

Program Information:

Title: Why Foxes Are Better Forecasters Than Hedgehogs

Location: San Francisco, CA, The Long Now Foundation

Date: Jan 26 2007

Good evening welcome to another seminar about Long Term Thinking sponsored by the Long Now Foundation. My name is Stewart Brand, Philip Tetlock is anticipated something Long Now has wanted to do by about 100 years, we have a site called longbets.org where people register predictions right out in public and people vote on them and some people bet money against them. And it's all very conspicuous and the idea is to develop an idea over 100 years or so. And who is actually good in that predicting, who is quite bad, what kinds of predications hold up and spend spending your time paying attention to what kinds are worthless.

Well Philip Tetlock jumped all that by just going ahead and doing a research on existing predictions by experts and finding out what shifted out from that. And he has got I think one of the fundamental ideas about Long Term Thinking especially as it applies to taking responsibility about what might happen soon in the future, please welcome Philip Tetlock.

Thank you very much Stewart. My topic today is good political judgment and it is a sign of the power of politics to fog our thinking that there is vastly more controversy surrounding the concept of good political judgment than there is surrounding the concepts of say good medical judgment, or good engineering judgment, or good financial judgment. The vast majority of intelligent people feel that there are reasonably solid benchmarks gold standards as it were for assessing good medical judgment. Patients either do well or they do poorly die, but did either they stand or they fall. Your portfolio either grows or it dwindles by contrast good political judgment feels to most people hopelessly subjective. You really don't expect Nancy Pelosi and George W. Bush to agree and you certainly expect in the international scene to observe dramatic differences and how the Iranian leadership looks at good judgment or how Osama Bin Laden views it or Kim Jong-il or Vladimir Putin.

So my first task in this talk is to persuade you that the concept of good political judgment is not hopelessly subjective. That it is possible to construct some reasonably objective benchmarks I call them trans ideological benchmarks, benchmarks that experts from wide range of ideological and theoretical orientations can at least grudgingly accept. And it's possible to learn a lot from applying these benchmarks for the purpose of assessing the judgmental accuracy by political experts.

There are two big categories of benchmarks of good judgments that we use there are knowledge indicators and meta-knowledge indicators. So we are interested both in, whether experts get it right or wrong. And we are also interested in the degree to which experts assign realistic subjective probabilities to their predications. So the degree to which experts know something about what they know and don't know. So do you know what you do, and do not know? And I am going to argue that there is really only way to find out that is to get in a habit of keeping score on one zone judgmental accuracy requires doing something rather cognitively unnatural I think I am speaking now as a psychologist because my disciplinary background is that of a psychologist. Its not an altogether natural mental act to mental process is that way. I'll give a quick example of what is it like to scrutinize ones mental process and I am going to use some non-political examples initially but we will slip quickly into political ones. This is a subjective probability scale for people have dichotomous choice. There are only two possible answers. So a 50% probability represents of course maximum uncertainty coin toss uncertainty, 100% probability represents complete confidence that you are right. And, the research that I am, I started doing in the 1980's and I was then pretty much exclusively an experimental psychology already on confidence and confidence calibration, measuring how realistically people could judge states of knowledge. And these involve constructing rather elaborate test and exercises in which people were asked questions like this about isotopes or the next question France being larger in area than Spain. And we are interested both in the accuracy of your answer and we are interested in the accuracy of your subjective probability assessment. And you might ask, well how would you assess the accuracy or subjective probabilities. You know we, and, I will answer that. There are two fundamental ways to do that; one is called calibration, as the ability to assign subject to probabilities to outcomes that correspond to their objective probabilities. Now to do that, you need to get people to take rather long and elaborate tests. Need to get them to make lot of predictions, one of the thing I was successful in doing was in getting some very thoughtful people to sit down and make a lot of different predictions. Now you're well calibrated if, when you say there is a 70% chance of something happening, the things that you are saying that 70% likelihood to occur about 70% of the time. And you are well calibrated if when you say a 100% probability the thing always happens. When you say 50% it only happens 50%, so there is rough correspondence between subjective and objective probabilities and those of you who have some statistical background will know that this is somewhat noisy measurement process, so you need to make the large numbers work for you, you need to get a lot of predictions to make this, to make the results reasonably reliable and in fact in the research I am going to be talking about, its based, most of the analysis are based on approximately 28,000 predictions over roughly 15 year period.

