Lecture 10: Psychology of probability: predictable irrationality.

David Aldous

October 5, 2017

Here are two extreme views of human rationality.

> *(1) There is much evidence that people are not rational, in the economist's sense [maximization of expected utility (MEU)]. Some would argue we need descriptive economics; I would argue that all should be taught about probability, utility and MEU and act accordingly [Dennis Lindley,* Understanding Uncertainty*.]*
>
> *(2) You mentioned research which revealed that shoppers often prefer "50% extra free" to a notionally more generous 33% reduction in price, and you cited this as evidence of irrationality or poor mathematical ability on the part of consumers. . . . . . . Since all value is subjective, if people value 50% extra free more highly than 33% off, then that is an end of the matter. Whether or not the resulting behaviour conforms to some autistic neoclassical idea of rationality is irrelevant. [Rory Sutherland, Ogilvy & Mather UK. Letter to* The Economist *July 21 2012.]*

The 2011 best-seller *Thinking, Fast and Slow* by Nobel Prize winning Kahneman gives a wide-ranging and very non-technical account of human rationality and irrationality. The key point is that we're not arbitrarily irrational but that our intuition is "predictably irrational" (title of popular 2008 Ariely book) in ways one can describe. The part of this field relevant to STAT 157 concerns "decisions under uncertainty", which necessarily involves issues of probability and utility.

Psychology research gets real data from real people, but the data mostly consists of subjects' answers to hypothetical **limited explicit relevant data** exam-style questions involving uncertainty. My personal view of this field is that we have a good understanding of how people think about such hypothetical questions, but it is less clear how closely this translates to their "real life" behavior, two obvious issues being

- real life does not present us with limited explicit relevant data
- your answer to a "what you would do if ..." question may well not be what you would actually do in real life.

A 2004 book *Cognition and Chance: the psychology of probabilistic reasoning* by Nickerson gives extensive summaries of the research literature and descriptions of experiments (surveys).

Course project: repeat some experiment on your friends.

Later I will describe two such course projects done by students in previous years.

Amongst many survey articles, *Cognitive biases potentially affecting judgment of global risks* (Yudkowsky, 2008) is relevant to a later lecture.

I don't have any new data to serve as "anchor" for today's lecture. Here is a famous example which reveals one important general principle. Text here copied from Wikipedia *Framing (social sciences)* – several later examples also copied from relevant Wikipedia articles.

Two alternative programs have been proposed to combat a new disease liable to kill 600 people. Assume the best estimate of the consequences of the programs are as follows. (information presented differently to two groups of participants in a psychology study).

info presented to group 1: In a group of 600 people,

- Program A: "200 people will be saved"
- Program B: "there is a one-third probability that 600 people will be saved, and a two-thirds probability that no people will be saved"

info presented to group 2: In a group of 600 people,

- Program C: "400 people will die"
- Program D: "there is a one-third probability that nobody will die, and a two-third probability that 600 people will die"

In group 1, 72% of participants preferred program A
In group 2, 22% preferred program C.

The point of the experiment is that programs A and C are identical, as are programs B and D. The change in the decision frame between the two groups of participants produced a preference reversal: when the programs were presented in terms of lives saved, the participants preferred the secure program, A (= C). When the programs were presented in terms of expected deaths, participants chose the gamble D (= B).

This example illustrates a general **framing** principle, observed in many contexts. Our heuristic decisions under uncertainty are strongly affected by whether our attention is focused on the possible benefits/gains or on the possible risks/losses.

The framing issue arises in many "risk" contexts

- medicine – whether to have an operation
- financial investments

Let me quickly mention another well known cognitive bias, called *Anchoring*. To quote Wikipedia, this is

the common human tendency to rely too heavily on the first piece of information offered (the "anchor") when making decisions. Once an anchor is set, other judgments are made by adjusting away from that anchor, and there is a bias toward interpreting other information around the anchor. For example, the initial price offered for a used car sets the standard for the rest of the negotiations, so that prices lower than the initial price seem more reasonable even if they are still higher than what the car is really worth.

As the last sentence implies, anchoring can be used as a negotiating tactic to gain advantage.

This bias is perhaps not so relevant to probability questions, but is loosely related to one of our Lecture 1 survey questions.

Here is data from a Lecture 1 survey question – asked in both the 2014 and 2017 class.

