

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 studies. 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 modeling which helped start the semiparametric 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

2 ELIE TAMER

1 *superb advisor and mentor to many graduate students. His legacy as a thinker,*
2 *researcher, and a teacher is an example in scholarship that is hard to follow.*

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

5 **Choice Modeling**

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

7 The first topic I want to ask about is choice modeling. I want to go back to what
8 I think is a landmark article, the 1975 multinomial maximum score.¹ What is the
9 intellectual inspiration for the article and this approach to choice modeling? And
10 in particular, did you think that there were some concerns with McFadden's con-
11 ditional logit in terms of its robustness to parametric restrictions which motivated
12 this work?

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

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

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

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

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

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

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

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

29 **Chuck:** Econometrics was a little bit behind statistics, but not that much be-
30 cause the Jennrich article was 1969. It was an *Annals* article, if I recall. So uni-
31 form laws of large numbers, which are the basic building blocks of everything to-
32 day... They didn't exist earlier. We knew basic ones like Kolmogorov–Smirnov,
33 but the uniform laws of large numbers that would allow you to do serious work on
34 nonlinear models were only coming into existence in the mathematical statistics
35 literature in the 1960s and early 1970s. What I was doing was on my own, and I
36 had essentially no formal statistics training. Anything I knew that I had learned
37 from Frank Fisher, who was wonderful on linear models. Frank was my advisor
38 and was a very serious thinker, but his world was entirely inside linear models.
39 That is what he taught, that is what everybody did.

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

41 **Chuck:** Yes, but that math was totally irrelevant. I was a physics major for a
42 year, then a joint economics and physics major for a while, but the kinds of math

4 ELIE TAMER

1 that you learn in physics were just not relevant. I didn't get tooled up mathemati-
2 cally until later. This was when I was an assistant professor and spent a semester
3 at Berkeley in the spring of 1976. Dan invited me there. By that time I realized
4 I needed more math. I almost killed myself, I sat through, obviously not taking
5 it for credit, but I sat through much of the first year Ph.D. sequence in math at
6 Berkeley.

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

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

17 What I was doing, going back to the question of maximum score, was just
18 mucking around asking why maximum likelihood works. In an intuitive way. I
19 remember just playing. I said, well okay, the likelihood function has some intu-
20 ition, but basically all the math was dealing with the log likelihood, not with the
21 likelihood function itself. I said, well, you look at the log likelihood function, you
22 maximize it, so the log likelihood function is just taking a monotone transfor-
23 mation of the probabilities, the log of the probabilities, and it was heuristically
24 intuitive that you'd want to maximize some monotone function of the probab-
25 ilities. And I started asking what other monotone functions might work? Instead
26 of maximizing the log of the probabilities, why not maximize the sum of the
27 probabilities? I couldn't make that work, so I started thinking about other ideas,
28 playing around with different functional forms. Looking at the structure of the
29 discrete choice model, I decided to focus on the assumption of IID disturbances.
30 I wanted to get away from the extreme value distribution, but I felt comfortable
31 with IID disturbances at the time. I saw that the IID assumption implied that the
32 higher $X\beta$ is in the systematic part of the utility function, the higher the choice
33 probability is going to be.

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

40 Actually, in the original version of maximum score, it was not just zero-one,
41 that you get one point versus zero. In the 1975 article, it's any ordinal transfor-
42 mation, an idea that no one has ever used since then as far as I know. The basic
43 reasoning concerned the ordering of the probabilities—as long as you have any

1 function that makes the ordering of the probabilities the same as the ordering of
2 the $X\beta$, it will work.

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

6 **Chuck:** Then it will work.

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

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

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

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

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

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

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

22 **Elie:** This is in contrast to the 1985 version³ which is beautiful. The 1975 article
23 needs a little thinking to penetrate. Forget the technical part.

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

26 **Elie:** But it was not your job market article?

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

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

32 **Chuck:** No. I know that in the more recent literature people say—who cares
33 about the betas? You care about prediction. But within the model I was working
34 with, the ordering of $X\beta$ gave the ordering of the choice probabilities. Having the
35 ordering of the choice probabilities would allow you to do partial predictions of
36 choice behavior. This was quite important to me later on in my article "Identi-
37 fication of Binary Response Models" in *JASA* in 1988.⁴ If you read that article,
38 you see that the interpretation of β was important to me, particularly what you
39 can get with β in terms of being able to extrapolate and do counterfactual predic-
40 tion. It depended upon which framework you're working in. In the case of binary

1 response I was working with quantile independence. With median independence,
2 you get a very simple ordering. If $X\beta$ is greater than zero, the choice probability
3 is greater than one half. That allows you to make a real prediction.

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

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

13 **Chuck:** Well, we're going to talk later about partial identification. Maybe it
14 wasn't entirely clear at the time, but there is partial identification of average ef-
15 fects embedded in maximum score. What you learn about β suffices to get an
16 ordering of the choice probabilities, but not to pin them down. So there is partial
17 identification of the choice probabilities. I think to this day that it is a useful form
18 of partial identification, including for the multinomial case.

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

24 **Chuck:** That motivated maximum score from the beginning. It was not only β
25 per se.

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

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

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

39 **Chuck:** Yes, we were all students. We were trying to model demand for higher
40 education. That was the term we were using at that time: "demand for higher
41 education." We started out trying to use neoclassical demand theory. It didn't
42 make sense, but it was what we knew. We were uncomfortable with neoclassical

1 demand because the issue wasn't how many units of schooling you are going to
 2 buy. It was rather what kind of school you choose, and whether you go to college
 3 or not. That is, do you go to Harvard or MIT or Northwestern or wherever? It
 4 was a qualitative rather than quantitative choice. But there was no literature at the
 5 time.

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



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

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

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

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

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

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

21 **Chuck:** This was my job market article. I became familiar with what Dan Mc-
 22 Fadden was doing and I eventually got comfortable with parametric maximum
 23 likelihood and discrete choice. But other economists found the conceptualization
 24 of discrete choice difficult, not just the technical aspects.

25 The idea of thinking of goods as attribute bundles would seem obvious. It
 26 wasn't entirely original to Dan. I talked with him about this and he told me that
 27 he was influenced by the work of Zvi Griliches on hedonic price indices and per-
 28 haps some other ideas. And then maximum likelihood. I'd go on the market and

1 I'd give seminars and people would say, "What's the dependent variable?" I said,
 2 "Well, a choice." But unless you can write it as $y = X\beta + \varepsilon$, people just didn't
 3 understand. I must have given pretty bad seminars.

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

6 **Chuck:** Yes, the choice set generation process. That turned into the third chap-
 7 ter of my thesis, on "The Structure of Random Utility Models,"⁶ which laid out
 8 the idea of choice set generation. That article is extremely hard to read too.

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

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

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

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

19 **Elie:** Even at JPE?

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

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

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

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

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

31 **Elie:** The Wisconsin days?

32 **Chuck:** Until the Wisconsin days. Maximum score in 1975 was a stan-
 33 dalone article. I did not try to extend it afterwards. I didn't really think much
 34 about it. My attention turned to other things. The work that had more success
 35 was on choice-based sampling, which was soon after 1975. That was accepted
 36 quickly in *Econometrica*.⁷ People understood it. I was also doing applied work in
 37 transportation—automobile choice modeling and other things like that. But max-
 38 imum score. . .

