elie**: There is also a feeling in economics that the low hanging fruits and that all major open questions have been worked out.

**Chuck**: The question is whether that's true. One could say that the low hanging fruit are gone and all that's left to do is the more technical work. On the other hand, the basic ideas in partial identification are extraordinarily simple – like the missing outcomes problem. It was no more than high school math to do the early partial identification work. Obviously, there is some very difficult math as you get deeper into it but, in terms of the basic ideas, there was unrecognized low-hanging fruit in partial identification. There could be unrecognized low-hanging fruit in other areas too. I can't say that we've solved all the easy problems in the world and all that's left to do is to do tougher and tougher math.

**Elie**: I hope there are still many low hanging fruits!

**Chuck**: Think about your own job-market paper, your *Restud* paper[9] and everything that followed from that. Of course, there are technical issues, but the basic conceptualization is simple. It could have been recognized much earlier, but it wasn't.

**Partial Identification**

**Elie**: Getting to partial identification now, let me just start with the obvious question first. It's fair to say that among the many big hits of yours, this is the biggest. Do you agree?

**Chuck**: Absolutely. I can partition my career into the period before 1988/89, the beginning of the work on partial identification, and afterwards.

**Elie**: Talk about what was going on in your mind right around when you were writing the JASA paper on the identification of discrete response models.

**Chuck**: In retrospect, there was a formal piece of partial identification analysis in the JASA paper. Also in retrospect, there was partial identification in the maximum score work, as the

---

[9] Tamer, E. (2003). Incomplete simultaneous discrete response model with multiple equilibria. *The Review of Economic Studies*, *70*(1), 147-165.

assumptions I made there partially identified choice probabilities. However, the concept wasn't consciously on my mind until a specific event in the spring of 1987. My conscious study of partial identification starts with a question from Irv Piliavin.

**Elie**: You actually thank Irv in the *Journal of Human Resources*[10] paper.

**Chuck**: Yes, and I also thank him in a couple of the books that I've written, in the preface. This is why I have found it so important to have real-world problems that motivate the econometrics. I was at Wisconsin. Irv was in the social work school and he...

**Elie**: Is he still alive?

**Chuck**: No, unfortunately. He died in the fall of 2009, a month before Art Goldberger died.

Irv and Art Goldberger were best friends. I got to know Irv through Art, because he was always coming to talk to Art and then we became friendly too. Irv was in social work. He didn't have a technical background. He did empirical work.

In the mid '80s, he was working on transitions of homelessness. He was studying a homeless population in Minneapolis and he had a specific problem. He had longitudinal data and wanted to estimate a transition probability. Conditional on being homeless at a particular date, what is the chance of being homeless six months later?

**Elie**: Descriptive duration model.

**Chuck**: Yes. He had a problem of missing data. It is hard to get a random sample of homeless people and follow them. He went to Minneapolis and obtained what he thought would be a random sample of homeless people. He tried to re-interview them six months later and he couldn't find some of them. He had attrition in a longitudinal study and was trying to figure what to do.

He came to me and he said: "There are only two ways I know to deal with this. One is to assume that attrition was random, but I don't believe that. The other is to assume a parametric selection

---

[10] Manski, C. F. (1989). Anatomy of the selection problem. *Journal of Human resources*, 343-360.

model, but I don't believe that either." Although Irv was not a technical researcher, he appreciated the issues conceptually. He then asked me: "Isn't there anything else that I could do?" He gave me an incentive. He came up with a month of summer salary.

**Elie**: You were senior faculty at that time.

**Chuck**: I was senior faculty but the incentive still mattered. I had some NSF money but I didn't have all of my summer covered.

**Elie**: No. It's funny that's even on the table.

**Chuck**: Well, he had the funds from grants, I suppose. He asked me to think this through carefully, not just superficially. When I began to do so, I realized that Irv's problem was just one instance of a very broad problem.

I had never worked on the selection problem. Of course, I knew the literature. I knew all of the arguments about the fragility of parametric selection models.

**Elie**: You were surrounded by people who have.

**Chuck**: Yes, Goldberger had written about this.

I had seen this all the time I was growing up. I don't know how many seminars I attended in which selection bias was a concern, but I hadn't given it much thought.

That summer I started fussing around, the way econometricians typically do. You start with a tight model that is point identified and then you try to weaken the assumptions. That can be fruitful. I don't want to argue against work of this type. I began with parametric selection models and tried to think about weakening them to semiparametric or nonparametric ones. This was 1987, so some work of this type was already being done. I decided that thinking in this way would not give a good answer to Irv's question. I wanted to do something that would be useful to him.

At some point, I stepped back and thought: "Let's go back to basics." This is a way of thinking that I have used multiple times in my career when I find myself stuck, and I am proud of it. I

made it the theme of an essay I wrote in the *American Economist* called "Unlearning and Discovery."[11]

**Elie**:  Yes, I am familiar with this.

**Chuck**:  This is a prime example of what happens when a literature develops. You develop certain ways of thinking about things and your brain becomes wired to think in those ways. That can be useful.  You make new contributions built on earlier ones, thinking in a particular way.

However, sometimes building on the literature does not work. There are times when it is more fruitful to step back and say: "Imagine I'm a baby. I don't know anything. Looking at this problem from the beginning and supposing that I know nothing, what is the essence of the problem?"

I wrote down the Law of Total Probability and had a revelation.  Suddenly the issue was clear. We don't know anything about the distribution of missing data.

**Elie**:  Did you instantly recognize this? It's a nice insight. Is it possible that you may have just handed the work to Irv and moved on.

**Chuck**:  I could have just gone on to do other things, but I decided not to. Once I wrote down the Law of Total Probability, I thought, "My God. This is the way the selection problem should have been viewed from the beginning. Why wasn't it? Maybe there's something wrong with my reasoning." But there was nothing wrong.  The bounds could be wide but the logic was correct.

Sometime later, I saw the connection to the problem of inference on treatment response and the fact that counterfactual outcomes are missing data.  Don Rubin had made the connection between counterfactuals and missing outcome data much earlier and he deserves credit for it. However, for whatever reason, he felt it essential to obtain point estimates and was willing to make assumptions strong enough to get them.

---

[11] Manski, C.F. (2010) Unlearning and Discovery. *The American Economist*, 55(1), 9-18.

The reason why I didn't just go on is that I decided it was very important. After I did the work for Irv that summer, he used it for his work on homelessness in Minneapolis to some extent. I sat on it for six months without writing a paper.

**Elie**: This is the *JHR*?

**Chuck**: This is the *JHR*. Robert Moffitt, who was the editor of the *JHR*, asked me to write it. Robert wanted to invite papers on econometric methodology that would be helpful for applications.

What was striking is that before I wrote the *JHR* paper, there was a considerable period before I was willing to talk about the work in public. The idea was extremely simple, but I could easily forecast that it would not be well received. I eventually did go public at a conference that Moffitt organized on "kinkometrics;" that is, the use of parametric structural models to study labor supply with kinked budget sets arising from progressive income tax schedules.

The conference was in Wisconsin at Wingspread, a conference center east of Madison. Everyone who worked on the topic was there, including Hausman and Heckman. They were using parametric structural models in which labor supply decisions involved selection of a segment of the kinked budget set, which determines the marginal tax rate you face. They asked me to discuss a paper, but I don't remember whose. I decided to use my discussion to lay out my ideas on how to approach the selection problem.

**Elie**: Dangerous.

**Chuck**: Yes, dangerous. First, what I had to say was simple, which would make it seem trivial to some. Second, I did not yet have a working paper, just slides.

Well, what happened was that I got screamed at simultaneously by Hausman and Heckman. I then knew for sure that what I had done must be good. I don't remember what Heckman said, but I do remember what Hausman said and I have since quoted him by name in my 2013 book *Public Policy in an Uncertain World*[12]. Jerry said: "You can't give the client the bound. The

---

[12] Manski, C. F. (2013). *Public policy in an uncertain world: analysis and decisions*. Harvard University Press.

client needs a point." He still believes this today. He is not upset with me for quoting him because that was and still is his view.

It took me six months to get to this point. Even though I was sure that I was right, I still worried about how to write the idea up. I thought, "How can I write a paper on something that is so simple? I'll be laughed at to write a paper on something that's this simple." Even then, the way you got papers published in econometrics was by doing tough proofs. I thought that if I write up something so simple, I won't be considered an econometrician. I did write it up for the invited *JHR* paper. This made things easier because I did not have to deal with the usual refereeing process. The more I thought about it, the more important I thought the idea was.

Of course, there were precedents to partial identification that were already in the econometrics literature. Probably the earliest was in the 1930s, the Frisch bound concerning linear regression with errors-in-variables. There was a little bit on partial identification in Frank Fisher's 1966 book on *The Identification Problem in Econometrics*, where he briefly considered identification of linear simultaneous systems with sign restrictions on parameters. There were other early precedents that you cited in your review article in the *Annual Review of Economics*[13]. However, there was no coherent body of work, just a scattering of findings.

**Elie**: Ed Leamer's work comes to mind.

**Chuck**: Leamer is the most interesting case by far. I have enormous respect for Ed. When I was a graduate student in MIT, he was an assistant professor at Harvard and gave a course on Bayesian econometrics that I attended. I learned quite a bit from him.

In the early 80s he wrote several papers that we would now call partial identification analysis. His interest focused on identification and estimation of linear regressions in the presence of classical measurement error and other data issues. In the mid '80s he basically stopped doing econometrics and changed careers, to work in international trade.

**Elie**: He had some papers on generalizing the Frisch bounds.

---

[13] Tamer, E. (2010). Partial identification in econometrics. *Annual Review of Economics*, *2*(1), 167-195.

**Chuck**: That paper was joint with Klepper. It is a nice paper in *Econometrica* in 1984[14]. They generalized the Frisch bound with errors-in-variables from simple regression to multivariate regression. Soon after, Ed stopped doing econometrics completely. Why? I don't know. He could have developed this further and he decided not to. I don't know why he didn't.

Regarding other precedents, I learned four or five years after my *JHR* paper that in 1954 Cochran, Mosteller, and Tukey[15] had done a simple analysis of missing outcome data in their book that evaluated the statistical methodology of the Kinsey Report on sexual behavior. They focused on estimation of the probability of a binary response in the survey that Kinsey used. They framed the issue not directly in terms of identification but rather as the mean square error of estimates of the probability when some outcome data are missing. They recognized that randomly missing data affect the variance of estimates but generate no bias issue. On the other hand, missing data generates potential bias in the absence of assumptions about the missingness process. They made a nice observation with practical importance, this being that the bias reduction from converting a non-respondent to a respondent often is a much better use of survey money than increasing the sample size, which only reduces variance. However, they never followed up on it. In fact, Cochran later wrote in his textbook that the bounds were distressingly wide and he didn't think them useful.

**Elie**: Is this Cochran the same sampling Cochran?

**Chuck**: Yes, it's the same sampling Cochran. I think he was at Harvard.

**Elie:** There is also a 1989 paper by Robins[16] that uses bounds.

**Chuck**: Jamie and I independently got the same result on intersections bounds for an average treatment effect. Our derivations made different assumptions. He was writing about randomized experiments with non-compliance and I was writing about use of observational data with an

---

[14] Klepper, S., & Leamer, E. E. (1984). Consistent sets of estimates for regressions with errors in all variables. *Econometrica: Journal of the Econometric Society*, 163-183.
[15] Cochran, W. G., Mosteller, F., & Tukey, J. W. (1953). Statistical problems of the Kinsey report. *Journal of the American Statistical Association*, *48*(264), 673-716.

[16] Robins, J. M. (1989). The analysis of randomized and non-randomized AIDS treatment trials using a new approach to causal inference in longitudinal studies. *Health service research methodology: a focus on AIDS*, *113*, 159.

instrumental variable assuming mean independence, this being in my 1990 AER Proceeedings article.[17]  However, the results we reported were algebraically the same.  My result was a sharp bound under the assumptions that I made.  Jamie's setup made stronger statistical independence assumptions and can yield a tighter bound, as Balke and Pearl showed later.  However, Jamie's analysis did not use the full power of statistical independence, so he obtained the same bound that I did. The work that Jamie and I did was independent and essentially simultaneous. He and I didn't know each other at that time.

Another precedent, which I now teach regularly, is the ecological inference problem, which is classic. It is a different identification problem. Duncan and Davis[18] derived a bound in the *American Sociological Review* in 1953. It's beautiful and simple. Yet another is the Peterson bound for the competing risk model, in 1976[19].

**Elie**:  It is a good paper but I think his result, at least the way I interpreted it, is saying any competing risks model is observationally equivalent to a competing risk model with independent arms.

**Chuck**:  The biostatisticians concluded that you might as well assume independence, but that's not the conclusion I would draw.

**Elie**:  Yes. He didn't say that exactly, but increasingly, in statistics you see statements like "Why do you care about correlation? Peterson showed us that it's equivalent to independence," which I think misses the point.

**Chuck**:  It misses the point. There are other literatures that are close but are not the same. When Joel Horowitz and I wrote our paper on contaminated sampling, we went back to Huber and the robust statistics literature on gross errors.  It is not formal identification, but it has commonalities.

---

[17] Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, *80*(2), 319-323.

[18] Duncan, O. D., & Davis, B. (1953). An alternative to ecological correlation. *American sociological review*, *18*(6), 665-666.

[19] Peterson, A. V. (1976). Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks. *Proceedings of the National Academy of Sciences*, *73*(1), 11-13.

**Elie**:  So you just stuck with it.

**Chuck**:  Yes, I stuck with it. I stuck my neck out, over and over. I didn't get much positive reinforcement.  I was able to get my work published, sometimes with difficulty, sometimes more easily, all through the '90s. However, few people paid attention to it. Even so, I stuck with it. It helped a lot that I had incredible support from Art Goldberger.  I also had good support from Joel Horowitz and from a few students I advised at Wisconsin, particularly John Pepper.

**Elie**:  I suppose in some ways it's a good thing, because people left you alone.

**Chuck**:  Yes, I had a whole field to myself. I had ten years to harvest low-hanging fruit. That's right, but it was not easy. It was lonely. I kept wondering: "Are they right, and I'm wrong?" I kept asking whether others may be right not to be interested in the work. Maybe I should stop doing it?  But I kept going back to the basic logic and concluding: "No. This makes sense."