(a) It was estimated that in 2013 there were around 1,400 billionaires in the world. Their combined wealth, as a percentage of all the wealth (excluding government assets) in the world, was estimated as roughly
1.5% 4.5% 13.5% 40.5%

(b) I think the chance my answer to (a) is correct is ....%

	2014 course	
response	number students	average guess P(correct)
1.5%	5	54%
4.5%	3	37%
13.5%	12	36%
40.5%	14	64%

	2017 course	
response	number students	average guess P(correct)
1.5%	4	55%
4.5%	0	
13.5%	9	51 %
40.5%	17	70%

2014 course

response	number students	average guess P(correct)
1.5%	5	54%
4.5%	3	37%
13.5%	12	36%
40.5%	14	64%

2017 course

response	number students	average guess P(correct)
1.5%	4	55%
4.5%	0	
13.5%	9	51 %
40.5%	17	70%

This data is interesting for several reasons.

(1) The figures are from Piketty's *Capital in the Twenty-First Century* who gives the estimate 1.5%. Other sources for 2015 give 2.8%.

(2) One can regard this as an instance of *anchoring*, because I placed the correct answer at one extreme of the possible range of answers.

(3) It is also a dramatic illustration of *overconfidence* in that the people most confident in their opinion were in fact the least accurate.

Wikipedia has a long *List of cognitive biases* [show] and Kahneman's book discusses many of those related to probability and utility. Only a few are mentioned in this lecture.

In the studies above, participants were simply asked questions. But structuring psychology studies as games (rather than as just answering questions) has several advantages, in particular making voluntary participation more appealing. So let me describe two game-structured projects done by students in this course in previous years, repeating on a small scale studies described in the academic literature.

[demonstrate Project 1 with cards]

Project 1. Set-up. From 2 decks of cards assemble one deck with (say) 34 black cards and 17 red cards. Get 50 tokens (or dimes or Monopoly currency notes).

Procedure. Show participant the deck, say it's a non-standard deck with different numbers of black and red cards, but say "I'm not going to tell you anything else – whether there are more black or more red". Say you will shuffle and deal face-up cards. Each time the participant must bet 1 token on the color of the next card – can bet on either red or black.

You do this quickly; at the end ask participant what strategy they were using to decide which way to bet.

Project 1. Set-up. From 2 decks of cards assemble one deck with (say) 34 black cards and 17 red cards. Get 50 tokens (or dimes or Monopoly currency notes).

Procedure. Show participant the deck, say it's a non-standard deck with different numbers of black and red cards, but say "I'm not going to tell you anything else – whether there are more black or more red". Say you will shuffle and deal face-up cards. Each time the participant must bet 1 token on the color of the next card – can bet on either red or black.

You do this quickly; at the end ask participant what strategy they were using to decide which way to bet. A common answer is "after a while I noticed there were more red than black cards – maybe around $2/3$ were black – so I bet on black $2/3$ of the time".

Analysis. At this point the participant may realize that in fact this strategy is not optimal. Once you decide there are more blacks than reds, you should **always** bet on black.

This error is called *Probability matching*. The brief Wikipedia article has the *please improve this article* note, so that's a **project**.

Our second project illustrates a less well known effect.

Project 2. Set-up. Take a bingo game, in which you can draw at random from balls numbered 1 - 75, or similar randomization device. Take 5 tokens.

Procedure. Tell participant you will draw balls one by one; each time, the participant has to bet 1 token on whether the next ball drawn will be a higher or lower number than the last ball drawn. After doing 5 such bets, tell participant "there will be one final bet, but this time you can choose either to bet 1 token, or to bet all your tokens". Finally, ask participant what was their strategy for which way to bet, and how did they decide at the final stage whether to bet all their tokens or just 1.

Results. Before the final bet, almost everyone does the rational strategy – if the last number is less than 37 they bet the next will be larger. What we're interested in is their rationale for *how much* to bet at the last stage. A surprising number of people invoke some notion of *luck* – "I was ahead, so didn't want to press my luck by betting everything at the last stage".

Conclusion. Even when "primed" to think rationally, people often revert to thinking about chance in terms of "luck".

Another common cognitive error is **base rate neglect**, which is the psychologist's phrase for not appreciating Bayes formula. In a famous example, subjects are asked the following hypothetical question.

> *A taxi was involved in a hit and run accident at night. Two taxi companies, the Green and the Blue, operate in the city. 85% of the taxis in the city are Green and 15% are Blue.*
>
> *A witness identified the taxi as Blue. The court tested the reliability of the witness under the same circumstances that existed on the night of the accident and concluded that the witness correctly identified each one of the two colors 80% of the time and failed 20% of the time.*
>
> *What is the probability that the taxi involved in the accident was Blue rather than Green knowing that this witness identified it as Blue?*

Most people answer either 80% or make some guess over 50%. The correct answer, via Bayes formula, is
[board]

Base rate neglect. (Wikipedia). If presented with related base rate information (i.e. generic, general information) and specific information (information only pertaining to a certain case), the mind tends to ignore the former and focus on the latter.