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

40 **Chuck:** Yes, I was doing structural parametric work. I know exactly when that

1 changed. It began to change when I was out of the country, in Jerusalem. That was
 2 the very beginning of the semiparametric literature. I called the second maximum
 3 score article “Semiparametric Analysis of Binary Choice.” I used the word “semi-
 4 parametric” because in 1982 or 1983, there had been an article using that term,
 5 I think by John Wellner. It was the first time I’d seen the word semiparametric.
 6 And I thought, “Ah! That’s what maximum score was doing.” It was semipara-
 7 metric. I decided to de-emphasize the maximum score criterion and emphasize
 8 the semiparametric aspect. I explicitly titled my article “semiparametric,” draw-
 9 ing the word semiparametric from the statistics literature, which was just coming
 10 into existence in the early 1980s. By that point, I was much more keyed in. I knew
 11 some of leading statisticians by that point. I met Wellner a few times. I was close
 12 with Peter Bickel, who had spent a semester in Jerusalem when I was there. He
 13 was doing adaptive estimation, which was also semiparametric. Following Peter, I
 14 wrote an article on adaptive estimation of nonlinear models. So I was getting into
 15 that frame of thinking—of trying, in a coherent way, to get away from parametric
 16 assumptions.

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

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

24 **Chuck:** I knew Roger Koenker’s article on regression quantiles. I didn’t make
 25 connections there because that was linear regression quantiles. But seeing what
 26 Jim could do for the Tobit model. . .

27 **Elie:** That’s interesting.

28 **Chuck:** Right.

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

31 **Chuck:** Yes. And I remembered that in my 1975 article, there is a small sec-
 32 tion specific to binary response that shows maximum score works with a median
 33 independence assumption. It is in the 1975 article, but I had not drawn any con-
 34 clusion from it. . . It was just a side comment as a special case. There was no open
 35 identification issue anymore when I began the 1985 article. The open issue was to
 36 be able to say something formally. . .

37 **Elie:** Now moving towards consistency.

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

41 **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 that I hadn't proved consistency in the first article...

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 articles 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 nontrivial. 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 articles 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.

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

2 **Chuck:** Think about your own job market article, your *Restud* article⁹ and
3 everything that followed from that. Of course, there are technical issues, but the
4 basic conceptualization is simple. It could have been recognized much earlier, but
5 it wasn't.

6 **Partial Identification**

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

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

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

14 **Chuck:** In retrospect, there was a formal piece of partial identification analy-
15 sis in the JASA article. Also in retrospect, there was partial identification in the
16 maximum score work, as the assumptions I made there partially identified choice
17 probabilities. However, the concept wasn't consciously on my mind until a spe-
18 cific event in the spring of 1987. My conscious study of partial identification starts
19 with a question from Irv Piliavin.

20 **Elie:** You actually thank Irv in the *Journal of Human Resources*¹⁰ article.

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

25 **Elie:** Is he still alive?

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

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

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

36 **Elie:** Descriptive duration model.

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

12 ELIE TAMER

1 He came to me and he said: “There are only two ways I know to deal with this.
2 One is to assume that attrition was random, but I don’t believe that. The other is
3 to assume a parametric selection model, but I don’t believe that either.” Although
4 Irv was not a technical researcher, he appreciated the issues conceptually. He then
5 asked me: “Isn’t there anything else that I could do?” He gave me an incentive.
6 He came up with a month of summer salary.

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

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

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

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

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

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

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

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

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

28 At some point, I stepped back and thought: “Let’s go back to basics.” This is a
29 way of thinking that I have used multiple times in my career when I find myself
30 stuck, and I am proud of it. I made it the theme of an essay I wrote in the *American*
31 *Economist* called “Unlearning and Discovery.”¹¹

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

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

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

41 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.

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

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

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

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

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

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

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

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

34 **Elie:** Dangerous.

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

37 Well, what happened was that I got screamed at simultaneously by Hausman
38 and Heckman. I then knew for sure that what I had done must be good. I don't
39 remember what Heckman said, but I do remember what Hausman said and I
40 have since quoted him by name in my 2013 book *Public Policy in an Uncer-*
41 *tain World*.¹² Jerry said: "You can't give the client the bound. The client needs
42 a point." He still believes this today. He is not upset with me for quoting him
43 because that was and still is his view.

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

10 Of course, there were precedents to partial identification that were already in
 11 the econometrics literature. Probably the earliest was in the 1930s, the Frisch
 12 bound concerning linear regression with errors-in-variables. There was a little bit
 13 on partial identification in Frank Fisher’s 1966 book on *The Identification Prob-*
 14 *lem in Econometrics*, where he briefly considered identification of linear simulta-
 15 neous systems with sign restrictions on parameters. There were other early prece-
 16 dents that you cited in your review article in the *Annual Review of Economics*.¹³
 17 However, there was no coherent body of work, just a scattering of findings.

18 **Elie:** Ed Leamer’s work comes to mind.

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

23 In the early 80s he wrote several articles that we would now call partial iden-
 24 tification analysis. His interest focused on identification and estimation of linear
 25 regressions in the presence of classical measurement error and other data issues.
 26 In the mid ‘80s he basically stopped doing econometrics and changed careers, to
 27 work in international trade.

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

29 **Chuck:** That article was joint with Klepper. It is a nice article in *Econometrica*
 30 in 1984.¹⁴ They generalized the Frisch bound with errors-in-variables from simple
 31 regression to multivariate regression. Soon after, Ed stopped doing econometrics
 32 completely. Why? I don’t know. He could have developed this further and he
 33 decided not to. I don’t know why he didn’t.

34 Regarding other precedents, I learned four or five years after my *JHR* article
 35 that in 1954 Cochran, Mosteller, and Tukey¹⁵ had done a simple analysis of miss-
 36 ing outcome data in their book that evaluated the statistical methodology of the
 37 Kinsey Report on sexual behavior. They focused on estimation of the probability
 38 of a binary response in the survey that Kinsey used. They framed the issue not di-
 39 rectly in terms of identification but rather as the mean square error of estimates of
 40 the probability when some outcome data are missing. They recognized that ran-
 41 domly missing data affect the variance of estimates but generate no bias issue. On
 42 the other hand, missing data generates potential bias in the absence of assump-
 43 tions about the missingness process. They made a nice observation with practical

1 importance, this being that the bias reduction from converting a nonrespondent
 2 to a respondent often is a much better use of survey money than increasing the
 3 sample size, which only reduces variance. However, they never followed up on
 4 it. In fact, Cochran later wrote in his textbook that the bounds were distressingly
 5 wide and he didn't think them useful.

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

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

8 **Elie:** There is also a 1989 article by Robins¹⁶ that uses bounds.

9 **Chuck:** Jamie and I independently got the same result on intersections bounds
 10 for an average treatment effect. Our derivations made different assumptions. He
 11 was writing about randomized experiments with noncompliance and I was writ-
 12 ing about use of observational data with an instrumental variable assuming mean
 13 independence, this being in my 1990 AER Proceedings article.¹⁷ However, the
 14 results we reported were algebraically the same. My result was a sharp bound
 15 under the assumptions that I made. Jamie's setup made stronger statistical inde-
 16 pendence assumptions and can yield a tighter bound, as Balke and Pearl showed
 17 later. However, Jamie's analysis did not use the full power of statistical indepen-
 18 dence, so he obtained the same bound that I did. The work that Jamie and I did
 19 was independent and essentially simultaneous. He and I didn't know each other
 20 at that time.

21 Another precedent, which I now teach regularly, is the ecological inference
 22 problem, which is classic. It is a different identification problem. Duncan and
 23 Davis¹⁸ derived a bound in the *American Sociological Review* in 1953. It's beau-
 24 tiful and simple. Yet another is the Peterson bound for the competing risk model,
 25 in 1976.¹⁹

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

29 **Chuck:** The biostatisticians concluded that you might as well assume indepen-
 30 dence, but that's not the conclusion I would draw.

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

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

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

39 **Chuck:** Yes, I stuck with it. I stuck my neck out, over and over. I didn't get
 40 much positive reinforcement. I was able to get my work published, sometimes
 41 with difficulty, sometimes more easily, all through the '90s. However, few people

1 paid attention to it. Even so, I stuck with it. It helped a lot that I had incredible
2 support from Art Goldberger. I also had good support from Joel Horowitz and
3 from a few students I advised at Wisconsin, particularly John Pepper.

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

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

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

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

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

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

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

30 **Elie:** Yes.

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

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

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

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

41 **Elie:** The articles and proceedings?