So calibration is one aspect of the accuracy of subjective probability judgment. Discrimination is the other key aspect I am, going to talk about today. And ask your ability to assign sharply higher probabilities to things that happen and the things that don't happen. And your perfect calibration score is going to look like zero, because there is no gap between subjective and objective probabilities. Your perfect discrimination score is going to look like a 100%. So here is an example of a graph in which on the X-axis you've got subjective probability judgments on the Y-axis you have objective frequencies. And this is an example of someone who never, this is an expert you might characterize as rather cowardly. This is an expert who never offers a subjective probability lower than 0.4 or never offers a subjective probability higher than 0.6. So the best to get out of this character is the minor shades or maybe. Nonetheless this character scores well on calibration because when the individual assigns a probability of 40% things happen about 40% at the time. When the person assigns to probability of 60% things happen 60% at the time. So if your data point is on the diagonal here you are very well calibrated. You are not very discriminating now. This would be an example of good calibration, very good calibration and much better discrimination because now the experts using a much wider range of the probability scale from 0.1 all the way to 0.9. And the next one is what we call the God profile. This is essentially omissions. This is what happens, when you assign a probability of zero things never happen. When you assign a probability of 1, things always happen and since you are omission and perfectly deterministic universe that is, is the way it is. So this is pretty permutable and I don't think I will be surprising anybody when I tell you that no human being looks like that. Okay this is just some technical stuff and probability scoring, which I can come back to if you are interested. I wanted to just say a few words about how I actually came to be doing this. When you do a longitudinal study of forecasting the sorts like that I'm doing, the project starts to feel as long as your entire life. I started doing this back in 1984, so I am 52 almost 53 years old now and I just got in tenure in 1984 Berkley. And I found myself on a committee, the national research council committee created to look at the dynamics of American-Soviet relations. And I don't how many if you remember 1983, early 1984 the Bulletin of the Atomic Scientist had pushed its clock closer to midnight and I think any other, at any other juncture than I think the Cuban Missile Crisis. (KAL07) it has been down and there was a lot of talk about the growing tension in the relation between the United States and the Soviet Union. And there were lot of foundation that we are interested in funneling money into research in this area and this committee existed for that reason. The interesting thing about the situation was that, Jonathan Schell wrote a book "The Fate of the Earth" in which he characterized Ronald Reagan as the demented bus driver queuing down a mountain road with sharp curves. And a lot of people who were liberals thought that Ronald Reagan was indeed bringing us pretty close to the apocalypse. So they are basically, and the lot of people who are conservative thought well Reagan is doing the right thing but the

most we can realistically hope in this situation is to contain the Soviet Union and the Soviet Union will retreat in the Neo-Stalinist mode and that will be that.

Stewart was actually sharing an interesting story about Peter Schwartz and Robert Gates. And I actually had some similar kind of story; I can share with you later.

