International Statistical Review

International Statistical Review (2020) doi:10.1111/insr.12395

Interview with Professor Adrian FM Smith

Petros Dellaportas[1] and David A. Stephens[2,3]

[1] *University College London, London, UK*

[2] *Athens University of Economics and Business, Greece*

[3] *McGill University, Montreal, QC, Canada*
E-mail: p.dellaportas@ucl.ac.uk

Summary

Adrian Smith joined The Alan Turing Institute as Institute Director and Chief Executive in September 2018. In May 2020, he was confirmed as President Elect of the Royal Society. He is also a member of the government's AI Council, which helps boost AI growth in the UK and promote its adoption and ethical use in businesses and organisations across the country. Professor Smith's previous role was Vice-Chancellor of the University of London where he was in post from 2012. He is a past President of the Royal Statistical Society and was elected a Fellow of the Royal Society in 2001 in recognition of his contribution to statistics. In 2003-04 Professor Smith undertook an inquiry into Post-14 Mathematics Education for the UK Secretary of State for Education and Skills and in 2017, on behalf of Her Majesty's Treasury and the Department for Education, published a 16-18 Maths Review. In 2006 he completed a report for the UK Home Secretary on the issue of public trust in Crime Statistics. He received a knighthood in the 2011 New Year Honours list. The following conversation took place at the Alan Turing Institute in London, on July 19 2019.

Key words: Bayesian Statistics; MCMC; Hierarchical models; Sequential Monte Carlo.

PD: Petros Dellaportas

AS: Adrian FM Smith

DS: David A Stephens

PD: What was academic life like when you were a PhD student and when you did your PhD with Dennis Lindley?

AS: You have to wind the clock back quite a long time to 1968, over 50 years, and the number of people studying PhDs was really tiny compared with today. In London at that time, across LSE (London school of Economics), University college (UCL) and Imperial, there were four of us who knew each other doing PhDs, and we know each other to this day. The scale of investment in things like postgraduate statistics was quite small.

When I did an undergraduate degree in Cambridge, there was virtually no statistics in the maths degree, and that was typical in the 1960s, so the research councils actually funded conversion Masters courses. I was a postgrad student at University College from 1968; I did an MSc, which

was essentially the whole of statistics in one year. Then there was funding for two years for a PhD. Two years! It is hard now to imagine the environment. The leading researchers in London (e.g. Cox, Lindley, Durbin), typically would have had just a handful of PhD students, so it was a very personal kind of relationship. I think Lindley at UCL never had more than three or four PhD students.

One of the consequences of that environment, particularly if you were working in what was seen as a slightly subversive, or controversial area, such as Bayesian statistics, was that you did feel you were part of a small band of people with a mission. But not at all as people would imply in saying 'You guys behave like you're religious' – it wasn't that at all. It was just that there were few of us, and so it was more an issue of sticking together and trying to support each other, because most people would not even listen to what your analysis led to. It was like you were doing this crazy stuff that only Lindley did – so it's obviously worthless to start with.

PD: Tell us more about your career path, before we ask you about statistics.

AS: While I was finishing the PhD at UCL – and I think it's very brave of him in retrospect – John Kingman, who had gone to Oxford as a professor, took a gamble on me. I hadn't even got my PhD. I got the appointment as a lecturer, and I started in October 1971. The RSS paper published with Lindley in 1972 wasn't read at the Society until December 1971. I didn't have a PhD, I hadn't given the paper, so certainly Kingman took one hell of a punt. Part of it may have been he really respected Dennis Lindley, because I think Dennis had taught him probability at some stage.

My first academic job was therefore in Oxford, a world-famous university and brilliant place for mathematics. They wouldn't give Kingman the title 'probability', so he was professor of stochastic analysis. He fought for a statistics post. Brilliant of him to get one and allow them to call it 'statistics' rather than 'applied stochastic analysis' or something else of that sort. My post in the Mathematics Institute was the first in history – after about 700 years of the existence of the university – to have the word 'statistics' in the title. There were just the two of us in probability and statistics, amid masses of group theorists.

At that time, there was a huge cultural bias, even more so in the States I would guess, in university departments – maths departments, stats departments – that it was all really about mathematics. In the US there were exceptions like George Box at Wisconsin, but that was controversial within the university; he was regarded as a subversive for trying to get statistics as an applied science. I grew up in a world where, if you wanted to be taken seriously, you needed to be seen to know a lot of mathematics. Certainly, when I went to Oxford, the undergraduate curriculum had only two courses that exposed people to probability and statistics – they were two quite short courses – plus a course on decision theory.