1 **Chuck:** The articles and proceedings. It is a simple, short article. Basically no
 2 one in the NBER labor group pays attention to it. The article gets citations from
 3 econometricians, but if you were to compare its citations or the one on monotone
 4 instrumental variables to that of Imbens and Angrist, there is an order of magni-
 5 tude difference.

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

9 We can act as empirical economists and say this is the revealed preference of
 10 the research community, that they'd rather use LATE than understand analysis of
 11 treatment response more fundamentally. When the LATE article came out, it was
 12 published as a note in *Econometrica*.²¹ I thought it well deserved to be published
 13 as a note. It was a cute mathematical result. I thought it was a one-off thing. I
 14 had published four years earlier the IV bounds on average treatment effects and
 15 I thought, "Well, it's clear the IV bounds are far more useful and interesting than
 16 this LATE thing."

17 **Elie:** You mean the articles and proceedings?

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

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

28 **Chuck:** Let me push on this a little bit. We were talking earlier about the prob-
 29 lem of econometrics getting too technical. Empirical people want to stay away
 30 because it's too technical and they see these regularity conditions that they can't
 31 interpret. However, my articles on bounds on treatment effects can't be blamed
 32 for being too technical. The ideas are extremely simple. There may be some in-
 33 tricacies regarding how to do confidence sets, but that is a second order issue. I
 34 really try to write for a wide audience. I wrote the 1995 book²² and the 2007 text-
 35 book to be easily accessible. There's really no excuse. The empirical fact is that
 36 the NBER labor economists act as if the bounds literature doesn't exist.

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

39 **Chuck:** There clearly has been an obsession with points, but it is frustrating
 40 to be left with this as a residual, psychological answer, in the absence of a rea-
 41 sonable logical answer. I knew the obsession with points from the beginning.
 42 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

1 partial identification has been resisted is an obsession with points. This remains a
2 common view in some fields even today.

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

7 **Elie:** Yes, it is the same conceptual problem as trying to use a confidence inter-
8 val.

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

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

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

23 **Chuck:** As a methodological question, I've tried to be open minded about this.
24 I know we're talking about the origins of partial identification, but I can date my
25 concern about what assumptions to make and what you get out of them to sev-
26 eral years earlier. When we were talking about maximum score, I said my second
27 round of work was influenced by Jim Powell and the LAD idea. That had a spe-
28 cific influence because of the quantile independence interpretation of maximum
29 score, but there was another event around the same time that, in a much broader
30 way, shook me up quite a bit back then. This was the discovery of nonparametric
31 regression.

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

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

1 I said at the World Congress that there is a vast frontier for econometricians
2 to explore, as you vary assumptions and see what you get from them. That idea
3 became important to me later and I eventually came to call it the “Law of De-
4 creasing Credibility” in my 2003 book on partial identification.²³

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

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

12 That’s the way I view the flexible functional form literature and some of the
13 work on discrete choice. Dan McFadden did this explicitly in the 1970s and
14 early ‘80s, going from the IID extreme value assumption to the generalized
15 extreme value model. Dan felt more comfortable thinking in terms of finite-
16 dimensional parametric models than I do, but he wanted to get away from the
17 strictness of the logit assumptions and so he went to the GEV model. The discrete
18 choice literature also did that with random-coefficients multinomial probit mod-
19 els, which everyone was talking about in the 1970s. The only reason they weren’t
20 applied very much was computational. Everyone would have preferred to do
21 that then.

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

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

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

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

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

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

34 **Elie:** Yes.

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

38 Regarding structural models, I have written two articles that pose things the way
39 I would like to do it. I have the article in the *IER* in 2007²⁴ on partial identification
40 of counterfactual choice probabilities.

41 **Elie:** McFadden’s special issue?

Chuck: Yes, in the special issue for Dan's retirement. Then there is the follow-up, the article 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* article. 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* article 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 article 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 article. I pushed it much harder in the income tax and labor supply context in the 2014 article.

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 article, 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 article, 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,

1 it's just mechanics now. But then, we start to worry about how much of this policy
2 response is due to functional forms.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

1 In the partial identification literature, this was important from the beginning.
 2 In my own work, it shows up most explicitly in my *Econometrica* article with
 3 Joel Horowitz on contaminated sampling.²⁶ We explicitly defined identification
 4 breakdown for that problem, which is a mixture decomposition problem. You
 5 get informative bounds if the prevalence of bad data is below some level. If the
 6 prevalence of bad data goes above some limit, the bound becomes a logical bound.
 7 It's not informative.

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

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

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

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

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

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

26 I only do two types of statistical research today. I sometime use large deviations
 27 analysis to obtain exact results for finite samples. I first used large deviations the-
 28 ory in my 1985 article on maximum score and I have found it to be useful in
 29 some recent articles using statistical decision theory to analyze randomized ex-
 30 periments. More generally, I do statistical decision theory, which is finite-sample
 31 and nonparametric.

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

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

38 **Elie:** There are hard statistical problems that people recently have tried to ad-
 39 dress. Really hard problems.

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

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

2 **Chuck:** My strong view is that I wish people would direct more of their ef-
3 fort towards work that is motivated by some real application. There may be hard
4 computational problems and statistical inference problems, but they would be mo-
5 tivated by the application.

6 An example of what I have in mind is the recent *QE* article²⁷ on insurance
7 purchases by Francesca with Levon. One can argue about whether you like their
8 specific model, but I think it is good work. When you asked earlier about structural
9 choice models with partial identification, I told you about my abstract work in the
10 *IER* article, and the one in *QE* on income tax and labor supply. I should mention
11 their *QE* article, which is a full-blown partial identification analysis of a structural
12 choice model.

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

18 **Elie:** Is it cars?

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

25 **Elie:** Yeah, yeah. There are many ways.

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

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

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

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

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

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

41 **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

1 have to ask yourself: “Can I actually do this?” You can write down whatever
2 abstract theorem you want, but if you don’t have an application in mind, then no
3 one puts you up against the wall.

4 If you have an application in mind, you are forced to confront it. The only
5 reason why I don’t give computation primacy as an issue is that computation
6 improves over time, both the hardware and software. The problems that people
7 view as hard today would have been out of the universe 10, 20, 30 years ago.
8 Problems that were killing us then are considered easy today.

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

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

17 **Elie:** So, we need to better confront the computational problem in the con-
18 text of a clear applied economics model. There are examples in the literature of
19 econometrics articles that try and do that.

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

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

30 **Chuck:** Where I think the journals can help is by being more realistic
31 and not requiring that someone nail these computational problems in their ar-
32 ticle. I think it would be better to ask “Do they make some progress on
33 it?” Then maybe someone will come along and do better. Go back to maxi-
34 mum score, Scott Thompson was my RA at Wisconsin in the mid ‘80s. We
35 had fairly primitive ideas on how to search for the maximum of the score
36 function.

37 **Elie:** Now it’s really easy.

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

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

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

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

7 **Elie:** Oh, yeah.

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

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

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

22 Your perspective on this may be different because you have an IO orien-
23 tation, and I think the IO people have been much more open to this. For
24 the kind of things that I do, I do some. John Pepper has done quite a lot.
25 John's work is good, he gets it published, but it has no impact on the so-
26 called mainstream. John and I have a new article on Right to Carry laws. It
27 uses bounded variation assumptions.²⁸ These are assumptions that relax mean
28 independence to suppose instead that a condition suitably close to mean in-
29 dependence holds. Articles like this get published. Maybe not in the *AER*
30 and not in the *QJE*, but serious applied partial identification articles can get
31 published.

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

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

40 **Chuck:** All they need to do is apply analysis that was done 20, 25 years
41 ago. There's no excuse whatsoever, as far as I can see. Partial identification has
42 been successful among econometricians. It has generated an enormous amount of

1 theoretical work and it has had some success in IO applications. But really, it
2 should be everywhere.