The dominant view he was a senior Soviet analyst in the mid 1980's. Were the dominant view was that nothing very good was going to happen. There were a couple of people, couple of outliers who thought that dramatic liberalization under Gorbachev like figure could be possible. By it was by and large certainly not the dominant prediction. So the National Research Council is suppose to be a nonpartisan value neutral scientific activity and they are very carefully, they do their diligence, they bring Richards Pipes in and they bring (Marshal Solomon) and the whole political spectrum was represented in that context. And then along comes Gorbachev and it was my first encounter with what I call outcome irrelevant learning situation. This is a situation in which no matter how different the X anti expectation of the experts were, everyone is in the position after the fact to explain what happens. So it really didn't matter in some sense, the liberals could argue, well no Reagan didn't appropriate a nuclear war. You know what Reagan wasted a lot of money in unnecessary defense spending and the Soviet Union would have evolved in this direction anyway. Allow, of course very few people were predicting at X anti and the conservatives of course were eager to claim credit, thought that Ronald Reagan had won the Cold War. Even though they had been on record before hand saying that Gorbachev was really Neo-Stalinist in disguise and was just a (indiscernible) a breathing spell bit of Robert Gates for example has identified with that, that point of view until quite late actually in the Gorbachev period. But it was an outcome irrelevant learning situation and it convinced me that there really would be some great value added if someone tried systematically to keep score on political experts. And that was what I proceeded to do, I just got in tenure, so I didn't have to worry about the gestation period of the project and move forward. Now there is a very interesting similarity I think, I, not only of course you studied the predictions about Soviet Union quite a while to go, I've also monitored some of the more recent predictions about the war in Iraq. And there is some striking similarities in the conceptual structure of the arguments that unfold, I don't think I had been adequately commented on. It's a little bit complicated but it's worth thinking about, because it illustrates I think, it gets to the core answer as to why the concept of good political judgment is so profoundly controversial or as my, like my former colleague (indiscernible), what is it about politics that makes people so dumb? This is a standard 2 by 2 table here and the rows are defined by ideological perspectives the top row represents a view of Reagan or George W. Bush is essentially good, strong, visionary leader. The bottom row represents the view of Reagan or Bush is essentially a dogmatic stabber in somewhat op trap kind and leader. The good historical trajectory represents what happens if things break in the direction that, in the direction of the, the policy of the United States is pursuing

at the moment and the bad historical trajectory is the things break against American policy. So in the top square, in the top left you have direct indications. So insofar the Cold War ended roughly as the, on American terms conservatives quick to claim that most of our policies work. Now, insofar as the Cold War had not ended on favorable terms for the United States if in fact you round up with the kind of situation had round up in Iraq, you would have observe a series of belief system defenses. You would observe them saying; you probably recognize some of these arguments in the context of Iraq debate from Neo-Conservative recognize just half on timing, you have to stay the course, you have to be patient or the downward counter fact defense, you think this is bad you should see what would have happened if we haven't, so that's the downward count of actual, the war, it's a horrible situation we are in but it would been even worse. But we made the right mistake defend, they weren't weapons of mass destruction in Iraq, but it was prudent to set of very low threshold for moving in because that was, because making the other mistake the mistake of not invading when there were weapons of mass destruction would have been far worse.

And Robert Gates of course invoked that argument with respect to Gorbachev in commenting on why he was very slow to recognize that Gorbachev was a genuine liberalizer, he essentially said, well we are making the right mistake. It was prudent to overestimate rather, it's more prudent to over estimate the Soviets than to underestimate. Overestimate their capacity, overestimate their maliciousness. And finally you would take a bad historical trajectory as just further support for, an essential view of the historical adversary.

So what you see is, if the Cold War had ended in an, if the Reagan policies had been associated with nasty outcomes, they probably, they almost certainly would have been invoke counter facts defenses of the forth that you see in the context of Iraq. Conversely you have a belief system dilemma for liberals and many liberals were in awkward position explaining why the Cold War ended despite the fact that they are predicting that Reagan's legacy would be a bad one. So they could invoke the not worth the price defense, they can invoke the other, the upward counter factual. Things could have been just as good or better even if we hadn't spent all this money. And they can invoke their version of that we made the right mistake defense.

So what you have is an interesting kind of intellectual stalemate in which because they are no control groups in history, because no body knows how histories would have unfolded, if it hadn't been President Reagan, if it been a two term Carter presidency with a two term Mondale follow up, who in the audience really knows whether the Cold War would have ended in roughly the same way. It's simply not something available for human empirical knowledge. And for that matter no body can really say for sure what would have happened in the world in which we hadn't gone into Iraq. It's just that the world that we are in seems so bad, that is hard to imagine alternative world. That would be worse, but it's not impossible, certainly not impossible.

So it was just experience with the outcome irrelevant learning situations in the

1980's that lead me to the view that participants across the opinion spectrum are vulnerable to occasional bouts, of audiologically induced insanity and it is crucial, it is a great public need for developing like epistemic or knowledge standards for judging judgement and to having standards for judging judgement that somehow transcend ideological boundaries. From my effort to develop these standards was described in great didactic detail in the book, but the key idea is to work out ground rules for keeping score that advocates of the competing points of view can agree in advance or fair and therefore they find it awkward to denounce those standards after the fact when they are in an awkward position.