That was the attraction, maybe, in going back to University College in 1974, because University College was one of the first to have something called a 'statistical science' department. There the culture was the opposite: you weren't taken seriously unless you were doing something applied. Even then, however, you were supposed to do something applied somewhat as a sideline, and what you were supposed to do mainly was to think mathematical thoughts. It has been an interesting shift of culture over time. It's the difference in culture between the *Annals of Statistics*, *Biometrika*etc. and the plethora now of applied journals. I grew up and evolved at the same time as that landscape was evolving. At some stages of my life, if I wanted to have

street credibility, I'd focus on doing mathematics, and at other times, I'd focus on important applications.

What transformed the landscape was computing and computer power. The changing nature of computer technology and power totally transformed things you could do. There were papers in the 1970s which laid out the power of Markov chain Monte Carlo (MCMC) in Bayes computation. They didn't say that explicitly of course, and it wasn't recognised or developed because you couldn't do the computation. Then, suddenly the technology allows you to do the computation and it changes the nature of the beast.

I have been part of, over time, the way the culture and the intellectual basis of the subject became transformed, partly by university expectations and structures, partly by technology, partly by what kind of research gets funded. If you're going to get into serious applications, you need teams of people, you need computers and you need then to get involved in science funding mechanisms and the rest. I think, over those 50 years, the whole landscape has totally transformed.

PD: Your first contribution was a paper of Bayesian hierarchical models. How did this emerge? Was this a topic of your thesis or, when you started your PhD, were you thinking of producing such a modelling framework?

AS: It was my PhD, and the starting point was that Dennis Lindley knew Mel Novick in Princeton who was in the educational testing service. Mel came to UCL on sabbatical and gave tutorials on the big issues in education. You've got pupils, you've got classes, you've got schools, you've got groups of schools in cities, what we now call multi-layer modelling, accounting for different sources of variability at different kinds of levels.

In terms of my involvement, I think one day Lindley said that it would be quite interesting to try and think through all this stuff that Novick was talking about through Bayesian lenses – that became my PhD. Looking back now, the content was very primitive, it's pathetically trivial, but as an idea it turned out to be the first step into the foothills of graphical models, deconstruction of joint probabilities and local conditionals, and the link with MCMC computation. I didn't have the faintest idea of that in 1968. I just wanted to get a PhD.

PD: There was some notion of 'iteration' in this paper. You had a little paragraph saying that we could iterate by maximising rather than simulating.

AS: Yes, I invented a version of conditional iterated modes, but that doesn't work too well in general … so I'm glad I didn't get hooked. However, there are a lot of these things where quick and dirty approximations and shortcuts – which are vulgar, not pretty and not mathematical – could be fantastic tools for accelerating the real thing. I just said to Dennis, 'Hey, I found this interesting thing'. It never was published. It's in a footnote in some monograph of his in 1969. I think it says, 'A. F. M. Smith pointed out to me the following interesting property'.

PD: I think it is mentioned in the 1972 RSS read paper also.

AS: If you're doing a PhD, and if you are in your early 20s, you don't really know anything about the world of Royal Statistical Society discussion meetings. I churned out the algebra and

Dennis one day said, 'I think we should write a paper'. To be honest, I might not have even known, when I started the PhD, that what you did was write papers. I might not have known that.

PD: Who presented the paper in 1972?

AS: I think I did. The discussion was very interesting. One discussant, Oscar Kempthorne, started off with a diatribe against subjective modelling, with 'You gentlemen are trying to destroy the processes of science', and then later on said 'and anyway, we've been doing this for years in my field'. Later, John Kingman, who had given me the job in Oxford, wrote me a little note and it said 'OK not equal to OK'.

How lucky can you get that you stumble into this stuff? If Novick hadn't been on sabbatical, who knows what I would have ended up doing. The next PhD that Dennis supervised was in dynamic programming for drug discovery in the pharma industry.

PD: Was José Bernardo at UCL at the same time as you?

AS: Yes, José was there. He nearly got thrown out after a week. If you had a car, in those days you used to be able to drive in and park in University College. One day, somebody hemmed him in, so he reversed and did a spin turn on the lawn and churned up all the grass. An official complaint was made from the groundsman and he nearly didn't start his PhD because of that! He was there, and so was Tom Leonard at roughly the same time. Tom was doing something similar to what I'd done, but with binary data, related to earlier work of Jack Good.

I'm sure you know that Jack Good was Turing's assistant at Bletchley. If you watch the film *The Imitation Game*, his character appears for a second. He was fantastic, mathematically. When they cracked the Enigma code at Bletchley, they had a terrible dilemma. What do you do with it? Because if you know that a naval convoy is going to be attacked and you disclose it, you disclose that you've cracked the code. What do they best do to pretend they don't know in order to minimise the number of deaths and to maximise the security around knowing the code? Jack was responsible for the relevant probabilistic calculations.