3 **From Jerusalem to Wisconsin**

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

5 **Chuck:** We moved to Jerusalem in 1979. We originally weren't intending to
6 stay. I was on leave from Carnegie-Mellon and we went there for a semester.
7 While I was there, they asked me to stay. Our children were young then, five
8 and two years old. And we thought, shall we go back to Pittsburgh or stay in
9 Jerusalem? Pittsburgh was safe and known, Jerusalem was an incredible adven-
10 ture, like no place else on earth, and we decided to do it. For me it was easier
11 being Jewish and knowing some Hebrew. For Kitty—not Jewish and not knowing
12 a word of Hebrew—it was a much bigger move. We decided we would go for four
13 or five years, see how it is, and then decide.

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

20 In 1982–1983, I went on leave to MIT. Dan McFadden was at MIT then, so
21 it was a natural place for me to visit. Our family was from the Boston area, our
22 parents wanted to see their grandchildren, so we came to MIT, and then we de-
23 cided what to do looking forward. Kitty and I had agreed that after we arrived in
24 Boston, we would have some distance from Jerusalem and then decide “Are we
25 going to go back there for good, or are we going to return to the United States?”

26 We decided to return to the United States and I put myself on the market. I
27 had offers from a bunch of places—from Wisconsin, Minnesota, Michigan, and
28 Northwestern. I went to Wisconsin.

29 **Elie:** Who was there?

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

39 The person I did know was Gary. We were not close friends, but we went
40 to high school together at Boston Latin School. And not only the same high
41 school. Gary and I were in the same classes for many years—calculus, Latin,
42 whatever, we somehow were tracked together. As undergraduate we were both

1 in Cambridge. I was at MIT both as an undergraduate and in graduate school.
 2 He was at Harvard undergrad and graduate school. We never saw each other
 3 as students. It was only when we both finished the Ph.D. that we found out
 4 that we both had decided to do econometrics. In the mid to late 1970s, I saw
 5 Gary periodically at conferences. So Gary and I knew each other. He was my
 6 main contact at Wisconsin, with Art Goldberger a more distant impetus to going
 7 there.

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

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

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

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

18 Art read my articles for three or four years. At some point, roughly around
 19 1986, he read one of my articles and said, "Okay, I'm not going to do it anymore.
 20 I've done all I can for you." And I said to him, "Does that mean I've graduated or
 21 you've given up on me?" He said, "You decide."

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

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

29 Art read and commented line by line, word by word, concentrating on get-
 30 ting across the message, the presentation. Of course, you can't do that kind
 31 of close editing unless you fully understand the concepts substantively. And
 32 Art was full of substantive contributions. This was during the period of my
 33 semiparametric work. There was a period in the mid-1980s when I went back
 34 to maximum score, and then there was a string of articles. It began with the
 35 1985 article and then the ones on maximum score with panel data and choice
 36 based sampling, a couple with Scott Thompson on computation and predic-
 37 tion, and finally the *JASA* article on identification. I had hit on a basic idea,
 38 for about four years I wrote articles on that theme, and Art would read all of
 39 them. I was doing other work as well, more applied work. Whatever it was, I
 40 would give it to Art, he would read it, and I would get incredible comments
 41 back.

42 After that initial period, for the rest of the time I was at Wisconsin, I would
 43 talk with Art about work as it was going on, but he didn't do line by line

1 reading anymore. Art became extremely important when I began working on partial
 2 identification. I dedicated my book on *Partial Identification of Probability*
 3 *Distributions* to him because he had faith in the idea and in the work. Even though
 4 I was able to get the early work on partial identification published, basically I was
 5 alone and no one was paying attention to the work. Art encouraged me to stick
 6 with it and keep going, regardless of whether people were interested in the work
 7 or not. He became very important.

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

16 Jim was in Madison for five years or so, and he did some of his best work
 17 then. He wrote his article on symmetrically trimmed least squares, which was a
 18 beautiful idea. He also co-wrote with Whitney an article that's not as well known,
 19 but also nice, on expectiles, which are the square-loss analog of quantiles.

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

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

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

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

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

36 **Chuck:** Well, people did write together sometimes. I think it's true that the
 37 prevalence of coauthoring has gone up over time, but we are not talking about the
 38 1600s here. We're talking about the 1980s. I think we each had our own agenda.
 39 I suppose that Jim and I might have coauthored since we were both focusing on
 40 semiparametric work from different directions, but at least it was on that topic.
 41 Gary also had related interests. He did his article on efficient method of moments
 42 and I had done my work on analog estimation, of which the method of moments
 is a special category.

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

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

9 **Elie:** Even with Art, did you ever write an article with him?

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

11 **Elie:** A book review.

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

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

14 **Chuck:** It is in the *Journal of Economic Literature* in 1995.²⁹ I initially did not
15 want to do this work. What happened was that *The Bell Curve* had come out in
16 print. Art, going back to the 1970s, had been a severe critic of the literature on
17 heritability, on attempts to decompose life outcomes into fractions explained by
18 genetics and by the environment.

19 There were two reasons why this was so important to Art. One was that the sta-
20 tistical and econometric analysis in the literature was horribly flawed. The second
21 reason was personal. The heritability research was put forward as ordinary neutral
22 social science, but some of it had a racial tinge. It was used to argue against anti-
23 poverty programs and particularly against helping African-Americans by making
24 the argument that they're just not bright and there is nothing you can do for them.
25 I won't say that everyone who wrote on heritability was a racist, but there was
26 a subset who seemed to be. If there was something that Art felt extraordinarily
27 strong about, it was about racial equality, particularly of blacks and whites.

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

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

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

40 Art and I both thought that the early reviews of *The Bell Curve* tended to be
41 superficial, mixing criticisms of the bad science with anger at the results that

1 Herrnstein and Murray reported and their objectionable policy recommendations.
 2 We wanted to do it right and we knew that this was going to take a long time. I
 3 remember saying to Art: “I’d rather do new research than review someone else’s
 4 research.” He said that it was for the good of society and really important. I agreed
 5 to do it. It took us several months.

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

11 For example, the book reported logistic regressions and concluded that genes
 12 are “more important” than environment. But what does it mean to say that genes
 13 are more important than environment? There is no common unit for measuring
 14 genes and the environment. A common practice in sociology, used in *The Bell*
 15 *Curve*, was to take the cross-sectional standard deviation of some real-valued
 16 measure of genes and environment as a constructed common unit and to define
 17 “more important” in these terms. That is, researchers would report how much
 18 an outcome like years of schooling or wage would change with a one standard-
 19 deviation change in the measure of genes and environment, and they would draw
 20 conclusions about relative importance this way.

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

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

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

42 We already talked about how partial identification came out of my contact with
 43 Irv Piliavin and his specific question about missing data in a longitudinal survey.

1 What we haven't talked about yet is that about a year later, the Institute for Re-
 2 search on Poverty (IRP) at Wisconsin needed a new director. I had never done
 3 poverty work. I was just a straight econometrician. But someone suggested to me
 4 that I become director. I decided to apply for the position and they asked me to do
 5 it.

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

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

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

25 The other reason why I did it was that by the late '80s, I concluded that I was
 26 reaching diminishing returns on the semiparametric work. I had a long stream of
 27 publications in the mid-80s on maximum score and related topics, culminating
 28 with the 1988 JASA article. After writing that article, I thought I understood the
 29 literature the way I wanted to and I became worried that I had run out of inter-
 30 esting work to do on the topic. Diminishing returns can be a serious problem for
 31 people doing econometric theory. Even the best research stream transitions from
 32 a period of excitement and productive work to one where it eventually runs out of
 33 steam. I had built up enough human capital in semiparametrics that I could have
 34 cranked out further articles for several more years, but I worried that they would
 35 be increasingly marginal. I was looking for a new source of ideas and thought that
 36 IRP might provide them.

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

1 new stream of research. All were generated by having to deal with real world
2 topics at IRP.