So we did this project over a long period of time, 18 year project, 284 experts in international affairs and they were signing prediction, roughly about a 100 each expert on average coming out to about 28,000. Virtually all of them had some postgraduate training, roughly 12 years of work experience and they included a mixture of academics, journalist some intelligence analysts and people in various think tanks. There are certain key requirements for developing good questions; you need to specify the outcome so clearly that they pass the clairvoyance test.

So if Stewart were a true clairvoyant I could hand my predictions to him and Stewart could tell you whether the predictions are right or wrong without the need to comeback to me and say, oh Tetlock what exactly did you mean by a Polish Peron or what did you mean by a back clash in South Africa or what did the kinds of, the types of vague verbiage you often get from experts. Once identified the possible features and in, sufficiently clearly, the pass of clairvoyance test then you get the experts to place subjective probabilities on each set and a lot of our questions have this kind of a format, so central government debt in country X is going to either hold between 35% and 40% of GDP, is going to fall below or rise above that range.

Now this is one of the more far sighted questions we asked, although this question you might know was not all that far sighted because it never really occurred to us that the communist party would collapse. You know the closes we came to anticipating the future here was - its control would weaken. But it's an illustration of what we are getting at. So we asked a lot of questions, 59 different nations, states passed some questions varying on EU, NATO, World Trade Organization lots of different things. Economic performance, growth, inflation, unemployment, policy priorities, defense spending, leadership changes, border conflicts, entry-exit from international agreements. And we had different types of forecasting arises for different variables so we had shorter ranges for the faster moving variables like stock markets and we had slower, we had very long time raises were things that move slowly like border changes for example. So some of our predictions haven't even come do yet, having to deal with border changes.

Okay, among other things we discover this; we discover that contrary to what most political participants would have expected, namely that there a side would win. We find that traditional divisions of political opinion don't matter all that much. It didn't matter all that much in terms of academia we had lots a

different theoretical contrast, and I am sound a little alien we had realist versus institutionalist, constructivist and we had, we had optimist verses pessimist, boomsters versus doomsters, didn't make much of a difference. We had liberals versus conservatives it didn't make all that much of the difference. The content of what people thought was not a very good predictor of how accurate they were. But was a good predictor of how accurate they were was how they thought. At least if you are quiet interested in his aggregate accuracy, what matters is principally how you think and this brings us to the point of our foxes and hedgehogs. And I was actually stunned by the consistency of the result. That expert's who may classified as foxes and I can explain how we classify experts in to these categories in moment.

But experts whom may classified as foxes beat hedgehogs and it was an advantage we could not make a goal it was very hard to make it disappear statistically. So we have tweaked in all sorts of different ways to see if it was not effective chance in some way or not effective different error avoidance priorities all sorts of different control mechanism and the fox advantage really does it does prove to be quiet robust.

But who are these foxes and hedgehogs well there is a essay by Isaiah Berlin came about 45 years ago in which he draws on a fragment of Greek poetry from 2500 years ago by the Greek warrior poet Archilochus which is roughly translated as the fox knows many things but the hedgehog knows one big thing. And he defines the ideal type hedgehog has an expert or professional or a thinker who relates every thing to a single central vision in terms of which all that they say has significance. So you could be a Marxist hedgehog or you could be a libertarian hedgehog. You could be a boomster hedgehog or you could be a doomster malthusian hedgehog. You could be a realist hedgehog; you could be a idealist hedgehog. The important thing is that you approach history; you approach current events in a deductive frame of mind. You have certain first principals and you try as hard as you can to absorb as many different facts in to the frame work of those first principals, that might sound like vagility but if you think about it for a minute from a philosophy of science point of view it also is parsimony that is what scientist are supposed to do they are supposed to explain as much as possible with as little as possible.