Sadly, all this fantastic knowledge that was generated at Bletchley Park was locked up under the Official Secrets Act for 50 years. After the war, Jack Good couldn't get a chair in Oxford because he couldn't lay claim to what he'd done. That's why he ended up at Virginia Polytechnic. Incidentally, he was very clever there, because when asked what salary he wanted, he said, 'I want one dollar more than the football coach'. Jack Good had more or less invented sequential analysis and probabilistic allocation models. Just imagine that he couldn't claim it or tell anybody about it.

When I went to Oxford in 1971, there was a unit called the Biostatistics Unit, where Maurice Bartlett was head. Then in the Economics Department, there was John Hammersley. Very clever probabilists and statisticians, but there was no unified focus of activity in Oxford at all. Cambridge had something called the Statistical Laboratory. It was actually always run by operations researchers and probabilists, not statisticians, but it provided a focus. I had a job for life in Oxford, but I left after 3 years, because it just wasn't a vibrant environment for my kind of statistics. But I did stumble across some clever people in the meantime, like Michael Goldstein, who was my first PhD student.

PD: Tell us more about the developing attitudes to statistics at that time.

AS: Not many people knew this, but a lot of what drove Dennis in terms of Bayes and axioms and the rest was to try and put statistics on a firm axiomatic basis, like mathematics. Maths was respected, because you take some axioms and you follow them through. This wasn't the case in statistics – it was a culture of 'Pearson said this', 'Neyman said that' and 'Fisher said the other' – a very unsatisfactory intellectual basis.

Lindley was very attracted as a young mathematician by Kolmogorov's book on probability, *Foundations of the Theory of Probability*, published in 1933. It was in German, and Dennis' wife spoke German, so the two of them translated Kolmogorov's probability book, in homage to 'That's the axiomatic way we should approach the subject'. I think Kingman really respected Dennis for his work on probability, because the other thing that Lindley did was fundamental mathematical work on queueing theory and differential equations, which made him a famous probabilist. And this famous probabilist wanted to put statistics on an axiomatic basis.

Dennis worked at the National Physical Laboratory in the war and then, I think, in the late 40s, early 50s, he was in Cambridge. At the same time, in Chicago, L. J. Savage was working on the axiomatics of statistics. David Wallace was chair of the department at Chicago, and he knew Dennis. They were having tea in the Ritz one day and he said, 'You should meet this guy Savage, who's working on . . . ', and this ended up with an invitation to Dennis to spend a sabbatical in Chicago. I think Savage published the book *Foundations of Statistics* in 1954. Seminal thinking can be quite lonely, so I think there was a sense of solidarity in them knowing there were at least two of them in the world working on the same set of things.

Historically, Savage had discovered, and Lindley was totally unaware of, the existence of de Finetti's work in the 1920s and 30s. And they were all unaware of the work of Frank Ramsey, who was the Archbishop of Canterbury's younger brother. Ramsey had done it all in about 1928, as a philosopher in Cambridge. Then suddenly, by the time I was around, enough of this was then known that there was a very influential book, Kyburg and Smokler's monograph *Studies in the Theory of Subjective Probability*, which was 1964. It drew together all this historical material, a tradition developed by people in different spheres completely independently – philosophically looking at the meaning of probability, or whether statistics could be axiomatised. Although the motivation was to put Neyman–Pearson–Fisher on an axiomatic basis, it turned out that the axioms say 'be a Bayesian'.

This was in the air when I was a PhD student. At around that time, Thomas Kuhn wrote a book called *The Structure of Scientific Revolutions*. It was about sometimes, every now and again, you just get a paradigm shift in intellectual thought. So, rather grandiosely, I thought 'Hey, I'm part of a paradigm shift!' and that made it worth putting up with being abused for being Dennis' student. 'Oh God, you're not working with that man!'

DS: You mentioned before about the importance of applied work. When did you feel a switchover from the importance of the axiomatic side?

AS: Criticisms of Bayesian thinking evolved from 'You guys are completely mad; subjectivity destroys objectivity and the processes of science' to 'Okay, well, we have to say there is something about this logical coherence that hangs together, but it's completely useless because you can't do anything with it in practice'.

If I had a thread to my statistical life's work, it was to recognise that ‘They’re right – we've got to address that challenge. We're never going to win the intellectual and practical argument unless we can compute answers'. So that became the challenge.