Seminar participants and referees were always saying, "You're making no assumptions. Nothing in, nothing out."  I would keep pleading against this simplistic view, saying: "The idea is to start with weak assumptions and see what you get." I would say over and over: "But that's only the beginning. You start building up, add assumptions, and analyze the identifying power."  I recommended what I think is a coherent research strategy, presenting a menu of results. You start with weak assumptions, add assumptions and see how the bounds shrink, and you finally get to the point where you get point identification.

This is explicit in my 1997 paper in *REStud* on the mixing problem[20].  I used the Perry Preschool Experiment to illustrate how adding assumptions shrinks the bounds. There's a coherent strategy going from weak assumptions to strong assumptions. I thought "Doesn't this make sense?" So, I kept doing it.

**Elie**:  The strange thing, too, is during that time, there is the IV Revolution which is motivated similarly by concerns with assumptions.

---

[20] Manski, C. F. (1997). The mixing problem in programme evaluation. *The Review of Economic Studies*, *64*(4), 537-553

**Chuck**: That's an issue that bugs me deeply. It particularly bugs me when I sometimes go to Labor Week at the NBER Summer Institute. I look at what goes on in these NBER labor papers and how everything is framed as a local average treatment effect. There is hardly a hint of partial identification work in that literature.

You've been more successful in the IO literature, right?

**Elie**: Yes.

**Chuck**: You have to look at these two empirical literatures entirely separately from one another. The empirical IO people have been far more sympathetic to the multiple equilibrium work that you've done.

I've never been invited in 25 years to give a seminar to any of the applied groups -- of course, I give econometric seminars, but not to any of the applied groups -- around MIT, Harvard, Princeton, and so on.

**Elie**: Isn't this curious, because your motivation is robustness. They motivate their literature also from the same place.

**Chuck**: The basic paper for my work is the 1990 paper. The very simple idea about...

**Elie**: The papers and proceedings?

**Chuck**: The papers and proceedings. It is a simple, short paper. Basically no one in the NBER labor group pays attention to it. The paper gets citations from econometricians, but if you were to compare its citations or the one on monotone instrumental variables to that of Imbens and Angrist, there is an order of magnitude difference.

The empirical people who are doing labor, health, education, applications, development, and so on, to this day, over twenty-five years afterwards, they don't even know this work exists.

We can act as empirical economists and say this is the revealed preference of the research community, that they'd rather use LATE than understand analysis of treatment response more

fundamentally. When the LATE paper came out, it was published as a note in *Econometrica*[21]. I thought it well deserved to be published as a note. It was a cute mathematical result. I thought it was a one-off thing. I had published four years earlier the IV bounds on average treatment effects and I thought, "Well, it's clear the IV bounds are far more useful and interesting than this LATE thing."

**Elie**: You mean the papers and proceedings?

**Chuck**: Yes, the *AER* papers and proceedings. In 1995, I wrote the book *Identification Problems in the Social Sciences* in like a baby language, to try to explain this for a general audience. I thought that the matter was clear, that my IV bounds would be viewed as revolutionary and that LATE would be viewed as a technical footnote. Clearly, that is not what happened. Now, 20 years later, the dominance of LATE has become stronger and stronger.

Of course, people in econometrics know the IV bound, and it has generated the inference work on intersection bounds by Kitagawa and by Chernozhukov, Lee, and Rosen. It has generated interest as a mathematical problem, but not in terms of application. I don't get it.

**Chuck**: Let me push on this a little bit. We were talking earlier about the problem of econometrics getting too technical. Empirical people want to stay away because it's too technical and they see these regularity conditions that they can't interpret. However, my papers on bounds on treatment effects can't be blamed for being too technical. The ideas are extremely simple. There may be some intricacies regarding how to do confidence sets, but that is a second order issue. I really try to write for a wide audience. I wrote the 1995 book[22] and the 2007 textbook to be easily accessible. There's really no excuse. The empirical fact is that the NBER labor economists act as if the bounds literature doesn't exist.

**Elie**: The only reason that I can see is this, they're just obsessed with numbers, points. They really would like to get a number.

---

[21] Imbens, G. W., & Angrist, J. D. (1994). Identification and Estimation of Local Average Treatment Effects. *Econometrica*, *62*(2), 467-475.

[22] Manski, C. F. (1995). *Identification problems in the social sciences.* Harvard University Press.

**Chuck**: There clearly has been an obsession with points, but it is frustrating to be left with this as a residual, psychological answer, in the absence of a reasonable logical answer. I knew the obsession with points from the beginning. People thought you had to have a point estimate. Jerry Hausman said it explicitly at the Wingspread conference in the late '80s. I agree that one of the reasons why partial identification has been resisted is an obsession with points. This remains a common view in some fields even today.

It is odd, in a way, because researchers are comfortable with confidence intervals to measure statistical imprecision in contexts with point identification. Probability theory gives many commonly used bounds, such as the Chebychev inequality, the Frechet bounds, and large deviations inequalities.

**Elie** Yes, it is the same conceptual problem as trying to use a confidence interval.

**Chuck**: So what's going on? Why is it that every empirical person is willing to recognize sampling uncertainty in a standard error or confidence interval, but is not willing to recognize partial identification?

**Elie**: Here's a question that I get asked. Many people in IO particularly say, "Well, I do have the structural model and I've got a whole list of assumptions. I want to take this partial identification approach. What should I do?" One approach is to just, you had mentioned this earlier, start with this fully structural model and somehow remove some of the assumptions in some ways, and then see what happens.

Admittedly, I guess this top-down approach is rather mechanical, but I think it's extremely important. You start from a fully structural model, take your paper, let's say, with Mundel and say you're worried about some assumption such as the IIA, or another assumption that seems to be of concern. This becomes a problem: which assumption to relax?

**Chuck**: As a methodological question, I've tried to be open minded about this. I know we're talking about the origins of partial identification, but I can date my concern about what assumptions to make and what you get out of them to several years earlier. When we were talking about maximum score, I said my second round of work was influenced by Jim Powell and the LAD idea. That had a specific influence because of the quantile independence

interpretation of maximum score, but there was another event around the same time that, in a much broader way, shook me up quite a bit back then. This was the discovery of nonparametric regression.

I know how I learned it, which is that Herman Bierens brought it into econometrics. It obviously wasn't original to Herman, but he did the technology transfer from the statistics literature. Herman got asked to give an invited paper in the 1985 Econometric Society World Congress. I was asked to be a discussant of his paper. He gave a survey paper on kernel estimation, which was new to economists at the time. In my discussion, I tried to think about what he was doing versus linear regression.

I came up with a metaphor of a production possibility frontier for econometrics. At one end, you can make strong assumptions and get strong results, but maybe you don't believe them. At the other end, you can make much weaker assumptions and pay in terms of weaker conclusions, but have more credibility. And I said that there is a whole middle ground between these polar cases.

I said at the World Congress that there is a vast frontier for econometricians to explore, as you vary assumptions and see what you get from them. That idea became important to me later and I eventually came to call it the "Law of Decreasing Credibility" in my 2003 book on partial identification.[23]

I reasoned that there are two ways you could approach the issue methodically. You can start with a strong model and then weaken it in this and that direction, and see what you can get from that. Or you can start at the other end, and then add assumptions.

My view is that you can learn from both approaches. The view that you were just expressing is starting with a strong structural model and then weakening it in different directions. I think you can learn a lot from that.

That's the way I view the flexible functional form literature and some of the work on discrete choice. Dan McFadden did this explicitly in the 1970s and early '80s, going from the IID extreme value assumption to the generalized extreme value model. Dan felt more comfortable

---

[23] Manski, C.F. (2003), *Partial Identification of Probability Distributions*. New York: Springer.

thinking in terms of finite-dimensional parametric models than I do, but he wanted to get away from the strictness of the logit assumptions and so he went to the GEV model. The discrete choice literature also did that with random-coefficients multinomial probit models, which everyone was talking about in the 1970s. The only reason they weren't applied very much was computational. Everyone would have preferred to do that then.

**Elie**: But nowadays I think the key thing is that with these fancy models, people didn't want to relax assumptions because they were worried they would lose point identification.

**Chuck**: ...we were doing this maintaining point identification.

**Elie**: Exactly, and that was an issue. If you relax, go to random coefficients, somebody will raise their hand and say, "How do you know you're still point identified?" You say, "Well, I have all these conditions that tell you that."

**Chuck**: Yes, you have papers like Ichimura and Thompson and so on.

**Elie**: Yeah. But now we can move beyond that. In fully structural models, we can actually have ways to not...

**Chuck**: There are two questions. Are you asking, what do I think of the work that tries to relax assumptions?

**Elie**: Yes.

**Chuck**: I say, it's fine, but it's not what I wanted to do. Maximum score was in that mold. But, I decided that I wanted to come from the other direction, to start from very weak assumptions and build up.

Regarding structural models, I have written two papers that pose things the way I would like to do it. I have the paper in the *IER* in 2007[24] on partial identification of counterfactual choice probabilities.

---

[24] Manski, C. F. (2007). Partial identification of counterfactual choice probabilities. *International Economic Review, 48*(4), 1393-1410.

**Elie**: McFadden's special issue?

**Chuck**: Yes, in the special issue for Dan's retirement. Then there is the follow-up, the paper in *QE* in 2014 on labor supply and income tax policy[25], which is a specific application. What it does is it starts out with very weak assumptions. It goes back to Samuelson's revealed preference arguments, which just assume that more is better in the utility function. In a multi-attribute utility function, you assume that more is better on each dimension and nothing else.

By the way, I teach Samuelson's revealed preference idea as an early instance of partial identification, which is not within the econometrics literature. Samuelson viewed himself as a theorist, but I now see it as a pioneering piece of partial identification analysis. It was a beautiful nonparametric piece of work. You observe the commodity bundle that a consumer buys with existing prices, or you maybe observe multiple bundles purchased in different price settings. You want to predict counterfactually what this consumer would do if he were to face different prices. Sometimes you get nothing, but sometimes you get an informative bound. It depends on the configuration of the data and what you want to predict. From today's vantage point, I view that as a partial identification problem.

That's where I started in the 2007 *IER* paper. In terms of what additional assumptions to make, I focused on an instrumental variable or exclusion assumption. I introduced a statistical independence assumption that is used in choice modeling all the time. It is that you observe behavior in different markets where people face different choice sets, but you assume that the distribution of preferences is the same across markets.

The IER paper asked abstractly what the power of this type of assumption is in terms of identification of counterfactual choice probabilities. Computationally, the identification region is a bound that solves a pair of linear programming problems. You did something similar in your paper with Bo on dynamic choice with an initial conditions problem, which also has a linear programming solution. That's what I did, and I gave a simple illustration at the back of the paper. I pushed it much harder in the income tax and labor supply context in the 2014 paper.

---

[25] Manski, C. F. (2014). Identification of income–leisure preferences and evaluation of income tax policy. *Quantitative Economics*, *5*(1), 145-174.

In terms of adding assumptions, my first step after basic revealed preference was to add exclusion restrictions and see what they buy you. Then, I went further, particularly in the 2014 paper, by imposing shape restrictions. Now, this is really getting closer to the kind of literature that you were talking about, which is more parametric. In the 2014 paper, I said, "Imagine everyone has a CES utility function." Everyone has a CES utility function in income and leisure, but with individual specific coefficients. I didn't want to make assumptions about the distribution of coefficients. The fact that every person is CES has additional identifying power. You add that assumption and it tightens the bounds further.

**Elie**: The reason why the other top down approach is also useful especially in IO is because people start with a clear policy question in mind. Should we shut down this merger? Should we do this? It's a prediction problem, also, of a particular kind, trying to predict effects of something that has not been implemented before. It's a complicated policy. If you have a fully specified model, of course, it's just mechanics now. But then, we start to worry about how much of this policy response is due to functional forms.

In some contexts, there seems to be a hierarchy of assumptions, depending on the problem and the literature, and so relaxing the most worrisome assumption first would be a good way to proceed.

**Chuck**: Your paper with Bo, I think, is a good example of this.

**Elie**: That's a fully parametric model.

**Chuck**: I thought it was a very good, interesting paper. You used a lot of structure there and you say, historically, the initial conditions problem has this particular place.

**Elie**: But you could come back and say, "Why are you doing random effects probit?"

**Chuck**: The cheap answer that people give is, there's so many assumptions that you could relax so we don't know which ones to, so why bother? We won't relax any. I'll just do this...

**Elie**: The question is to try and think harder about a systematic way to do inference that is believable and useful.

**Elie**: Let me just go to the next question, which is, in applied seminars, you hear this repeatedly. We don't care about identification because we can get bounds.

**Chuck**: Has this been going on?

**Elie**: Yes. I hear this all the time now. You raise your hand and you say are you worried about some of these assumptions and the answer is I could always get bounds if I relax the assumptions. It seems lazy.

**Chuck**: It's clearly lazy because it doesn't tell you what kind of bound you're going to get. If someone is saying, "I could relax this assumption but I'm not." I guess implicitly they're saying, "I could get a bound, but the bound's going to be fairly narrow."

It's the same thing with Monte Carlo analysis. Who ever reports a Monte Carlo analysis with their estimator and doesn't show it to be better than the literature?

If people only report robustness checks or Monte Carlo analysis that show that their stuff works and they don't push it to the breakdown point, then that's deception basically. I don't expect empirical researchers to prove theorems on breakdown points. But they can still try to weaken the assumptions and do sensitivity analysis to see where their results break down. There's nothing that prevents them from doing that.

If researchers get away with this, you can only blame the journals and the referees. Researchers could be forced by the journals to do more serious sensitivity analysis. They could use the language of partial identification and do formal analysis, or use the informal language of sensitivity analysis. The journals could force people to do more serious sensitivity analysis, but the fact is that they don't. There is a collective agreement to allow this kind of deception.

In the partial identification literature, this was important from the beginning. In my own work, it shows up most explicitly in my *Econometrica* paper with Joel Horowitz on contaminated sampling[26]. We explicitly defined identification breakdown for that problem, which is a mixture

---

[26] Horowitz, J. L., & Manski, C. F. (1995). Identification and robustness with contaminated and corrupted data. *Econometrica: Journal of the Econometric Society*, 281-302

decomposition problem. You get informative bounds if the prevalence of bad data is below some level. If the prevalence of bad data goes above some limit, the bound becomes a logical bound. It's not informative.

Or take my earlier work deriving bounds on quantiles when there are missing outcome data. For the median, there is a simple bound. If the prevalence of missing data is less than 50 percent, you get an informative, two-sided bound on the median. Once you go over 50 percent, you lose it.

These situations are much simpler than with structural models but I think one could make progress of this type.