So we will come back to that that value attention in a movement. The ideal type hedgehogs, Berlin defined this way, he said they pursue many ends often unrelated and even contradictory they entertain ideas that are centrifugal rather than centripetal with out seeking to fit them in to or exclude them from any one all embracing inner vision, those are the foxes. So foxes and hedgehogs, foxes are skeptical of big theories. You are not going to find very many foxes who are true believers. And interestingly the foxes, who did the best of my forecasting exercises were at least enthusiastic about participating and they were the most diffident about their ability to forecast. Because they really do see history as in substantial measures is quiet unpredictable. Whereas the hedgehogs were more enthusiastic about it they tend to be more enthusiastic about extending their favorite theories in to new domains and intend to be more

confident in their ability to predict.

Now this is some data which I discovered very quickly if you are on the perfect diagonal the straight line here represents perfect calibration so the were the curvy lines represent actual groups of human beings making thousands of predictions and these are aggregations of that. And the key thing to know here is that one of the lines stays further from the perfect diagonal and all the other lines so the line that strays furthest is the line in which hedgehog are making long term predictions within the domains of their expertise. Whereas the line that's closest to the perfect diagonal is foxes making short term predictions within the domains of their expertise. Now there is an argument that started to unfold about whether the foxes were doing better than the hedgehogs not because they are more perceptually accurate but rather because the foxes are, excuse they punt foxes they are chickens the foxes were unwilling to say any thing much more than maybe so the foxes were clinging around the subjective probability point to point five. So one way to test that if that were true then the foxes should not be as discriminating as the hedgehogs. The hedgehogs should - may lose on calibration but they should win on discrimination. So in this the little graph here what we do as we plot both calibration better calibration on the X-axis better discrimination on the Y-axis and as you move towards the upper right you move toward better and better performance and what you see is that the FST, FLT represent foxes making short term and long term predictions and you can see that they are doing better than hedgehogs making shot term and long term predictions HST and HLT and they are better on both calibration and discrimination. So I think that rules out the foxes were just chicken hypothesis. There is however one slightly disturbing result at least from the stand point of those who want to argue that the foxes are particularly brilliant and that is when you compare the fox accuracy of fox predictions to the accuracy of predictions you could have generated by using very simple statistical algorithms like predict no change or predict the most recent rate of change when you use that as your bench mark of comparison that's what we are calling MC mindless competition. And you can see the mindless competition is pretty close to he foxes and so it is not exactly a great indication of human intuitive judgment. Now you might ask what do hedgehog do well and the answer is they assign higher probabilities to big changes that do materialize so if you want to know who is really out front, really out front in predicting some of the bigger changes that have occurred over the last 20 years and there were couple of hedgehogs actually out front with respect to the Soviet Union and the decline of Soviet Union, we call them Ethnic Nationalist Fundamentalists. These were people who argued that the Soviet Union which essentially prisons house of nations it was inherently unstable and it was going to fall apart. Here in mind these people had been predicting the Soviet Union will fall apart back in 1960's, in the 1970"s as well, if like the broken clock is eventually right but they were worried about this, they are right, but they are right at a high price you are paying a high price in false positive predictions. So when you rely on the hedgehogs, the hedgehog predictors, you are going get a lot of false positives.

So the price of being a successfully predicting the disintegration of Yugoslavia or the disintegration of the Soviet Union is that you are also predicting the disintegration of Canada, of India, of Kazakhstan, of Indonesia, of Nigeria, of South Africa it's a long list. And that of course helps to explain what the hedgehogs are doing as well, they are assigning high probabilities to low frequency events which makes much more interesting than the foxes, but it also make them substantially more incorrect.

Now a friend of mine Nassim Taleb who am my, my suspect probably be here at some point, I just finished writing a book on Black Swans in which he talks about the radical unpredictability of the world and, if you want to know the advantage of being the hedgehog its that there its always, there is they usually at least a few of them who are out front in predicting outlandish outcomes. So actual blacks swans at the hedgehogs were in the forefront of expecting, at least a few hedgehogs, most hedgehogs were wrong but the, at least a hedgehogs who were out front. And the list here is collapse of the USSR in 1991 the Yugoslav Civil War, Saddam's invasion of Kuwait Rwandan genocide financial turbulence preceding the collapse of long-term capital management in 1998. The East-Asian currency crisis and Russian debt fault, the internet bubble the 90's this puncturing terrorist strikes all are known predicting 9/11 but terrorist increased incidents of terrorism.