In addition, people like me got fed up with nobody listening to what we said because it was ‘Bayesian and therefore, you’re obviously mad'. That led to a group of us organising the first Valencia Bayesian conference, to have a space in which we could toss around ideas among people who were interested in where it led, as opposed to telling you that you're wrong-headed. This was around 1978, after the fall of Franco and the opening up of Spain, in which context José Bernardo found a venue on the Spanish coast. I think there were about 80 people who came to that first meeting.

The organisers were myself, José Bernardo, Morris DeGroot and Dennis Lindley. DeGroot was there right from the beginning, because in Carnegie Mellon they always had close links between computer science and statistics and decision-making theory.

There were other strands of related thinking: Raiffa and Schlaefer at Harvard Business School also were prominent in decision theory, but they didn't argue about foundations. They just said the only way to make decisions is to maximise expected utility. The Raiffa and Schlaefer book devised an incredible notation, which distinguished prior, posterior, pre-posterior, predictives by keeping adding commas and dots to the usual algebraic notation. It was mathematically absurd, but the fact that Harvard Business School, whenever it did numbers, did Bayesian numbers was quite influential.

Then there were physicists and computer scientists, who traded under the maximum entropy label, which is more or less equivalent to using a particular kind of minimum information prior. Of course, it can't be true that the whole of mankind should be built on a particular prior that you've arrived at by standing on your head and doing somersaults. However, the fact was there were a lot of problems where that approach wasn't a bad starting point, so they were doing impressive applied work stuff in signal processing and imaging.

Following this first Bayesian Conference, in the late 70s, there was a subsequent conference organised in Cambridge, which had a very revolutionary title. It was called ‘Applied Bayesian Statistics’ – nobody had ever put those words together previously. From a personal perspective, it was at that meeting that I first met colleagues from Ciba-Geigy Pharmaceuticals, with whom I would spend the next several years tackling head on the challenge of making Bayesian statistics routinely applicable to complex applied problems.

One such canonical problem for pharmaceuticals is that you stick drugs into the body and you want to know how much to put in and how often, in order to try to maintain a target level of concentration. This is pharmacokinetics, which involves measuring the concentration in the blood over time. But then, given the concentration, you want to know the effect on the subject – pharmacodynamics. This leads to two different inter-related sets of equations, with perhaps data on two or three thousand people. If you model the curves that define the pharmacokinetics and the pharmacodynamics with, say, four or five parameter models for each individual, you end up with massively parametrised hierarchical non-linear models. A great challenge for Bayes and one of the first applications of Markov chain Monte Carlo.

PD: Which started again with a sabbatical, right?

AS: Alan Gelfand came on sabbatical from Connecticut. I was running the mathematics department at Nottingham, and I was on this committee and that committee. To have somebody around on sabbatical who, from eight o'clock in the morning until midnight, was free to explore ideas we threw up together was incredible. One day we're standing in front of a blackboard, in 1988 I think, when we thought 'Oh wow, we know how to crack this Bayesian computation problem'.

PD: You had suspected that the Geman and Geman paper had something?

AS: Yes, I was groping at it, and there was another one, Tanner and Wong, and of course, my original misguided insight into iterated conditional modes! It was all in the air, and Alan and I were thrashing around with all this stuff. The problem with the Geman and Geman sampler was that it was so specifically related to image reconstruction that I think that if there was an intelligent moment, it was to say, 'Hang on, it's got nothing to do with that, it's to do with the structure of . . . ' Once you've done that, you're there. But I kick myself that I didn't get at the same instant see it all in terms of general graphical models rather than the simple hierarchical stuff we were focussing on.

PD: Well, if it is very simple, it is very hard.

PD: Can you talk to us about your second contribution, which is the sequential Monte Carlo? That was driven by applications as well.

AS: Back in the 1970s, Lindley had funding from the defence agencies, effectively to look at non-linear tracking. If you take conventional models, which have signal and noise, the classical linear Gaussian solution is Kalman filtering, and there's no work to do. You turn the handle and there's a formula. But the world isn't linear and it isn't Gaussian. Many approaches to non-linear tracking were based on approximations to mixtures of Kalman filters. For example, so-called decision-directed learning, where you have a signal but don't know whether it's real or noise, you've only got a probabilistic view. You could toss a coin as to what you think it is and update on that basis. Or you could run it as two or three next steps with a probability on each and then collapse the mixture every so often, but eventually this combinatorial stuff gets completely out of hand. We never satisfactorily got anywhere with these kinds of approaches.

Then, when I was at Imperial in the 1990s, the defence agencies got back in touch saying 'We really do need to crack this'. So we began working on so-called bearings-only tracking, which turned into Neil Gordon's PhD thesis. By that time, of course, I saw everything through the lens of simulation and point clouds generated from distributions rather than the mathematical objects themselves. So the idea was to track the simulated point clouds by means of some further simulation tricks – now known as particle filtering.