**Elie**: We're trying. There's some work on this. It's not published but there's a lot of work.

**Chuck**: These are constructive things to do. They may require some tough analysis and hard computation. The econometricians who are now working so hard to get the nth order improvement on confidence intervals could be putting their energy into this instead. I think it would be more valuable.

**Elie**: The asymptotic theory for nonstandard problems is hard.

**Chuck**: I don't understand it anymore, and I stopped doing asymptotics. When these issues come up a colleague of mine often says: "Well, asymptotics is a metaphor." I snap back and say, "A metaphor, or is it hand-waving?" The word "metaphor" sounds positive and "hand-waving" sounds negative. There's a leap of faith about asymptotics.

I only do two types of statistical research today. I sometime use large deviations analysis to obtain exact results for finite samples. I first used large deviations theory in my 1985 paper on maximum score and I have found it to be useful in some recent papers using statistical decision theory to analyze randomized experiments. More generally, I do statistical decision theory, which is finite-sample and nonparametric.

**Elie**: And uniformity is built in here.

**Chuck**: Particularly with minimax and minimax-regret analysis, uniformity is built in. In contrast, I find that I can't evaluate most of the literature that uses local asymptotics to motivate confidence sets. I can name some thoughtful people, for whom I have high respect, who think it's a good thing to do, so they try to do it, but I can't evaluate that work.

**Elie**: There are hard statistical problems that people recently have tried to address. Really hard problems.

**Chuck**: But there only is a finite amount of effort that we all have available to allocate. Is this the right place for these very bright, young people to be putting in their effort, or is it not?

**Elie**: Especially that it's not making as much direct impact on empirical work.

**Chuck**: My strong view is that I wish people would direct more of their effort towards work that is motivated by some real application. There may be hard computational problems and statistical inference problems, but they would be motivated by the application.

An example of what I have in mind is the recent *QE* paper[27] on insurance purchases by Francesca with Levon. One can argue about whether you like their specific model, but I think it is good work. When you asked earlier about structural choice models with partial identification, I told you about my abstract work in the *IER* paper, and the one in *QE* on income tax and labor supply. I should mention their QE paper, which is a full-blown partial identification analysis of a structural choice model.

They are honest about it. They run into exactly the issue that you can't get rid of all the assumptions at the same time. If you do that, you are left with nothing. They had to decide which assumptions they would focus on. They had to ask what they thought, for modeling household insurance purchases, are the key behavioral issues.

**Elie**: Is it cars?

---

[27] Barseghyan, L., Molinari, F., & Teitelbaum, J. C. (2016). Inference under stability of risk preferences. *Quantitative Economics*, *7*(2), 367-409.

**Chuck**:  It was both cars and home insurance.  The ways they get identification is by observing and interpreting multiple insurance purchases. You observe what deductibles people choose across multiple lines of insurance. The assumption they relax is the standard expected utility assumption that probabilities of events enter linearly. They model one of the variants of behavioral economics that permit nonlinear probability transformations.

**Elie**:  Yeah, yeah. There's many ways.

**Chuck**:  There's many ways to do it, so they take a class of possibilities and they allow heterogeneous types of persons.  It's like your empirical work with different levels of rationality. There may be a mixture of types.

Their work is similar in that respect. They put structure on some parts, but there is not enough structure to get point identification of the mixture of types.

I like the way they present their work, in that they give a whole sequence of results.  They show what you get with weak assumptions about behavior. Then you add more assumptions, including some more homogeneity on preferences, and they get tighter results.  They explicitly have a partially identified structural choice model.

**Elie**:  Last thing I want to bring up, which I think is the most important at this stage of the literature, we talked about it today, is computations.

**Chuck**:  Most important, I don't know...

**Elie:** I think generally, we ought to pay more attention to whether something can be computed easily and give serious guidance as to how one can do so.

**Chuck**:  This is yet another reason why I think if you're going to do econometric theory, you want to have some empirical application in mind. Because then you have to ask yourself: "Can I actually do this?" You can write down whatever abstract theorem you want, but if you don't have an application in mind, then no one puts you up against the wall.

If you have an application in mind, you are forced to confront it. The only reason why I don't give computation primacy as an issue is that computation improves over time, both the hardware

and software.  The problems that people view as hard today would have been out of the universe 10, 20, 30 years ago. Problems that were killing us then are considered easy today.

**Elie**:  No. They improve over time and I think...

**Chuck**:  But there is improvement. It's not illogical. Let's not get into the algorithms literature on what is NP-complete and all that. I'm not even sure that those theories are useful, but there are levels of computational complexity. The other thing about computations is that there always are trade-offs. You can make something computationally hard or you can simplify and get an analytical solution. Where the boundary is keeps changing, depending on what is available computationally.

**Elie**:  So, we need to better confront the computational problem in the context of a clear applied economics model. There are examples in the literature of econometrics papers that try and do that.

**Chuck**:  It was explicit in the *QE* paper on labor supply and income tax, for which Matt Masten was my RA.  It is a massive linear programming problem.  Jörg Stoye and Yuichi Kitamura were doing work with a lot of overlap in the computational structure.  Jörg gave Matt their algorithms. There is a section of my published paper that goes through this linear programming problem and describes how the computational complexity changes as you vary the assumptions. The paper tried to be useful in explaining the computational problem and providing some guidance.

I usually try to stay away from hard computational problems and try to simplify things to get an analytical solution. That's just my style.

**Chuck**:  Where I think the journals can help is by being more realistic and not requiring that someone nail these computational problems in their paper.  I think it would be better to ask "Do they make some progress on it?" Then maybe someone will come along and do better.  Go back to maximum score, Scott Thompson was my RA at Wisconsin in the mid '80s.  We had fairly primitive ideas on how to search for the maximum of the score function.

**Elie**:  Now it's really easy.

**Chuck**:  But that has taken 25 years, right?

**Elie**: Now the computer scientists figured out how to rewrite the maximum score objective functions using hinge losses. It becomes convex and so easy to compute.

**Chuck**: There can be technological progress. That's why I'm more optimistic about computation. We can get computer scientists, who see this as their specialty, to work with us.

The underlying issue is that if you want the work to be useful for some semi-serious problem, can you compute in that context? I think we should make that the objective, instead of being totally macho in terms of the criteria we now use.

**Elie**: Oh, yeah.

**Chuck**: The field can set up norms. What do you have to do to show that something is worth publishing? This is different than what the econometric norms have been. They push people to doing more and more technical work.

**Elie**: Can you say something to the readers of this interview as to where and how one should look for new/novel questions in the partial identification literature.

**Chuck**: I find it hard to imagine almost any empirical problem, in either micro or macroeconomics, where we should feel comfortable with the assumptions needed for point identification. I'm not saying that you shouldn't report point identified results, but along with them, you should be pushing yourself and saying, "What sensible bounds might I get with weaker assumptions? Where does the analysis break down?" If I have a great disappointment about the literature on partial identification, it is that this kind of research hasn't happened much for the applied topics that I care about.

Your perspective on this may be different because you have an IO orientation, and I think the IO people have been much more open to this. For the kind of things that I do, I do some. John Pepper has done quite a lot. John's work is good, he gets it published, but it has no impact on the so-called mainstream. John and I have a new paper on Right to Carry laws. It uses bounded

variation assumptions.[28] These are assumptions that relax mean independence to suppose instead that a condition suitably close to mean independence holds. Papers like this get published. Maybe not in the *AER* and not in the *QJE*, but serious applied partial identification papers can get published.

It's not like it was in the '90s or the early 2000s, where young persons might think that they should stay away from partial identification because they would never have a career. That's not true anymore. It should be part of what everyone does and econometricians need to provide the tools to make this happen, working together with empirical people. However, I don't see that happening. Again, my perspective may be different from yours, but I go to the NBER Summer Institute. I see no movement there at all. It gets more and more entrenched.

**Elie**: That's not for a lack of theoretical work.

**Chuck**: All they need to do is apply analysis that was done 20, 25 years ago. There's no excuse whatsoever, as far as I can see. Partial identification has been successful among econometricians. It has generated an enormous amount of theoretical work and it has had some success in IO applications. But really, it should be everywhere.

**From Jerusalem to Wisconsin**

**Elie**: Can you talk about your move to Jerusalem and on to Wisconsin.

**Chuck**: We moved to Jerusalem in 1979. We originally weren't intending to stay. I was on leave from Carnegie-Mellon and we went there for a semester. While I was there, they asked me to stay. Our children were young then, five and two years old. And we thought, shall we go back to Pittsburgh or stay in Jerusalem? Pittsburgh was safe and known, Jerusalem was an incredible adventure, like no place else on earth, and we decided to do it. For me it was easier being Jewish

---

[28] Manski, C.F. and J.V. Pepper (2017). "How Do Right-to-Carry Laws Affect Crime Rates? Coping with Ambiguity Using Bounded-Variation Assumptions, *Review of Economics and Statistics*, doi: 10.1162/REST_a_00689, forthcoming in print.

and knowing some Hebrew. For Kitty–not Jewish and not knowing a word of Hebrew – it was a much bigger move. We decided we would go for four or five years, see how it is, and then re-decide.

On a day to day basis it was wonderful to live there. For the children it was fantastic, with a good community. The department was very good. At that point in time the Hebrew University Economics Department clearly would have been viewed as a top ten department if it had been in the United States. It was a remarkable place. It was not strong in econometrics, but as a general economics department it was very strong. We spent four years there.

In 1982-1983, I went on leave to MIT. Dan McFadden was at MIT then, so it was a natural place for me to visit. Our family was from the Boston area, our parents wanted to see their grandchildren, so we came to MIT, and then we decided what to do looking forward. Kitty and I had agreed that after we arrived in Boston, we would have some distance from Jerusalem and then decide "Are we going to go back there for good, or are we going to return to the U.S?"

We decided to return to the U.S. and I put myself on the market. I had offers from a bunch of places--from Wisconsin, Minnesota, Michigan, and Northwestern. I went to Wisconsin.

**Elie**: Who was there?

**Chuck**: Who was there was Gary. I had never met Art Goldberger. Regarding Art, after I moved to Wisconsin, getting to know him became one of the most important things that ever happened to me. There are things that happen, and you sort of wonder if you had taken a different path, how you wouldn't have even known some person. And you just can't imagine. I can't imagine not having gotten to know Art Goldberger in retrospect. But if we had chosen to go somewhere else, who knows? I had been in the profession for ten years by then, and for some reason I had never even met him at a conference. I knew his work of course, but we had never met.

The person I did know was Gary. We were not close friends, but we went to high school together at Boston Latin School. And not only the same high school. Gary and I were in the same classes

for many years---calculus, Latin, whatever, we somehow were tracked together. As undergraduate we were both in Cambridge. I was at MIT both as an undergraduate and in graduate school. He was at Harvard undergrad and graduate school. We never saw each other as students. It was only when we both finished the Ph.D. that we found out that we both had decided to do econometrics. In the mid to late 1970s, I saw Gary periodically at conferences. So Gary and I knew each other. He was my main contact at Wisconsin, with Art Goldberger a more distant impetus to going there.

When I first met Art, I found him incredibly intimidating. I had enormous respect for his written work.  In person, he comes across as really severe at first. No nonsense whatsoever. He would just rip you into shreds if you were to say something stupid.

I have to talk further about Art. As I was writing papers then, he took me under his wing, acting as a mentor. I was a full professor, but he would read my working papers and mark them up by hand, line by line, like a thesis advisor.

**Elie**: But he didn't do that with other people?

**Chuck**: He did it with his Ph.D. students, but I don't know if he did it for other faculty.

Art read my papers for three or four years. At some point, roughly around 1986, he read one of my papers and said, "Okay, I'm not going to do it anymore. I've done all I can for you." And I said to him, "Does that mean I've graduated or you've given up on me?"  He said, "You decide."

**Elie**: Were his comments more on the substance?

**Chuck**: They were both on the substance and on the language. I think that over time I've become a good writer, relative to some others. At that point, I was not. My early work, if I look back on it, was not well written. Think about the maximum score paper I wrote when I was 24 years old. I had no one who could read it and give comments. I couldn't get any feedback on it. But I still blame myself for the obscure writing.

Art read and commented line by line, word by word, concentrating on getting across the message, the presentation. Of course, you can't do that kind of close editing unless you fully understand the concepts substantively. And Art was full of substantive contributions. This was during the period of my semiparametric work. There was a period in the mid-1980s when I went back to maximum score, and then there was a string of papers. It began with the 1985 paper and then the ones on maximum score with panel data and choice based sampling, a couple with Scott Thompson on computation and prediction, and finally the *JASA* paper on identification. I had hit on a basic idea, for about four years I wrote papers on that theme, and Art would read all of them. I was doing other work as well, more applied work. Whatever it was, I would give it to Art, he would read it, and I would get incredible comments back.

After that initial period, for the rest of the time I was at Wisconsin, I would talk with Art about work as it was going on, but he didn't do line by line reading anymore. Art became extremely important when I began working on partial identification. I dedicated my book on *Partial Identification of Probability Distributions* to him because he had faith in the idea and in the work. Even though I was able to get the early work on partial identification published, basically I was alone and no one was paying attention to the work. Art encouraged me to stick with it and keep going, regardless of whether people were interested in the work or not. He became very important.

The next person to come to Wisconsin was Jim Powell. I don't remember when Jim and I first met. It was probably when he was right out of graduate school. He went to MIT as an assistant professor. When he was a midterm assistant professor there, he was worried about getting tenure. He had these superb papers on censored LAD and so on, but not that many papers. We got it in our heads that we might offer Jim a lateral move to Wisconsin and then bring him up for tenure soon after that. This made sense because we were strong in econometrics at that time. Jim liked Madison too. He moved and of course he did get tenure soon.

Jim was in Madison for five years or so, and he did some of his best work then. He wrote his paper on symmetrically trimmed least squares, which was a beautiful idea. He also co-wrote with

Whitney a paper that's not as well known, but also nice, on expectiles, which are the square-loss analog of quantiles.

I remember when Jim generated this idea as a classic Wisconsin moment. Jim is a very good technical econometrician, but what I really valued is that he had simple conceptual insights. Jim had the ability to see something brand new. Expectiles were a new idea. Of course, in retrospect they are a simple idea. We were already comfortable with asymmetric least absolute deviations, so why not do asymmetric least squares? But no one had thought about it. Jim did and he worked out the theory.

The classic Wisconsin aspect was that we all were paying close attention to naming things. Jim said, "I've got this idea. What should I call it?" We were in the hallway--Jim, Art, and myself--- and we probably spent a whole day thinking of different names. He settled on 'expectiles,' which I think is a good name. I remember that 'projectiles' was another name. Probably not so good.