But again lots of false positives, so predictions of the great American and then global depression of 1990's and Dow 36000 and the NASDAQ 5000. We actually had NASDAQ 5000 or Dow 1500 and NASDAQ 350 predicating a Neo-Stalinist school in Moscow that brings return of the Cold War. And of course predicting full bound civil wars in places such as South Africa, Nigeria, Pakistan, India, China, Korean Peninsula, Ukraine and Kazakhstan and even they use of nuclear weapons in the Indian Sub-continent or the Korean Peninsula the two most popular sites for making predictions of that sort. I mean I, I think I run out of time so I should be, how much more time do I have. Okay, I want to leave lots of time for questioning. But I want to say that, I think its tempting for me as a psychologist who as long interested individual differences to treat this as a horse race between foxes and hedgehogs. I think that's fundamentally the wrong way of looking at this. I think it makes more sense to think about hedgehogs and foxes existing in kind of interdependent intellectual ecosystems. So on a one hand you've got the hedgehogs, who are very good at seizing on big ideas and pushing them as far as reasonable and then offering good deal further. On the other hand you have the foxes almost playing a role of intellectual scavengers they are picking bits and pieces of big hedgehog ideas and they are creating a eclectic mismatches, that on average have more predictive power then the original ideas had. So I think this has interesting implications for optimal intellectual diversity in setting up teams. I think it also says; I will give a quick example of how that would work in the contexts of the Soviet Union. There were some people Jerry Hough I am thinking of a Sovietologist for example who were out front in predicting the Gorbachev would be a liberalizer before a lot of other people were. But the

people who are out front in predicting that Gorbachev would be a liberalizer tended to be people who saw the soviet system has having more internal legitimacy then it did. The people who thought that Gorbachev being a liberalizer, it would be impossible to have a liberalized emerge out of the (indiscernible) Patrick's and Richard Pipes of the world who saw the Soviet System has infallibly self reproducing Totalitarian mode, those people thought Gorbachev liberalization be impossible because the Kremlin leadership would recognized as liberalized thing will fall apart. So they wouldn't be stupid enough to do that.

So who would be best position both to predict Gorbachev liberalization and soviet disintegration? It wasn't the liberals who thought the soviet system had more legitimacy than that and it wasn't the conservatives who were ruling out the possibility of liberalization in the first place, it was a small group of foxes who had successfully manage to integrate in their mind some liberal cognitions and some conservative cognitions, they were willing to accept the liberal insight that the soviet leadership like most political leadership in human history was not totally monolithic and they were a genuine differences of opinion within it. So they were willing to accept that, but they also tended to were the more conservative, the soviet system didn't have legitimacy and that once to lid was lifted all hell would break loose.

So that's an interesting way of looking at perhaps the situation in Iraq, you could imagine various types of fox cognitions there. You could imagine fox cognitions that would, try this dissonant possibility for example it would well be that the Bush administration made a mass of mistake, it's certainly is the case that the conservatives we studied who were sympathetic to the invasion of Iraq in the first place would had been wrong on many, many predictions over the last 2 to 3 years. But it is also the case that nobody knows what's going to happen in the next 20 or 30 years. So foxes often have a way of taking out interesting contrarian positions that both liberals and conservative find works in. And a Fox position might be that, might annoy conservatives by saying that the Iraq was stupid and the other hand I might annoy liberals by saying its perfectly possible and in the next 10 or 15 years the situation in Iraq would might be far better and it would have been if Iraq had been under the leadership of Uday Hussein. So foxes have a way of annoying people across the political spectrum they have the way of annoying people who have entrenched theoretical positions. And those are the people who in my exercises tend to do better, they don't get everything right they get lot of things wrong. I think we live in an inherently probabilistic world, nobody is really expecting omissions. But it turns out that a somewhat contrarians self critical cognitive style does translate into more realistic subjective probabilities being assigned the possible futures. And if you think that there is an, that good policy depends on assigning more realistic probabilities the possible futures or if you think that making more money depends on assigning a realistic subject that probabilities possible futures has probably a result we are taking into consideration.

You have enough questions.