PD: You mentioned the book by Kolmogorov, but it seems that UK/USA statisticians did not have a lot of interaction with Eastern Europe at the time, right?

AS: One of the significant things in the world of statistics 50 plus years ago was the start of the programme of European Meetings of Statisticians, which alternated between East and West. There were groups of people reaching across the iron curtain who got to know each other through that and fertilised friendships and links and networks, which subverted the mainstream

politics of the time. Those links were really quite significant; people may now have forgotten what it was like when there was this East Europe/West Europe divide and people couldn't travel. The efforts you had to go to, to get people visas to go to conferences.

PD: When did these collaborations start?

AS: In the early 1960s. I was chair of the science programme committee when it was held in Bulgaria in Varna in 1979. I and others of my generation were deeply involved through the first few decades of these meetings.

DS: Can you talk to us about your current position as Director of the Alan Turing Institute, the UK's data science and AI centre? First, data play a dominant role in most domains of research these days, how does this affect the discipline of statistics?

AS: There's always been data, and people have always played with data to gain insights or make decisions. What's new is the sheer scale of the volume of data and data flows, coupled with incredible increases in computational power. That opens up the opportunity for novel mathematical and algorithmic ways of interrogating data for scientific insight and social and economic good. In a sense, it's still what statisticians always thought they were about but perturbed by the challenge of 'big data'.

PD: In what sense?

AS: Using data to get clever insights and make decisions to make the world a better place remains central. What's changed in it is that, for the most part, whatever the particular statistical approaches adopted have been, there was a huge weight placed on what you might call detailed modelling. The challenge now is: what's the role of modelling and human prior insight, as opposed to just letting the machines loose on the numbers? If Deep Mind can play chess better than any humans just by letting the machine learn from itself and from its own data, where does that leave modelling?

I think that's an interesting intellectual and practical challenge: the trade-offs between learning from enormous training data sets and seemingly mindless algorithms versus the use of human knowledge, prior knowledge, etc. On the other hand, the human inputs have got biases and prejudices. What's the trade-off between embedding them in the learning process and letting clever machines with enough data do their own thing and not necessarily reproducing those biases? This is a fantastically interesting set of issues.

DS: There is of interest, and concern, about the biases that are present in data as well though?

AS: Well, if you are training machines on data collected by humans, in some sense you have to model the human process that collected it and what the biases and the prejudices and the shortcomings are in that. But, turning this on its head again, could you use mathematics and technology to uncover those biases and correct the learning data sets?

In defence and security, depending who you think the good and the bad guys are, what we're all trying to do is subvert the opposition's training data sets by corrupting them. We understand outliers and robustness, so in a scatter plot if I give you a billion concentrated points and one at a distance, a fitted straight line would go through the distant one. Now that's easy to visualise

and understand, but what if we've got 10,000 dimensions and a hundred million data points on each, how would you spot that somebody's manipulated a subset in order to disrupt the data set?

DS: What seems to come up a lot now is a danger in implementing these sorts of methods without being aware of the kind of limitations that you're mentioning, because user-training in areas like robustness is lacking. I suppose it's a two-part thing: where do you see the role of theory, as opposed to heuristic thinking? And what is it that we should be training people in nowadays as statisticians to equip them with the right sort of skills?

AS: A lot of it is about awareness and it's a two-way process. For example, people are developing algorithms which might turn into wearables which monitor health, thereby putting medical devices on the market. But what are the protocols by which you should assess and regulate the process by which they're allowed to come to market? We have, of course, a great statistical history of the development of protocols around clinical trials in drug development. Phase one, phase two, etc., and then monitoring what happens if you put drugs in the public domain. There is both a need and an opportunity to learn from that kind of statistical thinking to approach issues relating to whether and how we might regulate devices based on artificial intelligence. More generally, there is tremendous knowledge and insight which has evolved in statistics, about pitfalls and elephant traps, which the data science and AI community should learn from.

There is probably a need the other way round, in terms of the education of some statisticians. You can't play very effectively in the world of 'big data' by old fashioned techniques of adding things, squaring them and dividing. You're in a totally different algorithmic space. Even if we might claim, using statistical language, that multi-layered neural nets are really just hierarchical non-linear search algorithms, a whole new world of mathematical and computational challenges has opened up.

DS: Do we bother training people in probability now?

AS: Yes, for sure. Some of the most interesting current algorithmic applications boil down to probabilistic prediction. At any given time, if you were the security services in any country, there would be large numbers of people who are subjects of interest. Whatever that number is, it would be many multiples of your ability to actually monitor and track individuals, so you need probabilistic triaging. The same thing applies to at risk kids at school. They turn up unwashed, they haven't had breakfast. If you had enough joined up local data you've got probabilistic predictive potential for individual children at risk. Probabilistically triaging is a universal problem and in the world of 'big data' algorithms, probabilistic search remains core.