**Elie**: Looking back, such an incredible group of people. With this group of people, you would expect more joint projects?

**Chuck**: That's actually interesting.

**Elie**: Maybe it wasn't common back in the day?

**Chuck**: Well, people did write together sometimes. I think it's true that the prevalence of coauthoring has gone up over time, but we are not talking about the 1600s here. We're talking about the 1980s. I think we each had our own agenda. I suppose that Jim and I might have coauthored since we were both focusing on semiparametric work from different directions, but at least it was on that topic. Gary also had related interests. He did his paper on efficient method of moments and I had done my work on analog estimation, of which the method of moments is a special category.

**Elie**: Also, Gary wrote that paper in the mid-1980s on the censored/selection models and identification at infinity with large support.

**Chuck**: Yes, his paper on the best achievable rate of convergence. You know, it's interesting. In retrospect, I don't even remember a case of potential coauthoring coming up. We commented on each other's work and we tried to help each other. But each of the ideas was being generated independently. I don't remember conversation where we entertained the idea of coauthoring. It wasn't out of hostility. It just didn't happen.

**Elie**: Even with Art, did you ever write a paper with him?

**Chuck**: Yes. Much later, but not an original research paper.

**Elie**: A book review.

**Chuck**: The review of The Bell Curve.

**Elie**: Yes, I remember it.

**Chuck**: It is in the *Journal of Economic Literature* in 1995.[29] I initially did not want to do this work. What happened was that *The Bell Curve* had come out in print. Art, going back to the 1970s, had been a severe critic of the literature on heritability, on attempts to decompose life outcomes into fractions explained by genetics and by the environment.

There were two reasons why this was so important to Art. One was that the statistical and econometric analysis in the literature was horribly flawed. The second reason was personal. The heritability research was put forward as ordinary neutral social science, but some of it had a racial tinge. It was used to argue against anti-poverty programs and particularly against helping African-Americans by making the argument that they're just not bright and there is nothing you

---

[29] Goldberger, A.S. and C.F. Manski (1995). Review Article: *The Bell Curve* by Herrnstein and Murray. *Journal of Economic Literature*, 33(2), 762-776.

can do for them.  I won't say that everyone who wrote on heritability was a racist, but there was a subset who seemed to be.  If there was something that Art felt extraordinarily strong about, it was about racial equality, particularly of blacks and whites.

In the late 1970s, when there was a lot of work and controversy, Art wrote this magnificent paper with the title "Heritability" that ripped the literature to shreds. It is probably the paper that Art was most proud of.  *The Bell Curve* came much later, in 1993 or so. To Art, this was holy war. You had to cream these guys Herrnstein and Murray. Their work was really bad.

**Elie**: Heckman and others came down on it too, right?

**Chuck**: Many people were incensed and many reviews of *The Bell Curve* were written.  John Pencavel, the editor of the *Journal of Economic Literature,* asked Art to write a review essay for the *JEL.* Art came to me and asked if we could do it together. I remember saying to him, "Oh come on, do we really have to do that? This is going to take a long time. If we're going to do it, we're going to have to really dig into this and do it right."

Art and I both thought that the early reviews of *The Bell Curve* tended to be superficial, mixing criticisms of the bad science with anger at the results that Herrnstein and Murray reported and their objectionable policy recommendations. We wanted to do it right and we knew that this was going to take a long time. I remember saying to Art: "I'd rather do new research than review someone else's research." He said that it was for the good of society and really important.  I agreed to do it. It took us several months.

Working with Art was arduous, but not in a negative way.  I am a nitpicker, but Art was much more of a nitpicker than I am. In the end, I am proud of our work and I think Art was as well. I think we wrote the most informative review.  We dug down into basic conceptual issues. I think the review article is worth reading even today because it raised basic questions.

For example, the book reported logistic regressions and concluded that genes are "more important" than environment.  But what does it mean to say that genes are more important than

environment? There is no common unit for measuring genes and the environment. A common practice in sociology, used in *The Bell Curve*, was to take the cross-sectional standard deviation of some real-valued measure of genes and environment as a constructed common unit and to define "more important" in these terms. That is, researchers would report how much an outcome like years of schooling or wage would change with a one standard-deviation change in the measure of genes and environment, and they would draw conclusions about relative importance this way.

Art and I wrote that this is nonsensical! The standard-deviation scaling depends on the population spread of each quantity, so your results on what is more important depend on what constitutes a standard deviation in the population under study, which is a feature of the marginal distribution of covariates. If you were to study populations with different degrees of heterogeneity, you would get different results. This was one of several simple points that we made in our review that I thought were important not only in assessing *The Bell Curve* but also a large body of research in the social sciences.

Returning to the Wisconsin group, beyond Art, Gary, Jim, and me, there also were periods in which John Rust and Ariel Pakes were there. John and I both arrived in 1983, he direct from his Ph.D. at MIT. We had lots in common at first because we were both doing discrete choice analysis, but John remained attached to parametric models and his interests became increasingly computational, so we drifted apart. Ariel was at Wisconsin for only a couple of years. He and I had been close in Jerusalem. Ariel and I talked a lot, but we never wrote together either.

**Chuck**: Thinking about Wisconsin from the late '80s on, there is something that we haven't discussed yet, except in a tangential way, but was extremely important. In our earlier conversation, I said that I wish econometric research would have a foundation in real world problems. This could be in many different areas, but there should be some real-world problem. This has been important throughout my career, so I want to talk further about it.

We already talked about how partial identification came out of my contact with Irv Piliavin and his specific question about missing data in a longitudinal survey. What we haven't talked about yet is that about a year later, the Institute for Research on Poverty (IRP) at Wisconsin needed a new director. I had never done poverty work. I was just a straight econometrician. But someone suggested to me that I become director. I decided to apply for the position and they asked me to do it.

I spent three years as director of IRP, from fall 1988 through spring 1991. This was when everyone other than Art was leaving. But my life was full in that period, so I wasn't lonely. IRP is a multidisciplinary social science institute. It had its roots in the Wisconsin economics department back in the 1960s from labor and public economists involved in the War on Poverty. Art had always been involved in it, because of his substantive interests. There also were people in sociology, social work, political science, even in the history department. They were doing lots of applied empirical work on policy evaluation, but there was no serious econometrics going on.

I knew that it was a risk to take the job, but I was willing to take it. I have taken professional risks my whole career. Maximum score was a risk. Going to Jerusalem and being isolated from most of the academic community was a risk. I've never been afraid of taking risks. But why would I risk moving from a reasonably central position within econometrics to dealing with all sorts of social scientists and policy types?

There were two reasons. One was that I felt that poverty research was important. I had always had interests in policy, as early as the thesis work on college choice. I thought, well, this will put me in an entirely new situation, and I'll have to deal with all kinds of different people, but it's important.

The other reason why I did it was that by the late '80s, I concluded that I was reaching diminishing returns on the semiparametric work. I had a long stream of publications in the mid-80s on maximum score and related topics, culminating with the 1988 JASA paper. After writing that paper, I thought I understood the literature the way I wanted to and I became worried that I

had run out of interesting work to do on the topic. Diminishing returns can be a serious problem for people doing econometric theory. Even the best research stream transitions from a period of excitement and productive work to one where it eventually runs out of steam. I had built up enough human capital in semiparametrics that I could have cranked out further papers for several more years, but I worried that they would be increasingly marginal. I was looking for a new source of ideas and thought that IRP might provide them.

My period at IRP turned out to be extraordinarily fruitful, as it generated three major new ideas. One was partial identification. At IRP, I was seeing work on policy evaluation and watching people argue with each other about the effects of welfare programs, all based on analyses that don't have much foundation. Another idea, which we haven't talked about yet, was analysis of social interactions. I learned about the research studying neighborhood effects on poverty from sitting in on sociology seminars at IRP. Third was the idea of measuring expectations. All of these ideas developed in the period 1988 to 1993. Each turned into a major new stream of research. All were generated by having to deal with real world topics at IRP.

**Work on Expectations**

**Elie**: OK. We're continuing now with the work on expectations. Can you tell us about when it all started?

**Chuck**: The work on expectations began roughly at the same time as the work on partial identification. I had given it some thought much earlier, about 1980, but I had never done anything about it. It was about 1988-89 when I began the IRP work that I went back to it. The impetus was that I was then in contact with demographers who study family formation and transitions.

The specific way it started was that I thought demographers were misinterpreting the responses to a question on the Current Population Survey asking women about fertility intentions. The question was: "Do you expect to have more children?" The allowed responses were "Yes, no, uncertain." I sat in on seminars in which demographers analyzed the data and used the responses

to forecast what future fertility would be. The practice was to compute the fraction of "yes" responses and conclude that this would be the fraction of women who will actually have children in the future.

I thought that this forecasting approach is logically incorrect if women are uncertain about future fertility. I reasoned as follows. Suppose that a woman has a subjective probability of having future children. Then she should respond "yes" to the CPS question if her probability is above one half and "no" if the probability is below one half. Thus, saying yes does not imply that she will have a child for sure. The matter is related to partial identification, in that the question response gives a bound on the subjective probability. A "yes" answer means that she has a probability between 0.5 and 1. A "no" means the bound is between 0 and 0.5. It is not clear how to interpret a response of "uncertain." In any case, the demographers' practice of using the fraction of "yes" responses to predict fertility could not be logically correct. The fraction of "yes" responses only reveals the fraction of women who have subjective probabilities over one-half.

It was a simple point, but apparently it was not in the literature. I dug deeper into the matter and wrote a full-scale paper[30] on "The Use of Intentions Data to Predict Behavior" that was published in *JASA* in 1990. I ended the paper by arguing that the CPS question is too crude to be useful. I recommended asking women to state the percent chance that they would have more children.

**Elie**: Probabilities, yes, so, you actually think people do have priors?

**Chuck**: Yes. Thinking about fertility intentions was part of what led me to conclude that we should ask probabilistic expectation questions more generally. The other issue that drove me to it is that for years I had been going to labor econometrics seminars on structural modeling of choice under uncertainty. I am thinking particularly of the work by Ken Wolpin, Zvi Eckstein, and Mike Keane--- the whole set of papers on dynamic choice of years of schooling under uncertainty. I valued their work as good applied structural econometrics, but I did not like their

---

[30] Manski, C. F. (1990). The use of intentions data to predict behavior: A best-case analysis. *Journal of the American Statistical Association*, *85*(412), 934-940.

practice of making assumptions about expectations and supposing that people have rational expectations.

In the schooling context, the assumption meant that people making decisions have objectively correct probabilistic expectations for the returns to schooling, conditioning on the information they have.  I asked myself: "How would they achieve rational expectations? Labor economists don't know the objectively correct returns to schooling, so how is it that ordinary people could?" It seemed clear to me that it was overly optimistic to assume rational expectations.

There is a second major problem with rational expectations assumptions. When economists use these assumptions in practice, they assume much more than that the agents in their models have objectively correct expectations. They also assume that they, the researchers, know what is objectively correct.  A researcher poses a stochastic model generating future events and assumes that the agents in the economy use the same model that the researcher poses.  This is pretty arrogant.  The macroeconomists, Lucas and so on, were very clear about this. They assumed that the agents in their models have the same model of the economy as the researcher does and that both of them are correct. I thought that this doesn't make sense and that we should think harder about how people form expectations.

I concluded that the structural econometric models assuming rational expectations are likely misspecified.  They were making strong and potentially wrong assumptions about expectations and estimating random utility models based on these potentially wrong assumptions. To make this point I wrote a second paper[31] on "adolescent econometricians."

The title was, "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?" I wrote it for an NBER conference on demand for higher education and it was published in an NBER volume. Between those two pieces of work, the *JASA* paper in 1990 and the Adolescent Econometricians paper in 1993, I decided that it would be better to collect expectations data than to continuing making assumptions about expectations.

---

[31] Manski, C. F. (1993). Adolescent econometricians: How do youth infer the returns to schooling? In *Studies of supply and demand in higher education* (pp. 43-60). University of Chicago Press.

I began to ask: "Why don't we have data on expectations?" I realized that we don't have data because economists are generally taught not to collect subjective data. This attitude has softened by now, but when I was in graduate school there was a strong dogma -- and I think "dogma" is the right word. The dogma was "Economists believe what people do. Economists do not believe what people say." Accepting this meant that only choice data are legitimate, so you can do reveal preference analysis. This view goes back to Samuelson's original revealed preference papers in 1938 and 1948.

**Elie**: Was that because people thought with expectation data people don't have the incentive to answer correctly?

**Chuck**: Yes.

**Elie**: Measurement error type?

**Chuck**: The reasoning that was given, and I don't want to downplay it, was to say that people don't have an incentive to take expectations questions seriously. They don't have an incentive, so why should you believe the response.

Now there is a literature on so-called "proper scoring rules" where you give people an incentive to respond correctly.

**Elie**: This comes up in experimental work, too.

**Chuck**: That's right. The experimentalists, particularly Andy Schotter at NYU, have been collecting probabilistic expectations data since the early 2000s. They can use proper scoring rules because in experiments you can give rewards based on what people do. On the other hand, proper scoring rules typically are not possible in surveys. In a survey, if you're asking people for their expectations for fertility or future jobs, there is no way to set up a proper scoring rule.

I decided that the incentive issue may be legitimate, but not so severe as to make data collection useless. I realized that the incentive issue applies to all survey questions, not just to questions about expectations. I thought, "If you are going to be too serious about the incentive issue, then we wouldn't have any micro data from surveys at all." I felt that surveys have been useful, but you have to be cautious. So we should try to ask about expectations. I always kept in mind that

the only alternative to measuring expectations was to make assumptions. It did not seem appealing to me to make assumptions about expectations without any data.

**Elie**:  How could data in it and by itself be actually bad?

**Chuck**:  Yes, how could data per se be bad?

*In 1978 at the Great Meadows National Wildlife Refuge in Concord Ma.*

**Elie**:  It is expectation data so it's not like you're saying this is what actually happened.

**Chuck**:  That was my view. There was enormous resistance at first. Several things happened more or less independently and simultaneously that got this literature going.

First, I was at Wisconsin, Jeff Dominitz was a graduate student, and I was advising him. I don't remember how this happened, but his thesis came to focus on the collection and analysis of expectations data. He did some work on his own, but much of the work we did together.