3 **Work on Expectations**

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

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

11 The specific way it started was that I thought demographers were misinter-
12 preting the responses to a question on the Current Population Survey asking
13 women about fertility intentions. The question was: "Do you expect to have more
14 children?" The allowed responses were "Yes, no, and uncertain." I sat in on semi-
15 nars in which demographers analyzed the data and used the responses to forecast
16 what future fertility would be. The practice was to compute the fraction of "yes"
17 responses and conclude that this would be the fraction of women who will actually
18 have children in the future.

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

31 It was a simple point, but apparently it was not in the literature. I dug deeper
32 into the matter and wrote a full-scale article³⁰ on "The Use of Intentions Data
33 to Predict Behavior" that was published in *JASA* in 1990. I ended the article by
34 arguing that the CPS question is too crude to be useful. I recommended asking
35 women to state the percent chance that they would have more children.

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

37 **Chuck:** Yes. Thinking about fertility intentions was part of what led me to con-
38 clude that we should ask probabilistic expectation questions more generally. The
39 other issue that drove me to it is that for years I had been going to labor econo-
40 metrics seminars on structural modeling of choice under uncertainty. I am think-
41 ing particularly of the work by Ken Wolpin, Zvi Eckstein, and Mike Keane—the
42 whole set of articles on dynamic choice of years of schooling under uncertainty.

1 I valued their work as good applied structural econometrics, but I did not like
2 their practice of making assumptions about expectations and supposing that peo-
3 ple have rational expectations.

4 In the schooling context, the assumption meant that people making decisions
5 have objectively correct probabilistic expectations for the returns to schooling,
6 conditioning on the information they have. I asked myself: “How would they
7 achieve rational expectations? Labor economists don’t know the objectively cor-
8 rect returns to schooling, so how is it that ordinary people could?” It seemed clear
9 to me that it was overly optimistic to assume rational expectations.

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

20 I concluded that the structural econometric models assuming rational expec-
21 tations are likely misspecified. They were making strong and potentially wrong
22 assumptions about expectations and estimating random utility models based on
23 these potentially wrong assumptions. To make this point I wrote a second article³¹
24 on “adolescent econometricians.”

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

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

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

41 **Chuck:** Yes.

42 **Elie:** Measurement error type?

1 **Chuck:** The reasoning that was given, and I don't want to downplay it, was to
 2 say that people don't have an incentive to take expectations questions seriously.
 3 They don't have an incentive, so why should you believe the response.
 4 Now there is a literature on so-called "proper scoring rules" where you give
 5 people an incentive to respond correctly.

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

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

13 I decided that the incentive issue may be legitimate, but not so severe as to make
 14 data collection useless. I realized that the incentive issue applies to all survey
 15 questions, not just to questions about expectations. I thought, "If you are going
 16 to be too serious about the incentive issue, then we wouldn't have any micro
 17 data from surveys at all." I felt that surveys have been useful, but you have to be
 18 cautious. So we should try to ask about expectations. I always kept in mind that
 19 the only alternative to measuring expectations was to make assumptions. It did
 20 not seem appealing to me to make assumptions about expectations without any
 21 data.

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

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



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

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

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

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

10 We did two things. One was a small pilot study on the returns to school-
11 ing. Jeff personally went to high schools in the Madison area. We got permis-
12 sion from high school principals and teachers to go into classrooms and elicit
13 from high school students their expectations of the returns to college atten-
14 dance. Our article was published in the *Journal of Human Resources* in 1996.³²
15 Since then there has been much more work eliciting returns to schooling, par-
16 ticularly by Basit Zafar and Pamela Giustinelli, both of whom I advised at
17 Northwestern.

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

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

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

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

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

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

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

3 **Elie:** Yeah, HRS.

4 **Chuck:** I view Tom as the father of measuring probabilistic expecta-
5 tions. Tom in 1966 in *JASA* published an article recommending collecting of
6 probabilistic data for buying intentions. This is important and Tom's work has
7 been underappreciated.

8 When I learned about it later—I only learned about Tom's work in the late '80s
9 when I started doing my own—I was embarrassed because I knew Juster from
10 times I had visited Michigan and I had no idea he had published this article until
11 25 years after he had done it.

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

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

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

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

41 Another independent source of data arose in Italy, where the Bank of Italy
42 began in 1989 to ask questions about income expectations in its Survey of House-
43 hold Income and Wealth. They asked households to forecast their incomes in the

1 year ahead, essentially to give subjective income distributions. Luigi Guiso and
 2 other Italian economists were involved in designing this survey and in analyzing
 3 the data.

4 So, in the early '90s there were at least three groups who independently thought
 5 that asking probabilistic expectations would be a good thing to do. We became
 6 aware of each other and the work started taking off. In the beginning, there was
 7 a lot of skepticism and it wasn't easy to get articles published. Many economists
 8 didn't want the expectation data because of the incentive issue we discussed. They
 9 said, "You just can't trust this."

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

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

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

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

32 It turns out, and the sociologists weren't aware of this, that some psychologists
 33 in the area of judgment and decision had been asking probabilistic questions for
 34 100 years, but that was a separate literature. Early on, the psychologists ques-
 35 tioned experts like weather forecasters, who would forecast the chance of rain.
 36 More recently they have questioned college students in experimental labs, and so
 37 on. The psychologists had been finding no problem asking probabilistic questions.
 38 Eventually I discovered the literature and got to know some of the psychologists
 39 who were doing this.

40 There are two differences between what the psychologists were doing and what
 41 we economists wanted to do. First, the psychologists would use convenient sam-
 42 ples of students, whereas the economists wanted to do national surveys to get
 43 representative samples. Second, the psychologists were intrinsically interested in

1 how people answer the questions, whereas the economists wanted to use the data
2 to help understand choice behavior. There were differences in what we did and
3 why we did it, but we still could learn from each other.

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

11 Jeff and I had a breakthrough in '97. We had a basic article³³ on the Survey
12 of Economic Expectations published in *Public Opinion Quarterly*, which is the
13 main journal in survey research. We also published an article that year in *JASA* on
14 measurement of income expectations.³⁴ Then people could say there are serious
15 journals that are willing to publish this work. It went on from there and has since
16 grown into a large literature.

17 I gave the Fisher-Schultz Lecture for the Econometric Society on the topic
18 "Measuring Expectations" and published the resulting review article³⁵ in *Econo-*
19 *metrica* in 2004. This year I was able to substantially update it when I was asked
20 to prepare an article on macroeconomic expectations for the 2017 NBER Macroe-
21 conomics Annual volume. The new article focuses on topics that should be partic-
22 ularly of interest to macro people, such as expectations of equity returns, inflation
23 expectations, and expectations of professional macroeconomic forecasters.

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

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

27 One of the first to use probabilistic expectation data in structural modeling was
28 Adeline Delavande, whose dissertation at Northwestern was on contraceptive be-
29 havior. She considered young women who are sexually active and choose among
30 different contraceptive methods. They make decisions under uncertainty. A cen-
31 tral uncertainty is the possibility of getting pregnant if they use one contracep-
32 tive method or another. There are also issues of sexually transmitted disease and
33 so on.

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

39 Social Interactions

40 **Elie:** Now, we get to your most cited article which is within this exciting area
41 of inference on models with social interactions. Can you speak about this work
42 and the *Restud* article³⁶ in particular.

1 **Chuck:** We talked earlier about how the three years I spent as director of IRP
 2 at Wisconsin affected my research greatly. One outgrowth was the work on partial
 3 identification, which came directly out of thinking about attrition in a longitudinal
 4 survey, and another was on measuring expectations. I described how that came
 5 out of reading about how demographers analyzed fertility intentions. These are
 6 applied problems that, as an econometrician, I would not probably have seen on
 7 my own.