An everyday example, at a wedding you might wonder about the chances this marriage will last 10 years. People's assessment of such things is strongly based on the specifics of the couple involved. But in most examples, you get an improved estimate by moving toward the "base rate" or "population average".

And there's some simple but surprising math here. Suppose n people guess the probabilities of a future event as $p_1, p_2, \ldots, p_n$. The true probability is an unknown value q. What would happen if everyone changed their guess to the average $\bar{p}$?

Intuitively, we can't say, because we have no idea what q is. What does math say?
[board]

Suppose n people guess the probabilities of a future event as $p_1, p_2, \ldots, p_n$. The true probability is an unknown value q. What would happen if everyone changed their guess to the average $\bar{p}$?

Math says: the average MSE would decrease, regardless of the value of q.

This is the "wisdom of crowds" idea, partly underlying the rationale of prediction markets. A more detailed analysis would then consider (over)confidence. **If** people could assess their likely accuracy correctly, then it would best to use a weighted average (more weight on "confident" people). Alas, as we have seen, ...

Base rate neglect. (Wikipedia). If presented with related base rate information (i.e. generic, general information) and specific information (information only pertaining to a certain case), the mind tends to ignore the former and focus on the latter.

A common textbook example concerns false positives in medical tests. Another contemporary example concerns facial recognition technology. Consider the following hypothetical setting.

In a city of 1 million inhabitants let there be 100 terrorists and 999,900 non-terrorists. Thus, the base rate probability of a randomly selected inhabitant of the city being a terrorist is 0.0001, and the base rate probability of that same inhabitant being a non-terrorist is 0.9999. In an attempt to catch the terrorists, the city installs an alarm system with a surveillance camera and automatic facial recognition software.

The false negative rate: *If the camera scans a terrorist, a bell will ring 99% of the time, and it will fail to ring 1% of the time.*
The false positive rate: *If the camera scans a non-terrorist, a bell will not ring 99% of the time, but it will ring 1% of the time.*

Suppose now that an inhabitant triggers the alarm. What is the chance that the person is a terrorist?

Someone making the 'base rate fallacy' would infer that there is a 99% chance that the detected person is a terrorist. In fact (Bayes formula) the chances they are a terrorist are actually near 1%, not near 99%.

The particular effects we have described are replicable in experiments, and we are confident that people do actually encounter such situations and make these kinds or errors in the real world.

But a critique of this style of work is that, by basing theory on hypothetical questions with limited explicit relevant data, one can get "out of touch" with typical real world issues. I will give two "critique" examples. The first involves reverse engineering a famous example. Consider two hypothetical scenarios.

Scenario 1. Police know, from existing evidence, that one of two suspects committed a crime.
Suspect 1 has attributes A and B
Suspect 2 has attribute A (don't know whether B).

(for instance, A = tall, B = wearing dark jacket at the time)

Then find a new witness who is pretty sure the criminal has attributes A and B.

Common sense and Bayes rule agree that, whatever the prior probabilities from other evidence, this extra evidence shifts the probability toward suspect 1.

Scenario 2.
One suspect – police sure this is criminal but are gathering extra evidence for trial. New witness says: pretty sure person has attributes A and B. Which is more likely:
(i) the criminal has attribute A (don't care whether B)
(ii) the criminal has attributes A and B.

In this abstract formulation, and to students who have taken a course in mathematical probability, as a simple matter of logic (i) must be more likely than (ii). But when the question is dressed up in a more colorful story then people often say that (ii) is more likely then (i).

This error is called the *Conjunction fallacy*.
The best-known hypothetical story here is "Linda the feminist bank teller", as follows.

Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Which is more probable?

(i) Linda is a bank teller.
(ii) Linda is a bank teller and is active in the feminist movement.

Most people say (ii) is more likely.

The standard explanation for the error in intuition is discussed in another Wikipedia article *Representativeness heuristic*. That is, the mental picture we get from the story fits (ii) better than (i),

To me, the key point is that the type of comparison we are asked to do in scenario 2 is very artificial – I can think of no "everyday life" setting where we would care about comparing two probabilities like (i) and (ii) for the same entity, so we have no experience of doing so. In contrast, comparing different entities – which menu item will I enjoy more? – is something we do many times a day, so an alternative explanation of this particular heuristic error is that we confuse scenario 2 with the much more familiar scenario 1.

So "Linda the feminist bank teller" works as a memorable example amongst the long list of counterintuitive examples within elementary probability (birthday problem, Monty Hall problem, . . .) but I view it as an artificial game (like Monty Hall) rather than a general principle (like the birthday problem, which spotlights the much more general fact that coincidences happen by pure chance more often than intuition suggests).