We did two things. One was a small pilot study on the returns to schooling. Jeff personally went to high schools in the Madison area. We got permission from high school principals and teachers to go into classrooms and elicit from high school students their expectations of the returns to college attendance.  Our paper was published in the *Journal of Human Resources* in 1996.[32] Since then there has been much more work eliciting returns to schooling, particularly by Basit Zafar and Pamela Giustinelli, both of whom I advised at Northwestern.

When Jeff and I did this early pilot study, a primary open question was whether high school students would be able to answer the questions in a sensible way. My wife, who was teaching high school in Madison, thought it was crazy to ask her students for probability assessments.

**Elie**:  Was this ever followed, to see what would happen? What happened after they actually went to college.

**Chuck**:  There was no follow up. It was a one-time survey, but the work had some influence on the field. The paper is reasonably well cited and the findings were viewed as positive, if only because we showed that you could ask questions and obtain responses that don't look crazy.

**Elie**:  You could compare it to the school's average?

---

[32] Dominitz, J. and C.F. Manski (1996). Eliciting Student Expectations of the Returns to Schooling. *Journal of Human Resources*, 31(1), 1-26.

**Chuck**:  You are raising the more general question of how, once you have these data, do you know how much credibility they have. How do you evaluate them? I'll try to get to that a little later.

The other thing that Jeff and I did, which was much more substantial, was to contact the demographers at Wisconsin who were running a national telephone survey every month. After a lot of back and forth, they allowed us to put probabilistic expectations questions on their survey. This became what Jeff and I called the Survey of Economics Expectations, which we ran monthly from 1993 to 2002. I got NSF funding to support it.  The phone survey gave us repeated random samples of the whole US population.  We asked a set of questions, including the chance you will lose your job in the next year, the chance you will be burglarized, the chance you will have health insurance and others.  We wrote a series of papers using the data and published them in *JASA*, *REStat*, *JHR*, *Public Opinion Quarterly*, and other journals.

At the same time at Michigan Tom Juster was beginning the Health and Retirement Study.

**Elie**:  Yeah, HRS.

**Chuck**:   I view Tom as the father of measuring probabilistic expectations. Tom in 1966 in *JASA* published a paper recommending collecting of probabilistic data for buying intentions. This is important and Tom's work has been underappreciated.

When I learned about it later-- I only learned about Tom's work in the late '80s when I started doing my own -- I was embarrassed, because I knew Juster from times I had visited Michigan and I had no idea he had published this paper until 25 years after he had done it.

This history is relevant. In the early 1950s George Katona at Michigan began doing consumer confidence surveys. At the time, he was called the father of psychological economics. This was not behavioral economics, this was different. During the 1950s there was a lot of interest in measuring personal expectations and subjective views of the economy. The idea was to have leading indicators for macro forecasting. The Federal Reserve funded it.  They wanted to learn the expectations of consumers to help forecast consumer purchase behavior.

Collecting data on buying intentions was particularly important.  They posed questions on buying intentions that were similar to what the demographers where doing with fertility.  They asked: "Do you expect to buy a car in the year ahead?" and the allowed responses were "Yes, no."  Juster had been involved with this in the 1950s and early '60s.  He concluded that the buying intentions question was not very informative and that instead we should ask people to state the chance that they will buy a car.  He posed this idea, did a small pilot study, and wrote a paper that was published in *JASA* in '66.

It was a nice piece of work, but economists paid no attention to it. On the other hand, people in marketing did.  Of course, people in marketing would be interested in buying intentions. They read Juster's paper and began asking his version of these probability questions. They called it the Juster Scale. That was in the marketing literature, but the economists paid no attention.  When I began my work much later, starting with my critique of the fertility question, I rediscovered Juster's idea on my own. When I found out about his work, I of course referenced him. What I did that was original in my own paper was to formalize the bounds and work out some implications.

In the early '90s, Juster initiated the Health and Retirement Study.  He was still interested in expectations, so he made sure there was a whole module of expectations questions put in the HRS.  That module has been an important part of the HRS from the beginning. He brought in Mike Hurd to work on this and other parts of the HRS, and then Bob Willis who later became director of the HRS.

Another independent source of data arose in Italy, where the Bank of Italy began in 1989 to ask questions about income expectations in its Survey of Household Income and Wealth.  They asked households to forecast their incomes in the year ahead, essentially to give subjective income distributions. Luigi Guiso and other Italian economists were involved in designing this survey and in analyzing the data.

So, in the early '90s there were at least three groups who independently thought that asking probabilistic expectations would be a good thing to do.  We became aware of each other and the work started taking off. In the beginning, there was a lot of skepticism and it wasn't easy to get

papers published. Many economists didn't want the expectations data because of the incentive issue we discussed. They said, "You just can't trust this."

The survey research people, who were predominantly sociologists and social psychologists, didn't want it for a different reason. They were happy to ask all kinds of subjective questions, but they would ask soft verbal questions. For example, the General Social Survey would ask about job loss, "Think ahead about losing your job in the next year. Do you think it's very likely, likely, or unlikely that you would lose your job?"

**Elie**:  Did they think people understood what probability really is?

**Chuck**:  They thought people wouldn't understand probability. When Jeff and I wanted to put the questions on the Wisconsin survey, sociologists were running the survey and there was strong resistance. They literally said: "Your questions will destroy the survey."  They said: "You'll ask people what's the chance you'll lose your job, and they'll say, what do you mean what's the chance? They'll get all upset, and they'll hang up the telephone."  The way we got the questions on the survey -- it was a negotiation -- is that we agreed to put our questions at the end of the questionnaire.  Then, if people hung up the phone, it would only be our data that would be lost. It wasn't a longitudinal survey so they weren't worried about attrition. We said OK, we'll put it at the end.

It turned out they were wrong. They were adamant that people could not answer our questions, and I would say, "How do you know? Has anyone tried it? Is there published research?"  They said, "We don't need to have research, everybody knows this."  But if everybody knows this and we don't need to have research, who knows if they are right? I persuaded them to let us try it, and they were wrong.

It turns out, and the sociologists weren't aware of this, that some psychologists in the area of judgment and decision had been asking probabilistic questions for 100 years, but that was a separate literature.  Early on, the psychologists questioned experts like weather forecasters, who would forecast the chance of rain. More recently they have questioned college students in experimental labs, and so on. The psychologists had been finding no problem asking

probabilistic questions. Eventually I discovered the literature and got to know some of the psychologists who were doing this.

There are two differences between what the psychologists were doing and what we economists wanted to do. First, the psychologists would use convenient samples of students, whereas the economists wanted to do national surveys to get representative samples. Second, the psychologists were intrinsically interested in how people answer the questions, whereas the economists wanted to use the data to help understand choice behavior. There were differences in what we did and why we did it, but we still could learn from each other.

During the '90s these groups found out about each other and we ran conferences, particularly at Michigan. We brought people together and things started taking off. There was kind of a support group, because we were getting beat on by the orthodoxy. The economists said you can't ask subjective questions at all and the sociologists said you can't ask probability questions. We thought both were wrong. Eventually we started getting papers accepted, which made it easier for later researchers.

Jeff and I had a breakthrough in '97. We had a basic paper[33] on the Survey of Economic Expectations published in *Public Opinion Quarterly*, which is the main journal in survey research. We also published a paper that year in *JASA* on measurement of income expectations[34]. Then people could say there are serious journals that are willing to publish this work. It went on from there and has since grown into a large literature.

I gave the Fisher-Schulz Lecture for the Econometric Society on the topic "Measuring Expectations" and published the resulting review article[35] in *Econometrica* in 2004. This year I was able to substantially update it when I was asked to prepare a paper on macroeconomic expectations for the 2017 NBER Macroeconomics Annual volume. The new paper focuses on

[33] Dominitz, J., & Manski, C. F. (1997). Perceptions of Economic Insecurity: Evidence from the Survey of Economic Expectations. *Public Opinion Quarterly*, 261-287.

[34] Dominitz, J., & Manski, C. F. (1997). Using expectations data to study subjective income expectations. *Journal of the American statistical Association*, *92*(439), 855-867.

[35] Manski, C. F. (2004). Measuring expectations. *Econometrica*, *72*(5), 1329-1376.

topics that should be particularly of interest to macro people, such as expectations of equity returns, inflation expectations, and expectations of professional macroeconomic forecasters.

**Elie**: These expectations data must be useful in structural models.

**Chuck**: That was an objective from the beginning. Remember, I said that I didn't like the rational expectations assumptions used in structural models.

One of the first to use probabilistic expectations data in structural modelling was Adeline Delavande, whose dissertation at Northwestern was on contraceptive behavior. She considered young women who are sexually active and choose among different contraceptive methods. They make decisions under uncertainty. A central uncertainty is the possibility of getting pregnant if they use one contraceptive method or another. There are also issues of sexually transmitted disease and so on.

I could go on and on to describe other uses of expectations data in structural modeling. To reiterate, one of the main original reasons for collecting expectations data was to be able to replace rational expectations assumptions with expectations data in structural econometric modeling. This started to happen in the past ten years and is becoming more common.

**Social Interactions**

**Elie:** Now, we get to your most cited paper which is within this exciting area of inference on models with social interactions. Can you speak about this work and the *Restud* paper[36] in particular.

**Chuck:** We talked earlier about how the three years I spent as director of IRP at Wisconsin affected my research greatly. One outgrowth was the work on partial identification, which came directly out of thinking about attrition in a longitudinal survey, and another was on measuring

---

[36] Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies, 60*(3), 531-542.

expectations. I described how that came out of reading about how demographers analyzed fertility intentions. These are applied problems that, as an econometrician, I would not probably have seen on my own.

The third was the work on social interactions. My interest in it came from seminars I would attend, usually given by sociologists but also by economists, on the impact of so-called neighborhood effects or peer effects on outcomes in low income families. We have spatial segregation in the United States, and you see that people who grow up in areas that are poor tend to remain poor themselves. You don't have good intergenerational mobility, so there has been a longstanding question about why this happens. I don't know how far back the research on this goes, but the literature I know got going in the 1960s. The general view was that this is not all individual level phenomena, but it is rather the effect of a person's environment on their outcomes through their social interactions.

There was a large body of work on schooling outcomes. The question was to understand the effect of where you live on how well you do in school, and then what happens when you go into the labor market and so on. When I came to IRP this work was still going on and I saw the estimates that people were producing.

They were estimating regressions. The regressions were of someone's outcomes---say how many years they attend school or whatever life cycle outcome they have---as a function not only of their own characteristics but as a function of the characteristics of the neighborhood or peer group or whatever environment in which they were embedded. The researchers would look at the coefficients in these regressions and do a typical thing. If they obtained a statistically significant coefficient in a certain direction, they would say 'there is a neighborhood effect.' These kind of neighborhood effects, where you estimate a regression of a personal outcome on characteristics of the neighborhood, were often called contextual effects. Context matters.

Sitting through these seminars, I realized that what was going on in the regressions was one thing, but the discussions that people were having were far more nuanced than what was in the regressions. Often, I would hear two different stories. One was that you have bad schooling

outcomes if you are in a school with lots of lower income black kids. The idea was that the fraction of lower income black kids in the school, or lower ability kids or whatever, influences your outcome. That's one story, and that's what contextual regressions were appropriate for.

But often people were talking in the discussion about a more subtle phenomenon that I eventually labeled an 'endogenous effect,' where my outcome is affected not by who I am around, per se, but by the outcomes of the other people around me. I thought about this and saw that there is a feedback. If my outcome is affected by the outcomes of people around me, I also contribute to the aggregate outcomes. Not only am I affected by the outcomes of people around me, but I contribute to the people around me. That is a simultaneity problem.

I could have written it down as simultaneous equations with an equation for each person. You would list the whole set of people, say a hundred people, and you'd have a hundred equations where each one interacts with the others. I decided not to do that because a simplified form of these models was being used in the literature and I could see that it had some special characteristics. The form assumed that my outcome depends on the mean outcome in the neighborhood, rather than in a more abstract way on everybody else. So I decided to think about this model where my outcome depends on the mean outcome.

Intellectually, the most interesting aspect of this to me to this day is that sociologists do not typically think about feedbacks. They pose unidirectional hierarchical models where A influences B, and B influences C, but with no feedback. The typical sociological model did not have feedbacks that generate an equilibrium. I decided to write down the simplest model, the linear-in-means model. To start it off, I purposefully made everything linear.

It was kind of odd. All the work that I was doing at the time was nonparametric. I had left behind any kind of parametric model, never mind linear parametric models, ten years before. But I wanted to communicate with the people who were doing the applied work. And I could see that if I were to write this down abstractly, nonparametrically, they weren't going to get it. It was also going to be much more subtle to analyze. So, I decided to play with the linear model. The linear model is simple, so I got simple results. Then I decided to write them up.

Of the huge number of citations this paper receives, probably $1 - \varepsilon$, with very small $\varepsilon$, is about the linear-in-means model. However, if you read the *Review of Economic Studies* paper, after going through the linear model, there is a substantial part of the paper in which I say, "The linear model is very special. Let's look at this more abstractly."

I wrote a section on parametric nonlinear models like logit models. Later on, Brock and Durlauf studied these models much more intensely. I observed that nonlinear models are different in that you could get multiple equilibria. I wrote a fully nonparametric section too. All of that is in the paper.

But what the paper is known for is the linear model. It is by far my most cited paper. I ascribe this to two things. First is the linear model, which makes it accessible to all kinds of people.

**Elie:** It also has to do, I think, with the nomenclature that you introduced there. It has become sort of a standard too.

**Chuck:** Yes. Endogenous effects, exogenous effects, and correlated effects. That of course is not specific to the linear model. That's a conceptual thing. I think it's an extremely simple exposition. It's the kind of thing that should have been recognized 30-50 years earlier. But somehow it didn't seem to have been, and particularly not in sociology. It comes back to the fact that sociologists just didn't understand how different things are when you have feedbacks. So that is right. The nomenclature was useful and is cited a lot.

Another reason for the citations may be the metaphor of the reflection problem, which I still love. I use it all the time. Whenever I'm trying to explain identification problems to this day, to explain identification problems from scratch, I use the idea of standing in front of a mirror and not being able to separate whether you are making the mirror image move or the mirror image is making you move. I find it works as a metaphor. I called it the reflection problem. That was good branding, I guess. When the paper came out, it didn't immediately take off. There was

some recognition of it and I gave seminars, but it wasn't right away a hit. It eventually went viral.

After this I wrote another paper on which I worked hard to go further, but I have to say it was not particularly successful. I knew that the models in the *REStud* paper were special – even the nonparametric form was special – because I assumed the interaction to be simultaneous. A lot of times you might think that there is some time lag in the way these things play out. Maybe what I do today depends on what people did yesterday and so on. Then you get feedback, but it occurs in a dynamic process.