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

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

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

32 Sitting through these seminars, I realized that what was going on in the regres-
 33 sions was one thing, but the discussions that people were having were far more
 34 nuanced than what was in the regressions. Often, I would hear two different sto-
 35 ries. One was that you have bad schooling outcomes if you are in a school with
 36 lots of lower income black kids. The idea was that the fraction of lower income
 37 black kids in the school, or lower ability kids or whatever, influences your out-
 38 come. That's one story, and that's what contextual regressions were appropriate
 39 for.

40 But often people were talking in the discussion about a more subtle phe-
 41 nomenon that I eventually labeled an "endogenous effect," where my outcome
 42 is affected not by who I am around, per se, but by the outcomes of the other
 43 people around me. I thought about this and saw that there is a feedback. If my
 44 outcome is affected by the outcomes of people around me, I also contribute

1 to the aggregate outcomes. Not only am I affected by the outcomes of people
2 around me, but I contribute to the people around me. That is a simultaneity
3 problem.

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

12 Intellectually, the most interesting aspect of this to me to this day is that soci-
13 ologists do not typically think about feedbacks. They pose unidirectional hierar-
14 chical models where A influences B, and B influences C, but with no feedback.
15 The typical sociological model did not have feedbacks that generate an equilib-
16 rium. I decided to write down the simplest model, the linear-in-means model. To
17 start it off, I purposefully made everything linear.

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

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

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

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

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

39 **Chuck:** Yes. Endogenous effects, exogenous effects, and correlated effects.
40 That of course is not specific to the linear model. That's a conceptual thing. I
41 think it's an extremely simple exposition. It's the kind of thing that should have
42 been recognized 30–50 years earlier. But somehow it didn't seem to have been,
43 and particularly not in sociology. It comes back to the fact that sociologists just

1 didn't understand how different things are when you have feedbacks. So that is
2 right. The nomenclature was useful and is cited a lot.

3 Another reason for the citations may be the metaphor of the reflection problem,
4 which I still love. I use it all the time. Whenever I'm trying to explain identifica-
5 tion problems to this day, to explain identification problems from scratch, I use the
6 idea of standing in front of a mirror and not being able to separate whether you are
7 making the mirror image move or the mirror image is making you move. I find it
8 works as a metaphor. I called it the reflection problem. That was good branding, I
9 guess. When the article came out, it didn't immediately take off. There was some
10 recognition of it and I gave seminars, but it wasn't right away a hit. It eventually
11 went viral.

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

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

26 Then I did have another opportunity to think further. I was asked to write an
27 article³⁷ for the *Journal of Economic Perspectives*. That came out in the *JEP* in
28 2000 and is my second most cited article. It is an expository article rather than
29 a technical one. However, it is not a review article. There is a lot of conceptual
30 thinking in it about how different fields do things. Writing the article was helpful
31 to me in getting my thoughts organized in a more coherent way.

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

36 **Elie:** Yes. That's right it is the article that also contains some interesting con-
37 ceptualization of what it means to have network effects in the specific language
38 of the counterfactual notation.

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

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

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

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

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

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

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

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

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

41 **Elie:** Network or social interactions literature has completely exploded. There
42 are empirical articles describing behavior far removed from you in a network

1 but that somehow seems to affect you through some process of diffusion. Some
2 people are excited about it.

3 **Chuck:** If you are thinking about real applications, it wouldn't surprise me
4 that you need to have a fair amount of structure to say something. The *Econo-*
5 *metrics Journal* article,³⁸ which is on treatment response with social interactions,
6 asks how far you can get with a reduced form. With monotonicity and other as-
7 sumptions that may be derived from a structural model, you legitimately have a
8 reduced form. And you can ask what kind of partial identification results you can
9 get. I worked hard on that article and think I got some results. Again, some of
10 what I think I accomplished was nomenclature and setting up notation rather than
11 theorems.

12 **Elie:** That's a nice article.

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

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

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

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

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

24 **Elie:** I agree.

25 **Work on Decision Theory**

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

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

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

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

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

37 The first article was the 1990 *AER* Proceedings article, which gave worst-case
38 bounds and IV bounds on average treatment effects. Often the bounds cover zero,
39 so it could be that treatment A is better than treatment B or that B is better than
40 A. When I would give seminars, people would say: "You get these bounds, but

1 what good are they?” They would argue that to make a decision, one has to make
 2 strong enough assumptions to at least identify the sign of the average treatment
 3 effect. They acknowledged that they might not like the assumptions they make,
 4 but they would argue that there is no reasonable way to make a decision if one
 5 does not know the sign of the treatment effect.

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

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

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

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

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

27 **Elie:** In Belgium or somewhere?

28 **Chuck:** No, this conference was in the United States, an annual conference
 29 on Uncertainty in Artificial Intelligence. I think you have in mind another group,
 30 mainly in Europe, called the Society for Imprecise Probabilities and Their Ap-
 31 plications. I also got involved with that group. However, my first article was not
 32 written for that group. It was for a group mainly of computer scientists who work
 33 on artificial intelligence and who try to develop algorithms that deal sensibly with
 34 uncertainty.

35 No one in econometrics was going to see that article, so the more observ-
 36 able starting point for my work occurred when Art Goldberger retired at Wis-
 37 consin. We had a conference for him. I wrote an article for the conference
 38 that expanded the ideas in the article for the AI conference and explicitly
 39 related partial identification of treatment response to treatment choice under
 40 ambiguity.³⁹

41 It is a simple conceptual article that I value for making the basic observations
 42 that led to all the other work that I have done since then. That is, the identification
 43 region for an average treatment effect generates what a decision theorist would

1 call the state space for the problem of treatment choice. When one's bound on
 2 the average treatment effect covers zero, both A and B are undominated treat-
 3 ment choices. Thus, partial identification generates a problem of decision under
 4 ambiguity. So, the revelation was to see that two topics that initially appear to be
 5 entirely separate, one in econometrics and the other in decision theory, really have
 6 a close relationship.

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

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

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

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

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

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

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

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

41 **Chuck:** Personally, I have been interested in real-world applied decision-
 42 making, but absolutely, standard estimation can be approached as a decision prob-
 43 lem. 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.

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

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

9 Returning to Wald, his view of the world was totally lost. After writing my ar-
10 ticle for the Goldberger conference and the special issue of the *Journal of Econo-*
11 *metrics*, I decided to try to do something serious on the subject. This led to what I
12 view as one of the most important articles I have written, the 2004 *Econometrica*
13 article⁴⁰ on "Statistical Treatment Rules for Heterogeneous Populations." The ar-
14 ticle took me several years to write. It firmly established the conceptual framework
15 I wanted to use, beginning from the general ideas of Wald and then focusing on
16 the minimax-regret criterion. I used large deviations theory to prove some results
17 that are useful in practice. The technical analysis went only part way to what I
18 would have liked to achieve, but Jörg Stoye later pushed the technical analysis
19 much further, publishing two brilliant articles in the *Journal of Econometrics* in
20 2009 and 2012.

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

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

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

33 **Chuck:** No, that's not my interest. You are talking about the asymptotic ver-
34 sion of Wald associated with Hajek and Le Cam. After I wrote my *Econometrica*
35 article on statistical treatment rules, Hirano and Porter wrote their 2009 *Econo-*
36 *metrica* article,⁴¹ which is in the asymptotic tradition. Then there is a new article⁴²
37 of Alex Tetenov and Toru Kitagawa, which is now forthcoming in *Econometrica*.
38 Their analysis is partly finite-sample and partly asymptotic. You are right that
39 the asymptotic minimax concept is a residue of the Wald theory, but it is a fairly
40 esoteric kind of analysis.

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

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

2 **Elie:** Yes, absolutely.

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

5 **Elie:** No.

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

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

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

17 **Elie:** No, it’s not common.

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

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

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

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

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

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

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

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

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

27 **Elie:** [laughs].