DS: Part of the reason I ask is because lots of universities have spent a lot of time putting together data science programmes, and it's hard to do that without basing it on fundamentals – you teach them some statistics, you teach them some computing, but they have to know some mathematics to understand both of those.

AS: I think that's a real and legitimate debate. It was always there implicitly in educating statisticians: how much maths and probability do statisticians need to know? If you're going to be into frontier developments of tuning multi-layered neural nets and back propagations, then you need some pretty hairy maths. How many people need to be in that space? I think it's the same argument as with stats. You can do fantastic good for the world being a clever applied medical statistician, even if you've never proved a mathematical theorem. But we still need a cohort of

primarily mathematically motivated individuals to push and explore the frontiers and validate empirically driven procedures.

Maybe what's perhaps equally important is that people understand the interaction between big data analysis and computer architectures and computer power, as well as understanding probability. They should know the difference in a GPU and a CPU. I don't think for most of my career I ever really knew anything deep about computing even though I was a state-of-the-art FORTRAN programmer in my ‘gap year’ pre-Cambridge.

These are shifting sands. What should be our educational strategy? Should every undergraduate course contain awareness of data science and artificial intelligence?

DS: The other aspect is communication, summarising data, being able to display data intelligently, and communicate the ideas, is fundamental as well. That is something many programmes currently – possibly when I was an undergraduate myself as well – was not really something we learned about. We learned the mechanics of statistics, and the theory of statistics, but not how to communicate statistics.

AS: Let me comment on this through a perspective from the Alan Turing Institute in the UK. Beyond the technological stuff, we recognise the fundamental importance of the social, behavioural, ethical wrap around to the technology. We systematically expose young researchers to the ethical and social issues. One of our core academic programmes is ‘safe and ethical AI’, which explores in depth issues of bias, fairness, explicability, transparency, robustness, reproducibility, etc. It's all very well having commentators say ‘I don’t know how they do it, they just beat you humans'. Maybe you get away with it in chess and Go, but if we are in a clinical decision-making context, you can't say, ‘I don’t know why, but I've got to go and cut him open because the algorithm said so'. Turing has a big programme in that space and we have a public policy programme, which inputs into government and government policy awareness of these issues and the communications challenges.

PD: Don't you think that because statistics is not ‘sexy’ any more, and because industry asks for people with the title ‘data science’ or ‘machine learning’ or ‘artificial intelligence’, the new graduates might not not know basic ingredients of selection bias, truncated samples, etc., that statisticians and economists and psychologists were taught for years?

AS: As per my earlier comment, whoever is convening new courses in the Data Science/AI/Machine Learning area, has much to learn and embed from the corpus of knowledge and insights acquired over the years by the statistics community.

PD: Data science and machine learning degrees are becoming very popular and statistics degrees are becoming less popular, I think.

AS: Do you fight this or join in? Personally, if I were still in a university and the idea of a data science degree was being mooted and I were running a statistics department, I'd put my hand up to take the lead. As an anecdotal aside: there was a point when Denmark was considering joining the EU and they decided not to. There was a football match between Denmark and Germany around the same time as the referendum, and Denmark beat Germany 4–2, so my friend the Danish statistician Steffen Lauritzen said, ‘If you can’t join them, beat them!' Back to your question: as the statistical community, you've got to join in these exciting new challenges.

DS: There is a 2001 paper by Breiman, *Statistical Modeling: The Two Cultures*. This paper is pretty dismissive of 'traditional' statistics and uses the same logic, and makes the case for 'algorithmic modelling'. As a statistician, you cannot sit back and say, 'We are the Principle people'.

AS: Historically, a lot of very sophisticated statistical modelling and tricks inherently only worked on 'small' data sets. The notion of 'small/big' data changes, obviously, but the kind of insight based on small data sets that you can 'see' has limitations. Just imagine the scale of data that's coming in from the SKA (Square Kilometre Array) radio telescopes in Australia and South Africa monitoring the cosmos. It's a challenge even to store the stuff, let alone look at it. I think the scale of data in this kind of big science is intellectually challenging and transformative for the statistical community.

PD: Let's take MCMC, which is, over the last 30 years, the fashionable thing to do in applied statistics. Do you think this will survive the era of big data in the next 20–30 years?

AS: In so far as it is basically clever exploration tricks based on structural knowledge of search spaces, yes some form of it will still be there.

PD: Simulation will stay somehow, right?