I was aware from my training in linear simultaneous equation systems about the difference between recursive systems vs. systems that look recursive but actually have dynamics with feedbacks in them due to serially dependent disturbances – these are things we used to teach in econometrics. I knew that dynamic systems can be subtle. I wrote a paper on the dynamic version and it was published in a chapter in a book from Santa Fe Institute. I don't think I really had any good insights in that second paper, which I wrote in the mid-1990s.

Then I did have another opportunity to think further. I was asked to write a paper[37] for the *Journal of Economic Perspectives*. That came out in the *JEP* in 2000 and is my second most cited paper. It is an expository paper rather than a technical one. However, it is not a review article. There is a lot of conceptual thinking in it about how different fields do things. Writing the paper was helpful to me in getting my thoughts organized in a more coherent way.

After the *JEP* article, I haven't worked very much on social interactions, but I have done scattered things. I wrote the paper that you know from the *Econometrics Journal*, studying identification of treatment response with social interactions. That was a few years ago, from that conference in Toulouse, you remember?

---

[37] Manski, C. F. (2000). Economic Analysis of Social Interactions. *The Journal of Economic Perspectives*, *14*(3), 115-136.

**Elie:** Yes. That's right it is the paper that also contains some interesting conceptualization of what it means to have network effects in the specific language of the counterfactual notation.

**Chuck:** And I did a paper on fertility interactions in Israel with Joram Mayshar, a friend. This was an applied paper, on the question of how family size gets chosen. It has a dynamic setup and some interesting result on situations in which there are multiple steady states. But I haven't been working much in the social interactions area, just a bit here and there.

The field as a whole has exploded, as everyone knows. I think it has led to some good econometrics. We now understand more about the specifics of the identification problem in the linear-in-means model and what changes when you go to different forms of models. I had discussed exclusion restrictions in the *Restud* paper, but the more recent papers have fleshed this out. So, we now have lots of work on the econometrics of social interactions. This includes the newer literature on network formation, to which I haven't contributed myself. I think this work is important. When we move away from the linear-in-means and other anonymous interactions to network models where people hold different positions in a network, the analysis becomes much tougher but we are starting to learn.

Thinking back, this was a case in my career when I was fortunate that I did something early on. And then everyone felt it necessary to cite me.

**Elie:** Well, a central piece in the paper is the negative result for that version of the linear in means model. Many of the cites to the paper essentially have the flavor of "identification is hard in social interactions models, see Manski".

**Chuck:** It is good that the literature has taken off, but I don't feel that it is as essential to me as some other things are. Partial identification is far more central to who I think I am. But I have found that some people know me just by the interactions work.

**Elie:** I think the empirical literature on social interactions is a little problematic. It is hard to motivate the use of a policy that manipulates one's peers. Some exist, but are not natural. Now, as a structural model, we need to work hard to explain how/why other people's outcomes enter your utility function? Can you write it down? Can you write me a model? Not problematic as a negative, but more as in complicated. And you really have to think about, what is this model, how are you interpreting parameters in relation to that model.

**Chuck:** I think that's the way it should be. It is subtle and it probably depends on the context. We haven't talked about work on oligopoly. The oligopoly literature, in which you've worked, is a part of the social interactions literature, where you have the interactions of the competing firms.

**Elie:** The problem is with models you're going to start making assumptions. I mean, you have the utility function, and you have to convince people why your utility has to contain other people's outcomes, which is fine. You write it as a game. But you know, then you get into assumptions about behavior etc. You just have to defend them.

**Chuck:** If it really is necessary to get to that level of detail to say something useful, then I'm not averse to doing that. But what I would be averse to is just making assumptions out of convenience. And then making counterfactual forecasts with them.

**Elie:** Network or social interactions literature has completely exploded. There are empirical papers describing behavior far removed from you in a network but that somehow seems to affect you through some process of diffusion. Some people are excited about it.

**Chuck:** If you are thinking about real applications, it wouldn't surprise me that you need to have a fair amount of structure to say something. The *Econometrics Journal* paper[38], which is on treatment response with social interactions, asks how far you can get with a reduced form. With

---

[38] Manski, C. F. (2013). Identification of treatment response with social interactions. *The Econometrics Journal, 16*(1).

monotonicity and other assumptions that may be derived from a structural model, you legitimately have a reduced form. And you can ask what kind of partial identification results you can get. I worked hard on that paper and think I got some results. Again, some of what I think I accomplished was nomenclature and setting up notation rather than theorems.

**Elie:** That's a nice paper.

**Chuck:** But if it turns out that for some purposes the fine structure really matters, then that's important. Right? It's important to understand it. It may limit our ability to do forecasting, but we certainly want to understand it, right?

**Elie:** Yes. I don't have a problem with that. I am just saying I think you really have to be wedded to economics. Your Restud paper arguably is social science generally. Anybody can understand what is going on. But then to really go into this, you really have to understand game theory then you have to understand what the utility is, where it comes from, and best response functions, etc.

**Chuck:** This may scare some people off, but that's okay.

**Elie:** It becomes very niche in economics too especially.

**Chuck:** But if that's what it takes to do it right, then it's okay.

**Elie:** I agree.

**Work on Decision Theory**

**Elie**: OK, so moving on to the big part of your last, what, more than 10 years of work...

**Chuck**: Yes, it's about decisions under ambiguity.

**Elie**: Maybe even 30 years, I guess.

**Chuck**: I would date it back thirty years or more. I always had a side interest in decision theory. There are a couple of early papers published in 1981 and 1988, but they're not really connected to the more recent work.

The work we want to focus on begins in the late 1990s. This comes directly out of partial identification of treatment response. There are many applications of partial identification, but from the beginning one that was important to me has been identification of treatment response.

The first paper was the 1990 *AER* Proceedings paper, which gave worst-case bounds and IV bounds on average treatment effects. Often the bounds cover zero, so it could be that treatment A is better than treatment B or that B is better than A. When I would give seminars, people would say: "You get these bounds, but what good are they?" They would argue that to make a decision, one has to make strong enough assumptions to at least identify the sign of the average treatment effect. They acknowledged that they might not like the assumptions they make, but they would argue that there is no reasonable way to make a decision if one does not know the sign of the treatment effect.

This is where my earlier interest in decision theory came in, because I realized they were wrong. You don't have to know the sign of an average treatment effect to make a reasonable decision. After all, what does a Bayesian do? If he doesn't know something, he puts a subjective probability distribution on it and maximizes subjective expected utility.

That would have been an easy answer to the critics—be Bayesian. However, I wanted to think about reasonable decision making without placing a subjective probability distribution on the value of an average treatment effect within its bound. What might one do?

Long ago I was aware of work on decision-making under ambiguity, stimulated by Ellsberg's paper in 1961. I had used maximin expected utility in my 1981 paper and was also aware of the minimax-regret criterion. Applied researchers were not using these ideas, but decision theorists knew them.

Chernoff's work, that of Jack Kiefer and many others, was done in the '50s. This was classical material, following Wald's book on statistical decision theory in 1950. And then of course Savage's book in 1954 and so on. This was a hot topic back then.

To begin, I wrote a short piece on the implications of partial identification for decision-making. The idea was simple, being that partial identification is a source of ambiguity. I initially wrote it for a conference on uncertainty and artificial intelligence.

**Elie**: In Belgium or somewhere?

**Chuck**: No, this conference was in the U. S., an annual conference on Uncertainty in Artificial Intelligence. I think you have in mind another group, mainly in Europe, called the Society for Imprecise Probabilities and Their Applications. I also got involved with that group. However, my first paper was not written for that group. It was for a group mainly of computer scientists who work on artificial intelligence and who try to develop algorithms that deal sensibly with uncertainty.

No one in econometrics was going to see that paper, so the more observable starting point for my work occurred when Art Goldberger retired at Wisconsin. We had a conference for him. I wrote a paper for the conference that expanded the ideas in the paper for the AI conference and explicitly related partial identification of treatment response to treatment choice under ambiguity.[39]

It is a simple conceptual paper that I value for making the basic observations that led to all the other work that I have done since then. That is, the identification region for an average treatment effect generates what a decision theorist would call the state space for the problem of treatment choice. When one's bound on the average treatment effect covers zero, both A and B are undominated treatment choices. Thus, partial identification generates a problem of decision under ambiguity. So, the revelation was to see that two topics that initially appear to be entirely separate, one in econometrics and the other in decision theory, really have a close relationship.

---

[39] Manski, C.F. (2000). "Identification Problems and Decisions under Ambiguity: Empirical Analysis of Treatment Response and Normative Analysis of Treatment Choice. *Journal of Econometrics*, 95(2), 415-442.

It is interesting that, in the special issue of the *Journal of Econometrics* that came out of Goldberger's conference, Gary Chamberlain also had a paper on decision theory. However, as you might expect, Gary's paper was orthodox Bayesian. Gary and I differed in our approaches but we shared an interest in the connection of econometrics to decision making, which was rare among econometricians.

The main topic I want to discuss now is not the difference between Bayesian decision theory and work on decisions under ambiguity of the type I do. The main topic is to ask why we do statistical and econometric research.

In the 1940s and 1950s, a prominent view among statisticians was that we do empirical analysis as an input to decision-making. That is, we face decision problems with incomplete knowledge, so we do empirical work to learn about the things we don't know. Then we use the empirical evidence to make decisions.

This was Abraham Wald's view. During the 1940s, he published a series of articles and then brought his work together in his book on *Statistical Decision Functions* in 1950. Wald's perspective was frequentist, but he was able to subsume Bayesian decision theory as a special case of his abstract framework.

I became aware of Wald's statistical decision theory when I was an assistant professor at Carnegie Mellon. I did not learn about it from the economists but from faculty in the statistics department. They were steadfast Bayesian, but they were well read and aware of the history of the subject. I am thinking, for example, of Morris DeGroot, who wrote a great book on *Optimal Statistical Decisions*. As much as I learned from DeGroot, I probably learned more from Jay Kadane, who had a rather general perspective on decision theory. For DeGroot, Kadane, and other Carnegie Mellon statisticians, the reason you do empirical research is to inform decision making.

**Elie**: Of course decision problems cover familiar situations like what estimator to use.

**Chuck**: Yes. In fact, that is part of what they did. They viewed choosing an estimator as a decision problem. We teach regularly that the mean is the best predictor under square loss. That is a decision problem.

**Elie:** It is clear that decision theory per se is fundamental to many areas of science, especially statistics, economics and econometrics.

**Chuck**: Personally, I have been interested in real-world applied decision-making, but absolutely, standard estimation can be approached as a decision problem. A classic book giving that perspective is by Ferguson in 1967. It sets up the familiar idea of viewing a point estimate as a best point predictor under some loss function.

What has always been harder is to motivate hypothesis testing. The question is: "What decision problem would motivate a classical hypothesis test?" This is hard to do except in special settings. Perhaps the only really clean one is the setting of the Neyman-Pearson lemma, where one has a simple null and a simple alternative hypothesis. One can view Neyman-Pearson as showing the optimality of likelihood ratio tests, without explicit reference to a decision. Alternatively, one can view Neyman-Pearson as studying a simple decision problem. You have to choose between two actions. Choosing one is equivalent to accepting the null hypothesis and choosing the other means accepting the alternative. Having simple nulls and alternatives means having a state space with just two elements.

The point I want to make is this. I think that econometrics and statistics took a bad turn after the 1950s, because the fields gradually lost interest in decision making as the motivation for empirical research. The general Wald theory lost ground by the 1960s. Bayesian statistical decision theory remained active for longer, but in a way that became detached from the frequentist world view of Wald. More recently, Bayesians have become less concerned with decision making. Today they focus on the computational problem of transforming priors into posteriors via the likelihood function, typically without reference to a decision problem.

**Elie:** Can you comment on your view of Bayesian decision theory?

**Chuck:** I have no problem with Bayesian statistical decision theory in circumstances where you feel comfortable using a full prior. To me, this is related to partial identification, in the sense that placing a fully specified subjective distribution on the state space assumes a lot of knowledge. Bayesian knowledge is probabilistic rather than deterministic, but it still puts a lot of structure on the world.

The question is, when in practice do we feel comfortable specifying a full prior? Obviously, I have found this sensible in my work measuring expectations, where I have regularly elicited subjective distributions from survey respondents. In that research, we typically ask persons about matters that are close to their lives---job loss, future income, etc. I tend to be less comfortable when researchers assert full priors on the parameters in econometric models. The topics of the research may be distant from researchers' experience, so they often use priors that are convenient rather than credible. They often use so-called "noninformative priors," even though every prior actually is informative.

Regarding the history, publications on Bayesian statistical decision theory continued through the '60s and '70s and then they started to drop off. There was some work in the '70s and '80s on robust Bayes analysis, where you move away from the idea of a full prior by allowing multiple priors on the state space, but this did not catch on. There has been a rebirth of Bayesian work in more modern times, but it is not about decision theory at all.

**Elie**: It's about computations.

**Chuck**: It's about computation, exactly. When people say, "Oh, Bayesian, everyone's doing Bayesian work now," they mean that they are using MCMC. MCMC is simply a computational devise for going from prior to posterior. You can do that without any decision problem in mind. Even though a lot of people use the mechanics of going from prior to posterior, it is not common today to find Bayesian updating embedded in an actual decision problem. Economic theorists use Bayesian updating this way, but empirical researchers do not.

Returning to Wald, his view of the world was totally lost. After writing my paper for the Goldberger conference and the special issue of the *Journal of Econometrics*, I decided to try to do something serious on the subject. This led to what I view as one of the most important papers I have written, the 2004 *Econometrica* paper[40] on "Statistical Treatment Rules for Heterogeneous Populations." The paper took me several years to write. It firmly established the conceptual framework I wanted to use, beginning from the general ideas of Wald and then focusing on the minimax-regret criterion. I used large deviations theory to prove some results that are useful in

---

[40] Manski, C. F. (2004). Statistical treatment rules for heterogeneous populations. *Econometrica*, *72*(4), 1221-1246.

practice. The technical analysis went only part way to what I would have liked to achieve, but Jörg Stoye later pushed the technical analysis much further, publishing two brilliant papers in the *Journal of Econometrics* in 2009 and 2012.

When I was writing that paper, whose form changed a lot through multiple versions beginning in 1999, I found it hard to give seminars on the topic. This remains so even today. Most econometricians know nothing about the Wald theory. I have to take the first half hour of a seminar to explain basic ideas like admissibility, finite-sample minimax regret, and so on.