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

34 I found that minimax regret works nicely for treatment choice. At the same
 35 time, I was aware of a classical axiomatic argument against it, made by Chernoff
 36 in a 1954 *Econometrica* article. Chernoff noted that minimax regret does not al-
 37 ways satisfy an axiom called independence of irrelevant alternatives. The axiom
 38 says that if I have three alternatives (A, B, and C) and if I choose A over B and
 39 C, then I should still choose A over B if C is eliminated from the choice set. This
 40 seems a sensible requirement for a decision rule, but I later concluded that the
 41 appeal is superficial. I argued this in my 2011 article⁴³ in *Theory and Decision* on
 42 “Actualist Rationality.” Chernoff’s critique of minimax regret was one paragraph
 43 long. He just said: “Look, minimax regret doesn’t obey independence of irrelevant

1 alternatives. Therefore, it is unacceptable.”

2 **Elie:** He has a little more on his criticism in his book with Moise called Ele-
3 ments of Decision Theory.

4 **Chuck:** Is there? I didn’t notice.

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

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

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

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

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

16 **Chuck:** It’s just standard. . .

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

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

23 After publishing that article, I wanted to begin analyzing some cases with
24 identification problems. For example, you may have finite-sample observational
25 data or an experiment with noncompliance. I have an article in the *Journal of*
26 *Econometrics* in 2007 that reports some partial results.

27 I should talk about Jörg’s work. His thesis, leading to his 2009 *Journal of*
28 *Econometrics* article,⁴⁴ was on classical randomized experiments. For a class of
29 interesting cases he nailed what I was not able to show in my 2004 article. That
30 is, he constructively derived the form of the minimax regret decision rule. I had
31 made some progress using large deviation bounds to derive upper bounds on re-
32 gret, which enabled me to partially analyze the regret properties of plugin treat-
33 ment rules, which I called “empirical success rules.” However, I couldn’t solve
34 the full minimax regret problem. I didn’t know how to do it.

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

40 **Elie:** A game.

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

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

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

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

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

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

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

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

19 The second article of Jörg, you should know. It is in the special issue⁴⁶ of
20 the *Journal of Econometrics* based on my 60th birthday conference. I think the
21 second article is even better than the first. It considers data generation processes
22 that have interesting identification problems as well as finite-sample imprecision.
23 He uses the game theoretic device to determine the minimax regret solution when
24 you have both issues.

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

36 The discontinuity result is specific to the treatment choice setting rather than
37 general. I know that now because Jeff Dominicz and I have a new article⁴⁷ in the
38 *Review of Economic Studies* on the problem of estimating the mean of a popula-
39 tion in the presence of missing data. In this case, regret is mean square error and
40 so the minimax regret problem is to find a statistical decision function that min-
41 imizes maximum mean squared error. The article is called "More Data or Better
42 Data?" because we consider a setting in which you can collect data in two ways.

1 You can use an expensive sampling process to obtain a complete random sample
2 without any identification problem, only statistical imprecision. Or you can col-
3 lect data more cheaply but then experience some missing data, which generates
4 an identification problem.

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

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

16 I should mention one further new article,⁴⁸ which I co-wrote with Alex Tetenov
17 and that appeared in *PNAS* in 2016.

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

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

28 Part of the problem with hypothesis testing is this asymmetry. Why should you
29 be willing to accept a twenty percent chance of a Type II error but only a five
30 percent chance of a Type I error? It is only a convention. Another problem with
31 hypothesis testing is that it takes no account of the magnitude of losses. If you
32 make a Type I error, you choose Treatment B when Treatment A is better, so you
33 suffer a certain loss. If you make the other error, there may be a different loss.

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

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

38 **Elie:** It's simple.

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

42 What Alex and I did is to say: Let's look at this choice of sample size problem
43 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 article 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 ε , then regardless of the true state of nature, I am within ε 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 article are sufficient conditions to enable near optimality for any specified ε .

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 articles, but not yet a groundswell of research.

A View Towards the Future

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. Articles 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

1 self-contained field. We are supposed to be useful to empirical research in eco-
2 nomics.

3 I know that I don't need to say this to you because I know you feel the same
4 way. I think we both know that a large fraction of what goes on is not like that.
5 For the future, getting rid of that negative is very important. I think that's the most
6 important task for econometrics in the years ahead.

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

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

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

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

14 I think there's a crisis in econometrics.

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

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

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

29 **Elie:** No.

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

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

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

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

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

38 **Elie:** In a computable way.

39 **Chuck:** It depends what computable way means. Was it Galton who invented
40 the word regression? I think it was Galton in the 1870s. As far as I can see,

1 conditional prediction, regression, and machine learning are all synonyms for the
2 same thing.

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

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

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

14 **Elie:** Same population.

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

17 If you go back to the linear regression literature, an ancient question con-
18 cerns how to proceed when the number of observed covariates is large relative
19 to sample size. Every undergraduate learned about the relevance of "degrees of
20 freedom." There were some poorly motivated suggestions early on, like step-
21 wise regression, but people were aware of the issue. Then there were shrink-
22 age estimators, which were better motivated. And there were variable selection
23 procedures.

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

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

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

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

42 **Chuck:** The assumptions.

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

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

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

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

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

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

15 **Elie:** What about counterfactuals?

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

6 **Elie:** You mean when you add more X's, then Y_1 and Y_0 become less indepen-
7 dent of D?

8 **Chuck:** Yes.

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

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

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

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

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

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

25 **Elie:** Not beyond conditional exogeneity.

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

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

33 **Chuck:** That's fine, in a way. That's a dynamic, that's the view of drug ap-
34 proval. It's sequential learning by experimentation. That's different. It is not in-
35 trinsically connected to big data. It is just sequential experimentation.

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

38 **Chuck:** If you have the ability to do sequential experimentation, that's a
39 fine thing to do. I don't connect that to big data or machine learning. We
40 talked about DeGroot, who wrote a Bayesian book on *Optimal Statistical*

1 *Decisions*. This was 1970, long before the words “big data” and “machine learn-
2 ing” entered.

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

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

9 I know some relevant articles. For the treatment choice problem, there are
10 articles in *Biometrika* and other journals on what are called “adaptive random-
11 ized trials,” which is sequential assignment of people to different treatments
12 where they do Bayesian updating of the treatment allocation. You don’t know
13 much at the beginning, so you do a 50-50 split. As you learn from observ-
14 ing trial outcomes, you change the allocation. These are good formal analy-
15 ses. We may not teach the material in econometrics, but that work is out there.
16 However, it is not machine learning nor big data. It is just sequential Bayesian
17 analysis.

18 I am hopeful that the bubble will burst soon. This is anecdotal, but I have talked
19 to serious noneconometricians who have been subjected to this hype and they ask:
20 “What’s going on?”

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

28 To relate this to partial identification, often a researcher estimating a point-
29 identified model makes some small change to the model and finds that nothing
30 much changes. They say that they have performed “robustness checks.” Have you
31 ever heard anyone say they failed a robustness check? No. In all the seminars,
32 everyone passes their robustness tests. I think that is *prima facie* evidence that
33 they are not pushing hard enough. Anything will break down if you push it hard
34 enough. They don’t push it. It’s the same thing when computer scientists say with-
35 out proof that some machine learning algorithm “works” and don’t tell you when
36 it doesn’t work.

37 **Elie:** You need theory.

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

41 I can relate this to work on partial identification. We know that when assump-
42 tions are weak we don’t get much. People then criticize us and say, “The bounds
43 are too wide.” I always reply: “But it is important to understand the identifying

1 power of assumptions. To do that, you have to know what you get with minimal
2 assumptions and then you see what you get as you add assumptions.” It is an im-
3 portant part of the work to learn when you can’t say much, when identification
4 breaks down.

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

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

12 **Chuck:** It could be.