AS: I think so. There will now be so-called exoscale computing power that will do things a million times faster than you would have ever thought possible. But there will always be an interplay between computer architectures and the relation of the architectures to the computational challenges of the age. I think statistical computational insight will still be relevant.

And this goes back to the earlier question about 'modelling'. A major boost to AI learning came in 2010, when deep neural nets were clearly getting close to human capability. But in the context of the competitions of the time, if you're still distinguishing tigers and giraffes, you need words and concepts to define tigers and giraffes – and this, of course, is still prior knowledge, so, at least in this context, I think there's an exaggeration if we're predicting the end of modelling. Perhaps the more subtle issue is how much energy do you put into ''modelling', as opposed to just let the algorithm learn for itself.

PD: But you still feed the system, right? It's a Markov decision process with some costs and some policies. You still have a prior there.

AS: Somewhere along the lines, there's a kind of optimisation function. In chess, it's winning. In protein folding, the object of the thing might be that you've come up with a classifier and so if the protein does this your probabilistic prediction is that it probably has this function. In judging whether the algorithm has worked, you implicitly have a model, an outcome model, but you might not have an internal model. We know why we're interested in protein folding, but we have no hypothesis as to how it works, and that is an interesting search problem.

Here is another example. Once upon a time, people got ill and you said what disease have you got and here's the intervention. As you live longer, it's almost impossible to have a single thing going wrong with you, what goes wrong goes wrong as clusters of morbidity and we do not currently have medical techniques, hypotheses, interventions to address this. What if we had an

enormous data set in the UK and let loose a machine learning algorithm? These are new arenas where we currently lack medical science hypotheses.

DS: I've been working a bit recently on one particular health trajectories application. I know how to start modelling that from a statistical or probabilistic perspective, to build in features of things that I might find useful, like competing risks, missing informative. However some of sometimes the' first instinct is 'hit it with a machine learning model'. I would say that it is the role of the statistician to point out all of these things to the people that are already known in statistics that have worked very well in certain places and may work in others.

AS: Of course. At Turing, one of the great things about the Institute is that it's a melting pot across disciplines.

DS: Another part of the training that we didn't mention earlier on is wrangling.

AS: The algorithmic contribution depends on – and is somewhat in-between – the data and the computing. These book ends are both, for mathematically minded types, often seen as unglamorous ends of the spectrum, but pulling the whole thing together is essential. In particular, we need more recognition of and investment in data wrangling and software engineering skills.

PD: Streaming video imaging, text and audio analysis – these are the three technological advances that seem to be driving the need for applications in artificial intelligence

AS: I would add flows within networks like the internet – it's not video, it's not audio, but it's traffic. And as we increasingly put sensors into household appliances, we'll need to come to terms with the 'Internet of Things'.

PD: This is a provocative question: could it be the case that this is all a babble and that's it? All these nice research projects will not link to anywhere.

AS: No, I think it's the other way round. It's now here for keeps, for better and possibly for worse.

PD: Actually, my question was more about AI itself than applications – deep neural nets kind of stuff. The magic of this kind of model, maybe it's a bubble?

AS: Well, there's certainly a mystique around, but it is intriguing. Achievements with Go and chess at a certain level are trivial but it is intriguing. If a similar approach came up with something in protein folding and structure and function, then it would have achieved something that the whole range of medical and mathematical and computing science had failed to do for 40 years. But what if you could now go back into the chess and Go achievements and actually understand the successes. These things are non-transparent. Maybe now you haven't the faintest idea what it did. But what if you put in all sorts of trackers of the movement and the weights through the neural learning process and turned that itself into a data set for exploration. Could you get insight into what the non-transparent algorithm did? Brave new world? This is intellectually challenging. How do you keep track of this stuff? What do you keep track of? What do you throw away? We're operating on huge data sets, but the operation itself generates a huge new data set. Could we monitor and explore that? We're currently in the foothills of where all this might lead.

DS: Which are, in your opinion, the big global public policy problems in which Turing institute can play a vital role?

AS: One is the physical planet and climate change in all its manifestations. I think we've reached a point where instead of just trying to understand implications for sea level or global average temperature, we need to turn our understanding into probabilistic predictions for agriculture, food and water supply etc. moving from scientific insight to policy options and implications. I think now we should be investing into modelling effects and mitigations as opposed to understanding it.

All that in a way goes back to what the engineering community regard as systems thinking. You've got to see it all as inter-related. You can say that forever but it would be useless, because actually to link all these things – unbelievable linkage of data sets – now you can begin to do it.

DS: The things you are mentioning now are in large part, and especially when it comes to mitigation, about policy – government policy, public policy. You mentioned before you had some buy-in.