The Wald theory is taught by only a few of us. I teach it and Gary does. I think that Chris Sims may teach it. Wald otherwise is not taught in econometrics courses. This also holds in statistics. Whenever I talk with statisticians, I ask them whether statistical decision theory is taught in Ph.D. statistics courses and the answer almost always is negative.

**Elie:** You are not talking about using this decision theory framework to frame asymptotic arguments about optimality.

**Chuck**: No, that's not my interest. You are talking about the asymptotic version of Wald associated with Hajek and Le Cam. After I wrote my *Econometrica* paper on statistical treatment rules, Hirano and Porter wrote their 2009 *Econometrica* paper[41], which is in the asymptotic tradition. Then there is a new paper[42] of Alex Tetenov and and Toru Kitagawa, which is now forthcoming in *Econometrica*. Their analysis is partly finite-sample and partly asymptotic. You are right that the asymptotic minimax concept is a residue of the Wald theory, but it is a fairly esoteric kind of analysis.

**Elie**: The people that work it view it as the thing to do, but it's clearly using that framework to say, "What's the best we can do?"

**Chuck**: In a particular well-defined sense.

---

[41] Hirano, K., & Porter, J. R. (2009). Asymptotics for statistical treatment rules. *Econometrica*, *77*(5), 1683-1701.

[42] Kitagawa, T., & Tetenov, A. (2017). Who should be treated? Empirical welfare maximization methods for treatment choice. *Econometrica*, forthcoming.

**Elie**: Yes, absolutely.

**Chuck**: The question is, "Who would be teaching that?" Certainly, a first-year course is not going to teach that.

**Elie**: No.

**Chuck**: I wonder, even in a second-year course.

**Elie**: There is a literature in mathematical statistics and econometrics on optimality (some small sample) in the context of estimation and inference. So, for example, the choice of a procedure that leads to the shortest confidence interval, or the choice of an estimator that is best in that it is minimax, etc. Whenever one uses this language of optimality when making a decision, it seems to me that there is no escaping using tools and connections to vintage statistical decision theory. Generally, though, this literature has no inputs or serious connections to empirical work.

**Chuck**: It has no inputs. You're right. Still, when you're giving seminars, you might be lucky that someone that does that stuff is in the audience.

**Elie**: No, it's not common.

**Chuck**: This is an example of the more general point that science does not always progress monotonically. Sometimes there are areas of research that are important and well-known in a certain period of history. Then interests change and they are forgotten. They become forgotten because people stop teaching them and then you get generations of students who are unaware of them.

It's up to us, because we teach the graduate courses, to decide what topics we'll cover and not cover. From the mid-2000s on, I have devoted a week or two of my course each fall to the implications of partial identification for treatment under ambiguity. I devote a week or so to the Wald theory. Chapters 11 and 12 of my text *Identification for Prediction and Decision* cover the material. There are only ten or eleven weeks in a quarter, so there's only so much you can do.

Returning to the 2004 *Econometrica* paper, the topic there is not identification at all. It is about use of the outcome data from a classical randomized experiment to inform treatment choice. The issue is finite sample precision, not identification.

**Elie**: Can you talk further about the conceptual problems that you analyzed there?

**Chuck**: There are two classic ways of using data from a randomized experiment to make decisions. One is the plug-in method, where you use your estimate of the average treatment effect. This is the usual analogy principle, replacing population means by sample averages. The approach has well-known good asymptotic properties based on laws of large numbers and central limit theorems, but they are only asymptotic properties, not finite sample ones. I was particularly concerned with how plug-in methods work with reasonable sized samples.

The other way of using experimental data is to perform a hypothesis test. You view one treatment as the status quo and the other one as an innovation, let the null hypothesis be that the average treatment effect is zero, and let the alternative be that it is larger than zero. You choose the status quo unless you reject the null hypothesis. This makes little sense from the perspective of decision theory, but it is a common practice.

What I did in the 2004 paper was to ask: "WWWD?" That is, "What Would Wald Do?" Part of the reason why it took so long to write the paper was that I had to think about how to set up the decision problem. In retrospect it's not hard, but it took a while because I had to understand the Wald framework more formally than I had previously and then see how to use it in my setting. Some of the issue was simply to define notation in a helpful way.

Once you structure the decision problem, the first technical question is to ask what statistical decision functions are admissible. A statistical decision function is simply a mapping from the available data to a decision. I called it a statistical treatment rule in my paper. It turns out to be hard to prove admissibility except in some special cases. A classic one is in a 1956 *Annals* paper by Samuel Karlin and Herman Rubin, who consider a specific context in which the monotone likelihood ratio property holds. More generally, it is easy to show that Bayes rules, maximin, and minimax-regret rules are admissible whenever they yield a unique solution.

The second technical question is to choose a specific decision criterion for study and work out its properties. I first thought about maximin and could easily show that it is poorly behaved in the context of treatment choice. Basically, it instructs you to entirely ignore the data. I showed this on my own and found later that Savage had hinted of the result way back in 1951, within his *JASA* review of Wald's book.

**Elie**: [laughs]

**Chuck**: The source of the problem is the ultra-pessimism of maximin. I next thought, "What about minimax regret?" I started playing with it and found that it is much better behaved. In a way, the process by which I converged on minimax regret was similar to what I did in my original maximum score work. I mucked around for a while, entertaining various decision criteria, and trying to learn their properties.

I found that minimax regret works nicely for treatment choice. At the same time, I was aware of a classical axiomatic argument against it, made by Chernoff in a 1954 *Econometrica* article. Chernoff noted that minimax regret does not always satisfy an axiom called independence of irrelevant alternatives. The axiom says that if I have three alternatives (A, B, C) and if I choose A over B and C, then I should still choose A over B if C is eliminated from the choice set. This seems a sensible requirement for a decision rule, but I later concluded that the appeal is superficial. I argued this in my 2011 article[43] in *Theory and Decision* on "Actualist Rationality." Chernoff's critique of minimax regret was one paragraph long. He just said: "Look, minimax regret doesn't obey independence of irrelevant alternatives. Therefore, it is unacceptable."

**Elie**: He has a little more on his criticism in his book with Moise called Elements of Decision Theory.

**Chuck**: Is there? I didn't notice.

**Elie**: Yes there is especially through simple examples.

---

[43] Manski, C. F. (2011). Actualist rationality. *Theory and Decision*, *71*(2), 195-210.

**Chuck**: I've only looked at the *Econometrica* article. Anyway, he says, "It doesn't obey this axiom. This axiom is self-evident, therefore minimax regret is stupid," and that's it.

**Elie**: There is no identification problem whatsoever because you have the randomized experiment.

**Chuck**: Yes. In my 2004 paper, it was just a matter of finite-sample imprecision. I thought this was the right place to begin my formal work because classical randomized experiments are a leading case of data collection and are simpler than cases with identification problems.

**Elie**: Is the randomization there coming from flipping the coin?

**Chuck**: It's just standard...

**Elie**: People have outcomes under treatment and non-treatment that are fixed, and the only randomness is just flipping the coin.

**Chuck**: Yes, this is the standard setup. Subjects are randomized to one treatment or another. Each subject has a deterministic potential outcome under each treatment, but you observe only the outcome for the realized treatment. You don't observe the other.

After publishing that paper, I wanted to begin analyzing some cases with identification problems. For example, you may have finite-sample observational data or an experiment with non-compliance. I have a paper in the *Journal of Econometrics* in 2007 that reports some partial results.

I should talk about Jörg's work. His thesis, leading to his 2009 *Journal of Econometrics* paper[44], was on classical randomized experiments. For a class of interesting cases he nailed what I was not able to show in my 2004 paper. That is, he constructively derived the form of the minimax regret decision rule. I had made some progress using large deviation bounds to derive upper bounds on regret, which enabled me to partially analyze the regret properties of plugin treatment

---

[44] Stoye, J. (2009). Minimax regret treatment choice with finite samples. *Journal of Econometrics*, *151*(1), 70-81.

rules, which I called "empirical success rules." However, I couldn't solve the full minimax regret problem. I didn't know how to do it.

Jörg had an unusual background, in that he was an econometrician who also had background in decision theory and game theory. Jörg was well aware of a device, which was obscure to me because I never had formal training in game theory, of representing (and this goes back to Wald in the '40s) the solution of a minimax problem...

**Elie** a game

**Chuck**: ...as a game. You may know this well, because you have studied game theory.

**Elie**: There's a book in the '50s by Blackwell and Girshick.

**Chuck**: Blackwell and Girshick[45], yes, in 1954.

**Elie**: That's all in there.

**Chuck**: It's all in there. It's some very early stuff.

**Elie**: It's a hard book but nonetheless much of the conceptual material is through the lens of a statistical game.

**Chuck**: That's exactly right.

**Chuck**: It was an extraordinary period. This approach was originally developed to solve for minimax actions, but it works for minimax regret as well. It doesn't give you a full recipe to solve these problems but it gives an idea that sometimes succeeds. You represent the decision problem as a two-person game in mixed strategies against a malevolent nature. Wald had shown that these games can be represented in Bayesian terms, the strategy of nature being the so-called "worst-case prior." You basically have to guess what the worst-case prior is and then plug it in and then see if it solves the game. You have to be creative, using intuition as to what the worst-case prior is going to be. That is what Jörg did.

---

[45] Blackwell, D. A., & Girshick, M. A. (1979). *Theory of games and statistical decisions*. Courier Corporation.

The second paper of Jörg, you should know. It is in the special issue[46] of the *Journal of Econometrics* based on my 60th birthday conference. I think the second paper is even better than the first. It considers data generation processes that have interesting identification problems as well as finite-sample imprecision. He uses the game theoretic device to determine the minimax regret solution when you have both issues.

If you are dealing both with finite-sample imprecision and with identification, and you ask how the decision rule should weigh those two issues, it turns out that the functional form of minimax regret has a discontinuity. If the sample size is small enough, the finite sample imprecision is dominant and you ignore the identification problem. Once the sample size gets large enough, then the identification problem becomes dominant. It is as if a switch flips. The identification problem becomes the concern and you don't worry anymore about statistical imprecision. You might think the result would be more continuous. Intuitively, as sample size goes up, the identification issue becomes more dominant than statistical imprecision. However, in the context of this treatment choice problem, the switch is discontinuous. That was surprising. Jörg proved it.

The discontinuity result is specific to the treatment choice setting rather than general. I know that now because Jeff Dominitz and I have a new paper[47] in the *Review of Economic Studies* on the problem of estimating the mean of a population in the presence of missing data. In this case, regret is mean square error and so the minimax regret problem is to find a statistical decision function that minimizes maximum mean squared error. The paper is called "More Data or Better Data?" because we consider a setting in which you can collect data in two ways. You can use an expensive sampling process to obtain a complete random sample without any identification problem, only statistical imprecision. Or you can collect data more cheaply but then experience some missing data, which generates an identification problem.

---

[46] Stoye, J. (2012). Minimax regret treatment choice with covariates or with limited validity of experiments. *Journal of Econometrics*, *166*(1), 138-156.

[47] Dominitz, J., & F. Manski, C. (2017). More Data or Better Data? A Statistical Decision Problem. *The Review of Economic Studies*, 84(4), 1583-1605.

The question is how to allocate a data collection budget if the objective is to minimize maximum mean square error. Our results are more continuous than what Jörg found. If you have a larger budget, you should pay more attention to the identification problem. A larger budget enables you to draw larger samples, which reduces the problem of statistical imprecision. The transition is continuous, not an on-off switch.

Comparing Jörg's work with my paper with Jeff highlights a general point that findings in statistical decision theory can depend strongly on the mathematical form of the objective function. The objective function of welfare maximization that has been assumed for analysis of treatment choice is different than mean square error. You have to analyze these problems separately.

I should mention one further new paper[48], which I co-wrote with Alex Tetenov and that appeared in *PNAS* in 2016.

This paper returns to the setting of a classical randomized experiment. Instead of asking how to use the data from an existing experiment to make a decision, the new paper asks how to choose sample size for a new experiment.

The standard approach, particularly in medical trials, is to choose sample size to give desired statistical power for a hypothesis test that compares treatment A and treatment B. Typically, you permit a five-percent chance of a Type I error and you choose sample size to obtain a twenty percent chance of a Type II error for a specified alternative hypothesis. This convention is widespread. You go to the website of the Food and Drug Administration or to other institutional websites and they tell you to choose sample size this way.

Part of the problem with hypothesis testing is this asymmetry. Why should you be willing to accept a twenty percent chance of a Type II error but only a five percent chance of a Type I error? It is only a convention. Another problem with hypothesis testing is that it takes no account of the magnitude of losses. If you make a Type I error, you choose Treatment B when

---

[48] Manski, C. F., & Tetenov, A. (2016). Sufficient trial size to inform clinical practice. *Proceedings of the National Academy of Sciences*, *113*(38), 10518-10523.

Treatment A is better, so you suffer a certain loss. If you make the other error, there may be a different loss.

**Elie**:  [laughs] Yeah, you may die, you have one loss.

**Chuck**:  Yes.  One loss may be that you take a few pills that you don't need to take and you have a few side effects.  The other loss may be that you die. Hypothesis testing doesn't deal with this asymmetry.

**Elie**:  It's simple.

**Chuck**:  Yes, it's simple. I've been railing against this for a long time. The Bayesian statistical decision theorists have railed against hypothesis testing for much longer, for the same reasons.

What Alex and I did is to say: Let's look at this choice of sample size problem from the Wald perspective. Formally, let the criteria be to get maximum regret below some predetermined level. A larger sample size enables you to reduce maximum regret.  We changed the semantics in the paper because maximum regret seems abstract.  We use the term "near optimality." Nearness to optimality is measured by maximum regret, which is the maximum loss relative to the optimum that is achievable, in a situation where you don't know the true state of nature.

If you knew the true state, you would make the optimal choice. The problem is that you don't know the true state, so you can't optimize. If am able to find a decision rule that makes maximum regret less than a specified $\varepsilon$, then regardless of the true state of nature, I am within $\varepsilon$ of the optimal decision. That's formally what maximum regret means.  Near optimality is just a different way of phrasing it.  The formal results in the paper are sufficient conditions to enable near optimality for any specified $\varepsilon$.

To sum up, we might ask where the new literature applying statistical decision theory stands today. I think we are still in an early stage.  There have been a set of important papers, but not yet a groundswell of research.