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

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

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

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

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

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

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

12 NOTES

- 13 1. Manski, C.F. (1975) Maximum score estimation of the stochastic utility model of choice. *Jour-*
 14 *nal of Econometrics* 3(3), 205–228.
- 15 2. McFadden, D. (1973) Conditional logit analysis of qualitative choice behavior. In *Frontiers in*
 16 *Econometrics*, pp. 105–142. Wiley.
- 17 3. Manski, C.F. (1985) Semiparametric analysis of discrete response: Asymptotic properties of the
 18 maximum score estimator. *Journal of Econometrics* 27(3), 313–333.
- 19 4. Manski, C.F. (1988) Identification of binary response models. *Journal of the American Statisti-*
 20 *cal Association* 83(403), 729–738.
- 21 5. Kohn, M.G., C.F. Manski, & D.S. Mundel (1976) An empirical investigation of factors which
 22 influence college-going behavior. In *Annals of Economic and Social Measurement* 5(4), 391–419.
- 23 6. Manski, C.F. (1977) The structure of random utility models. *Theory and Decision* 8(3), 229–254.
- 24 7. Manski, C.F. & S.R. Lerman (1977) The estimation of choice probabilities from choice-based
 25 samples. *Econometrica* 45(8), 205–228.
- 26 8. Powell, J.L. (1984) Least absolute deviations estimation for the censored regression model. *Jour-*
 27 *nal of Econometrics* 25(3), 303–325.
- 28 9. Tamer, E. (2003) Incomplete simultaneous discrete response model with multiple equilibria. *The*
 29 *Review of Economic Studies* 70(1), 147–165.
- 30 10. Manski, C.F. (1989) Anatomy of the selection problem. *Journal of Human Resources*, 343–360.
- 31 11. Manski, C.F. (2010) Unlearning and discovery. *The American Economist* 55(1), 9–18.
- 32 12. Manski, C.F. (2013) *Public Policy in an Uncertain World: Analysis and Decisions*. Harvard
 33 University Press.
- 34 13. Tamer, E. (2010) Partial identification in econometrics. *Annual Review of Economics* 2(1), 167–
 35 195.
- 36 14. Klepper, S. & E.E. Leamer (1984) Consistent sets of estimates for regressions with errors in all
 37 variables. *Econometrica: Journal of the Econometric Society* 1, 163–183.
- 38 15. Cochran, W.G., F. Mosteller, & J.W. Tukey (1953) Statistical problems of the Kinsey report.
 39 *Journal of the American Statistical Association* 48(264), 673–716.
- 40 16. Robins, J.M. (1989) The analysis of randomized and nonrandomized AIDS treatment trials using
 41 a new approach to causal inference in longitudinal studies. *Health Service Research Methodology: A*
 42 *Focus on AIDS* 113, 159.
- 43 17. Manski, C.F. (1990) Nonparametric bounds on treatment effects. *The American Economic Re-*
 44 *view* 80(2), 319–323.
- 45 18. Duncan, O.D. & B. Davis (1953) An alternative to ecological correlation. *American Sociological*
 46 *Review* 18(6), 665–666.
- 47 19. Peterson, A.V. (1976) Bounds for a joint distribution function with fixed sub-distribution func-
- 48 tions: Application to competing risks. *Proceedings of the National Academy of Sciences* 73(1), 11–13.

- 1 20. Manski, C.F. (1997) The mixing problem in programme evaluation. *The Review of Economic*
- 2 *Studies* 64(4), 537–553
- 3 21. Imbens, G.W. & J.D. Angrist (1994) Identification and Estimation of Local Average Treatment
- 4 Effects. *Econometrica* 62(2), 467–475.
- 5 22. Manski, C.F. (1995) *Identification Problems in the Social Sciences*. Harvard University Press.
- 6 23. Manski, C.F. (2003) *Partial Identification of Probability Distributions*. Springer.
- 7 24. Manski, C.F. (2007) Partial identification of counterfactual choice probabilities. *International*
- 8 *Economic Review* 48(4), 1393–1410.
- 9 25. Manski, C.F. (2014) Identification of income–leisure preferences and evaluation of income tax
- 10 policy. *Quantitative Economics* 5(1), 145–174.
- 11 26. Horowitz, J.L. & C.F. Manski (1995) Identification and robustness with contaminated and cor-
- 12 rupted data. *Econometrica: Journal of the Econometric Society* 1, 281–302.
- 13 27. Barseghyan, L., F. Molinari, & J.C. Teitelbaum (2016) Inference under stability of risk prefer-
- 14 ences. *Quantitative Economics* 7(2), 367–409.
- 15 28. Manski, C.F. & J.V. Pepper (2017) How do right-to-carry laws affect crime rates? Coping with
- 16 ambiguity using bounded-variation assumptions. *Review of Economics and Statistics*, forthcoming.
- 17 doi: 10.1162/REST_a.00689.
- 18 29. Goldberger, A.S. & C.F. Manski (1995) Review article: *The bell curve* by Herrnstein and Murray.
- 19 *Journal of Economic Literature* 33(2), 762–776.
- 20 30. Manski, C.F. (1990) The use of intentions data to predict behavior: A best-case analysis. *Journal*
- 21 *of the American Statistical Association* 85(412), 934–940.
- 22 31. Manski, C.F. (1993) Adolescent econometricians: How do youth infer the returns to schooling?
- 23 In C.T. Clotfelter and M. Rothschild (eds.), *Studies of Supply and Demand in Higher Education*,
- 24 pp. 43–60. University of Chicago Press.
- 25 32. Dominitz, J. & C.F. Manski (1996) Eliciting student expectations of the returns to schooling.
- 26 *Journal of Human Resources* 31(1), 1–26.
- 27 33. Dominitz, J. & C.F. Manski (1997) Perceptions of economic insecurity: Evidence from the sur-
- 28 vey of economic expectations. *Public Opinion Quarterly* 92, 261–287.
- 29 34. Dominitz, J. & C.F. Manski (1997) Using expectations data to study subjective income expecta-
- 30 tions. *Journal of the American statistical Association* 92(439), 855–867.
- 31 35. Manski, C.F. (2004) Measuring expectations. *Econometrica* 72(5), 1329–1376.
- 32 36. Manski, C.F. (1993) Identification of endogenous social effects: The reflection problem. *The*
- 33 *Review of Economic Studies* 60(3), 531–542.
- 34 37. Manski, C.F. (2000) Economic analysis of social interactions. *The Journal of Economic Per-*
- 35 *spectives* 14(3), 115–136.
- 36 38. Manski, C.F. (2013) Identification of treatment response with social interactions. *The Econo-*
- 37 *metrics Journal* 16(1).
- 38 39. Manski, C.F. (2000) Identification problems and decisions under ambiguity: Empirical analysis
- 39 of treatment response and normative analysis of treatment choice. *Journal of Econometrics* 95(2),
- 40 415–442.
- 41 40. Manski, C.F. (2004) Statistical treatment rules for heterogeneous populations. *Econometrica*
- 42 72(4), 1221–1246.
- 43 41. Hirano, K. & J.R. Porter (2009) Asymptotics for statistical treatment rules. *Econometrica* 77(5),
- 44 1683–1701.
- 45 42. Kitagawa, T. & A. Tetenov (2017) Who should be treated? Empirical welfare maximization
- 46 methods for treatment choice. *Econometrica*, forthcoming.
- 47 43. Manski, C.F. (2011) Actualist rationality. *Theory and Decision* 71(2), 195–210.
- 48 44. Stoye, J. (2009) Minimax regret treatment choice with finite samples. *Journal of Econometrics*
- 49 151(1), 70–81.
- 50 45. Blackwell, D.A. & M.A. Girshick (1979) *Theory of Games and Statistical Decisions*. Courier
- 51 Corporation.
- 52 46. Stoye, J. (2012) Minimax regret treatment choice with covariates or with limited validity of
- 53 experiments. *Journal of Econometrics* 166(1), 138–156.

- 1 47. Dominitz, J. & C.F. Manski (2017) More data or better data? A statistical decision problem. *The*
- 2 *Review of Economic Studies* 84(4), 1583–1605.
- 3 48. Manski, C.F. & A. Tetenov (2016) Sufficient trial size to inform clinical practice. *Proceedings*
- 4 *of the National Academy of Sciences* 113(38), 10518–10523.