AS: We've got a public policy programme which is a resource for conversations across all government departments regarding policy issues requiring insights from data, use of data, algorithms. There is great potential for public good, but also issues around whether we need new kinds of laws and bodies to regulate AI.

There are, of course, issues around political considerations versus straight use of data and technical optimisation. Who chooses the objective function? Who chooses the weights in a cost–benefit analysis?

DS: I had climate science and policy in particular in mind because in some parts of the world, in Canada, there is a perceived tension between carbon pricing imposed at federal and provincial level and a desire to grow the economy.

AS: Well, there are timescales. Do you impoverish the current generation in order for life to be better for their grandchildren? Do we vote on this or do we hand it over to technical gurus? None of this is independent of political process and policy, which an institute like Turing shouldn't get into in a partisan fashion, obviously, but we can try to flag where the boundaries are. It's about saying what is the issue, how do you want to frame it and how do we ensure appropriate public awareness and debate

DS: And not everybody thinks the same thing.

AS: No, and it's not just short-term prosperity versus long-term survival. It's often also an issue of prosperity versus security. In all these contexts, what we at the Turing Institute do is we give neutral advice based on our expertise across a wide range of disciplines.

PD: But is it the responsibility of the Turing Institute to protect citizens from cases like Cambridge Analytica?

AS: We obviously don't have the authority to protect but . . .

PD: Advise?

AS: Very much. We're in that space.

PD: The academic system of writing papers and getting reviews and getting publications has changed in the era of artificial intelligence. Now there is this system of conferences, short papers, quick decisions. What do you think about this?

AS: Okay, let's backtrack a bit. Think of Andrew Wiles' proof of Fermat's Last Theorem – maybe it ran to several hundred pages? The conventional model is that it's peer-reviewed and people guarantee that it's all right and put their stamp and say we believe this, it's a proof. However, can this process really hold up for the more complex, deep, abstract mathematical stuff? Who is going to give up a year of their life to plough through these things? And how many people do you need to review it before you have confidence? Who are the people who independently are going to devote their lives to refereeing? Their livelihood depends on pursuing their own research. So I think in many areas there's interesting debate as to whether you just put it out there and the crowd will decide whether it's good, bad or indifferent.

This example is in pure mathematics but many things depend on vast computer program searches, like the original 'proof' of the four-colour theorem. I think this involved checking out a few hundred cases. But what if it had been hundreds of thousands? Can we really expect people to check through millions of lines of code?

PD: What about statistics? These days there are prestigious machine learning/AI conferences that publish beautiful statistics articles much faster than, say, in *Biometrika* or *JRSSB*.

AS: Maybe we're confusing several things. You've got the sociology of universities and careers and promotions and that's in a box over here. In a box over there is real science, implications, effects on the world. Surely getting stuff to have the maximum exposure by people who might maximise the use of it should be the objective. That could be in complete conflict with sticking it in a minor journal that nobody reads.

DS: Statistics used the pre-print model, and certainly maths. Physics works as well on an archive-like basis.

AS: There are perhaps two different worlds. You could argue that the *Annals of Statistics* exists to archive theorems which are eternal truths. But when you publish an applied paper, presumably it's because you think you have insights that will be useful and will change the world. Locating this in a journal on a library shelf seems perverse. The world is 'out there'. I think the whole paradigm may be changing, not just with computer science and AI-type stuff, but also in life sciences and molecular biology, where advances are going faster and faster. Once upon a time, you get a paper accepted and maybe it comes out six months, nine months, a year later. If the speed of advance of the field is monthly, the work becomes redundant or outdated – so you present online or at conferences.
Let's go back to the status of a maths proof. By convention, a number of respected peers read it and said it was all right. That's all it can be, unless you can mechanise it. In mathematical theorems, there's a whole branch of logic which is about applying a computer program to check that the proof is correct. But there are limits on the size of the thing you can do that with. The relevant stuff in AI is called verification. There's a whole piece of computer science which will

be taking an AI machine learning algorithm and seeing if you could verify that it actually did what it claimed to be trying to do. We're back to Alan Turing, in a sense. Anything that is a series of steps, algorithm or otherwise, could or should be subject to another algorithm which checks that it can do what it's doing. It's back to computability. It's quite deep stuff.

But I think the era of peer-reviewed theoretical stuff is under threat, for good or ill. Maybe this poses a problem for universities and promotion structures. You could imagine a world where you have a website and you say, 'our university is thinking of promoting x from this position to this position on the basis of his or her research activity' and then everybody in the world is invited to comment.

PD: It would be interesting to write an AI program that does this.

AS: An academic version of Tinder – likes and dislikes?

PD+DS: Adrian, many thanks for this interview!