**A view towards the furture**

**Elie**: If we switch now to the future, I think probably the readers of this would like to know your predictions of what's going on with the field. [laughs]

I think we'd like to see what you see coming for future generations. What's the role of, say big data that you see or you've seen so far? What things are happening in econometrics that may be impacted by these modern approaches to analyzing data?

**Chuck**: Before even mentioning the terms big data or machine learning, there are some things that I think are wrong, and a lot could be done to fix them. It would be useful simply to change people's orientation. I worry enormously about the amount of effort that econometric theorists put into more and more obscure forms of local asymptotic theory.

It used to be that you did it with at least some hint of an application, but now you can do it without having anything in mind except the most toy application. I take a very strong position on this, that a lot of this is mathematics for the sake of doing mathematics.

I think it is dangerous for econometrics as a field. Students who come on the market as econometricians are well-trained in mathematical statistics, particularly in various forms of asymptotics, but they often have no background in economics. The field becomes inward looking. Papers are accepted for publication based on how tough the proofs are rather than on whether the theorem is of interest. It leads more and more to a situation where econometrics is isolated from the rest of economics departments, where economists don't have a clue what's going on when an econometrician gives a seminar on the job market. They can't make any sense of it, and they can't find any reason to care about it. They begin to wonder whether they should hire econometricians.

Viewing this future, the first thing that I think should be done is to get econometrics back to being useful for applied economic research. We should not be a self-contained field. We are supposed to be useful to empirical research in economics.

I know that I don't need to say this to you, because I know you feel the same way. I think we both know that a large fraction of what goes on is not like that. For the future, getting rid of that

negative is very important. I think that's the most important task for econometrics in the years ahead.

**Elie**: Other than the obvious, do you think there may be some things that could be done, like a foundational conference where the elders of the profession can come together and decide the kinds of questions that the young need to be looking at, or along these lines?

**Chuck**: It might be useful. The elders are not going to agree with each other.

**Elie**: That's OK, but at least...

**Chuck**: ...but at least to thrash it out in public.

I think there's a crisis in econometrics.

**Elie**: We need to be especially careful also now with the big data push.

**Chuck**: I still do not have a clue what big data means. I have been through multiple iterations of bigger and bigger, richer and richer data in my own career. As an undergraduate in the late 1960's and in graduate school in the early '70s, I knew that before the late 1960s, there were no microdata sets except for the CPS. There was a revolution in the late '60s and early '70s when the PSID, the NLS datasets, and others were begun. These certainly qualified as big data back then. I always thought that a Nobel Prize should have been given to the originators of the PSID and the NLS data. This work was not sexy, but the people who created those data sets provided the infrastructure for doing modern microeconometric work.

Of course, over time you get more and more data, richer and richer data, you learn more and more. If we think about IO, at some point you start getting the scanner data and so on. Is that big data? I don't know what...

**Elie**: No.

**Chuck**: Now they mean Facebook data. They also refer to administrative data, but we always had administrative data.

**Elie**: I think probably, the angle is...

**Chuck**:  Lots of covariates? I don't know what the...

**Elie**:  The angle is more about emphasis on prediction problems from the same population that can be done on a much much larger scale.

**Chuck**:  This goes into so-called machine learning which, of course, drives me crazy because machine learning is just regression.

**Elie**:  In a computable way.

**Chuck**:  It depends what computable way means. Was it Galton who invented the word regression? I think it was Galton in the 1870s. As far as I can see, conditional prediction, regression, and machine learning are all synonyms for the same thing.

Over time what is computable increases smoothly. I remember Art Goldberger telling me about using slide rules and various tricks to invert a 3x3 matrix. The hard computational problems in one period become trivial later on.  Progress of this sort has been taking place for a long time.

**Elie**:  What about the emphasis on, literally, just prediction, because we can't do anything else, so why not do this, which we know we can do really well? The more data you get, the better we get because things get trained. Prediction is really the bedrock of this data revolution.

**Chuck**:  There are two things to say about that. One is that, in the context of straight prediction, in the same population, there's no identification issue, just straight...

**Elie**:  Same population.

**Chuck**:  Same population. The issue then is simply a question...OK, so you get a very large sample size and you get a lot of covariates.

If you go back to the linear regression literature, an ancient question concerns how to proceed when the number of observed covariates is large relative to sample size.  Every undergraduate learned about the relevance of "degrees of freedom."  There were some poorly motivated suggestions early on, like stepwise regression, but people were aware of the issue.  Then there were shrinkage estimators, which were better motivated.  And there were variable selection procedures.

The issue is, what exactly is new now? No matter how large a dataset you have, if you also have more covariates, you still run out of steam. Some people argue that Lasso provides a new approach to deal with these issues. However, I think Lasso is going backwards and I don't understand its appeal. I am not surprised that mathematical statistics find it interesting to do the math of Lasso, but I am shocked that they think that it solves the basic problem in a useful way. When I say that it is going backward, I have in mind that Lasso assumes linear models. I grew up trying to get away from linear models. As I mentioned earlier, I found kernel nonparametric regression to be a revelation.

**Elie**: Can you talk more about that?

**Chuck**: So we began doing nonparametric work and came to recognize the curse of dimensionality, which I think is fundamental. The Lasso literature attempts to circumvent the curse of dimensionality by assuming linear models and sparsity. Then they can get some results. However, even considering within-population conditional prediction, for what class of problems should one feel comfortable with linear models and sparsity? I find it hard to find economic applications in which these assumptions are credible.

**Elie**: Are you objecting to the mechanics of it or to the...?

**Chuck**: The assumptions.

**Elie**: ...making the center of the enterprise prediction only?

**Chuck**: I haven't even come to that yet. The point I am making now is that even if the objective is within-population conditional prediction, the Lasso approach rests on poorly motivated assumptions of linearity and sparsity.

There clearly are cases in which prediction is a useful objective. It is appropriate when one is doing personalized risk assessment in medicine or credit scoring of loan applications.

**Elie**: You see an X-ray and you want to predict whether somebody has a disease or not.

**Chuck**: Clearly there are situations with stable environments where you want to do conditional prediction. Let's agree that there are cases in which conditional predication from a stable population is something you want to do.

Let's say that conditional prediction is important for many applications. Is there a revolution here? No, there's some new stuff, but I don't see any revolution.

**Elie**: What about counterfactuals?

**Chuck**: Regarding counterfactual forecasting, of course, that's out-of-population. If there is going to be a policy change, a within-population prediction says nothing per se.

Something that makes no sense to me is the work that connects machine learning and causal inference. The two phrases have nothing to do with each other as far as I can see.

**Elie**: This is through the use of conditional exogeneity.

**Chuck**: If you want to do conditional exogeneity and if you take this view that if you condition it on enough stuff...

**Elie**: That's what it is. That's how they connect causal inference to...

**Chuck**: I don't take that seriously. If that's all it amounts to, then it's just...

**Elie**: How would they solve the identification problem?

**Chuck**: They can't, obviously. It's just a marketing job that they've...

**Elie**: The idea is that as we are able to include more and more covariates, then conditional exogeneity becomes more palatable.

**Chuck**: I associate that claim with your colleague Don Rubin. Rubin has pushed this his whole career. I have had so many discussions with serious people on this over the years. We ask: "For what class of real-world problems is there a credible basis for thinking that as you add more covariates you get closer to random treatment selection?"

**Elie**: Yes, maybe, biology, disease. I don't know.

**Chuck**: I don't know, but I haven't seen a formal...

**Elie**: If you observe all the genes.

**Chuck**: The problem is, is that if you just add more covariates, then, basically, the unobserved stuff should be going to zero.

Fine. Let the unobserved stuff go to zero, but whatever is unobserved does not necessarily get closer to being random, as far as I can see. You can easily come up with counter examples. I give counter examples in my graduate text book and Goldberger did in his book. In these examples, when you add more covariates you move away from random treatment selection.

**Elie**: You mean when you add more X's, then Y1 and Y0 become less independent of D?

**Chuck**: Yes.

**Elie**: As you add X's.

**Chuck**: As you add X's. You can work out some clean examples. I did one with the reservation wage model of labor supply in my graduate text.

**Elie**: You seem to be asking for a formal model.

**Chuck**: The claim from the Rubin camp is that, as a general matter, if you add X's you get closer to exogeneity. It seems an article of faith. I don't see whether it does or it doesn't. If someone wants to come up with a good line of reasoning in a particular disease context, I would be happy to see some formal analysis. However, Rubin's claim dates back to the 1970s and I still have not seen formal analysis that justifies it.

**Elie**: The large-scale data and analysis that's happening also is estimating demand function without using instruments, because we have so much X's. We see how many websites you went to, where you went, so we can just regress quantity on price and include all these X's that are there and with that we've controlled the endogeneity.

**Chuck**: Is there a formal argument?

**Elie**: Not beyond conditional exogeneity.

**Chuck**: OK, so in terms of the future of econometrics, I think part of the future of econometrics should be to encourage skepticism about the claims that are being made regarding big data and machine learning. Econometricians should be skeptical and not get taken in.

**Elie**: Of course, in computer science there is a view that we don't care about causal questions, because we can run experiments as much as we want, and if we get it wrong today, we'll get it right, tomorrow because we learn.

**Chuck**: That's fine, in a way. That's a dynamic, that's the view of drug approval. It's sequential learning by experimentation. That's different. It is not intrinsically connected to big data. It is just sequential experimentation.

**Elie**: They don't even care about statistics. They say, "We're doing experiments. We can learn."

**Chuck**: If you have the ability to do sequential experimentation, that's a fine thing to do. I don't connect that to big data or machine learning. We talked about DeGroot, who wrote a Bayesian book on *Optimal Sequential Decisions*. This was 1970, long before the words "big data" and "machine learning" entered.

**Elie**: Why don't we do any of that? At least, formally, in econometrics, I haven't really seen a lot of this.

**Chuck**: Wald was one of the inventors of sequential analysis. In economic theory, the Bayesian perspective on sequential learning is routine, embedded in dynamic programming. It is hard to do sequential analysis from a frequentist perspective. Bayesian analysis is much cleaner.

I know some relevant papers. For the treatment choice problem, there are papers in *Biometrika* and other journals on what are called "adaptive randomized trials," which is sequential assignment of people to different treatments where they do Bayesian updating of the treatment allocation. You don't know much at the beginning, so you do a 50-50 split. As you learn from observing trial outcomes, you change the allocation. These are good formal analyses. We may not teach the material in econometrics, but that work is out there. However, it is not machine learning nor big data. It is just sequential Bayesian analysis.

I am hopeful that the bubble will burst soon. This is anecdotal, but I have talked to serious non-econometricians who have been subjected to this hype and they ask: "What's going on?"

We should be skeptical when computer scientists claim that machine learning "works." Anyone serious should ask: "How do you assess the claim that it works?" Computer scientists say that they don't need a theory. They say that they have good empirical experience that it works." However, you typically find that when someone says that it works, they selectively tell you what works and they leave the experiences in which it didn't work in the file drawer. There is publication bias.

To relate this to partial identification, often a researcher estimating a point-identified model makes some small change to the model and finds that nothing much changes. They say that they have performed "robustness checks." Have you ever heard anyone say they failed a robustness check? No. In all the seminars, everyone passes their robustness tests. I think that is prima facie evidence that they are not pushing hard enough. Anything will break down if you push it hard enough. They don't push it. It's the same thing when computer scientists say without proof that some machine learning algorithm "works" and don't tell you when it doesn't work.

**Elie**: You need theory.

**Chuck**: You need some theory. It's not a theorem as far as I know, but I believe that there is no nonparametric regression method that is uniformly best. Some are better in some cases and some are better in others.

I can relate this to work on partial identification. We know that when assumptions are weak we don't get much. People then criticize us and say, "The bounds are too wide." I always reply: "But it is important to understand the identifying power of assumptions. To do that, you have to know what you get with minimal assumptions and then you see what you get as you add assumptions." It is an important part of the work to learn when you can't say much, when identification breaks down.

These guys who just say, "It works," they don't do that. It's like they're in a marketing business, not science. There is no one marketing a product who says, "My product sometimes doesn't work." I don't think it would be a healthy future for econometrics to join the marketing game.

**Elie**:  There are some aspects of this literature on learning and predictions that is succeeding empirically. For example, the use of these learning algorithms in medicine, radiology etc.

**Chuck**:  It could be.

**Chuck**:  I can say something relevant about this. I have a new paper with the title "Credible Ecological Inference for Medical Decisions with Personalized Risk Assessment," This is a formal partial identification paper.  It is forthcoming in *QE*.

The paper addresses part of the radiologist's problem.  The radiologist will get the results of the X-ray or the scan and look at them. An old-fashioned radiologist will look at them and use what's called clinical judgment, which is subjective. He will look at the results, see things that show up, ask: "Is there a tumor or is it benign?" and reach some conclusion.  That is clinical judgment. The contrast would be using a statistical rule, in which you observe certain attributes of the X-ray and make a formal conditional prediction. Doing that is not new. Use of statistical prediction in medicine goes back to the 1940s at least.

**Elie**:  No. It's just a question of how to put a picture into the prediction problem.

**Chuck**:  Yes, it's how to measure the attributes of the picture. There is a psychological literature that goes back all the way into the 1940s and early '50s that compares the success of radiologists using clinical judgments with conditional prediction.  The statistical predictions back then were made with linear regressions or logit models, not with modern machine learning algorithms. Even so, the general conclusion of the psychologists' studies was that the statistical predictions were more accurate than clinical judgment.  A classic was a book by Paul Meehl in the '50s. Meehl and the other psychologists concluded that clinical judgments aren't trustworthy. They conjectured that the radiologists and other clinicians pay too much attention to obscure details in the evidence that they obtain from patients.

That was the conclusion back in the 1950s.  Clinical judgment may not improve over time. It is still the same subjective process today.  However, formal prediction methods can get better and the amount of data used to train them can get better.

**Elie**:  It's not going to eliminate radiologists, but at least it's going to be weighted more heavily.

**Chuck**: That's fine, but I don't view it as something revolutionary. I view it as a continuing process of improvement. In the 1950s, we only had X-rays on film and now we have more accurate digital X-rays. Now we have scans, MRIs, and CT scans. There is steady improvement in the quality of the data as the resolution of the scans improves too. We also have larger data sets over time, because more people have been subjected to scans. I think it's perfectly sensible to think prediction should improve in a continuous way. What I don't understand is the recent hype that says this has been a revolution, and that now we can make magically accurate prediction whereas before there was nothing. I see this as a much more continuous process, because there has been steady progress in estimation methods, algorithms and speed of computers and data availability.