A second critique, applicable to textbook exercises as well as these studies in psychology, is that questions based on “limited explicit relevant data” don’t address the complexities of “real life” probability, from which our intuition presumably developed.

Here is a hypothetical, though not so unrealistic, story.

In casual conversation with a stranger in the next airplane seat, you ask "do you have children?" and the response is "yes, two boys". Based on this, what is the chance that the two boys are twins?

This is a fairly realistic conversation. Of course, if you cared about twins you could just ask, so thinking about probabilities here is artificial, but let's do it anyway. We're looking for an approximate answer, and I won't discuss all the common sense approximations being made.
First we need the empirical frequency of a birth giving twin boys, which turns out to be about 5/1000. Second, amongst 2-child families in general, the frequency of "2 boys" must be about 1/4. Now we can estimate the population ratio

$$\frac{\text{(number of families with twin boys, no other children)}}{\text{(number of families with two boys, no other children)}} \approx 2\%.$$

Interpreting ratio of frequencies as a conditional probability,

$$P(\text{twins}|\text{2 boys, no other children}) \approx 2\%$$

and one might give "about 2%" as the answer to the original question.

This answer is drastically wrong! The information we have is not "2 boys, no other children" but instead the information is the specific form of the response "yes, two boys". A person with twins might well have mentioned this fact in their response – e.g. "yes, twin boys". And a person with non-twins might have answered in a way that implied non-twins, e.g. "Sam's in College and Jim's in High School".

The best answer to the original question that I can devise is

$$2\% \times \frac{p}{q}; \quad \text{where}$$

p is the chance a person with twins answers in such a way that you can't infer twins;
q is the chance a person with non-twins answers in such a way that you can't infer non-twins.

Common experience is that people with twins might well actually tell you, so let me guess $p = 1/8$ and $q = 1/2$, leading to my best guess "0.5%" for the desired probability.

The point: for doing probability calculations, and even more so in doing statistical analysis of data based on human responses, it can be misleading to use information without thinking how the information was obtained.

Information mediated by human choice may be true but not statistically reliable.

Here's another example. What information about the world do we learn from today's news headlines?

[show BBC page]

What information about the world do we learn from today's news headlines?

- A few events that happened
- and an almost infinite number of events that didn't happen but **would have been headlines** if they had happened.

Our minds naturally pay more attention to events that happen. But in any kind of Bayesian updating, one should also pay attention to what doesn't happen. This is obvious in the "geopolitical forecasting" context but less obvious in other contexts.

Utility theory is the classical model, referred to in our two opening quotations, for rational agents making decisions. "Utility" is a deliberately vague word, meaning a numerical measure of desirability of outcomes. First consider ordinary non-random settings. Why do we buy things?

Theory says you have a "utility for money" function $U(x)$, generally assumed increasing but concave. You also have a utility U^* for owning any particular item. If you have money x and the item costs y then buying the item will increase your utility if

$$U(x-y)+U^* > U(x).$$

Utility theory says, as a definition or axiom, that the "rational" way to make decisions is to take the action that maximizes utility.

Utilities clearly are subjective, and theory allows them to be arbitrary (different people buy different things), but theory then insists you should behave "rationally" in this particular sense.

In principle you have a utility for everything. For instance time (air travel: direct flight or 2-stop).

Note $U(x)$ not assumed straight line – losing all you money, or doubling your money, are not equal-but-opposite – so being as risk-averse as you wish is counted as "rational".

One might think this theory can explain everything, but it doesn't. For instance, utility theory says you should have the same utility for owning an item, regardless of whether or not you actually own it (price willing to pay to buy should be same as price willing to accept to sell). But in fact there is the

> **endowment effect:** *people place a higher value on objects they own than objects they do not own.*

To demonstrate this effect in the real world one wants a situation where one can eliminate possible extraneous reasons. Here is a very ingenious example.

Duke University has a very small basketball stadium and the number of available tickets is much smaller than the number of people who want them, so the university has developed a complicated selection process for these tickets that is now a tradition. Roughly one week before a game, fans begin pitching tents in the grass in front of the stadium. At random intervals a university official sounds an air-horn which requires that the fans check in with the basketball authority. Anyone who doesn't check in within five minutes is cut from the waiting list. At certain more important games, even those who remain on the list until the bitter end aren't guaranteed a ticket, only an entry in a raffle in which they may or may not receive a ticket. After a final four game in 1994, economists Ziv Carmon and Dan Ariely called all the students on the list who had been in the raffle. Posing as ticket scalpers, they probed those who had not won a ticket for the highest amount they would pay to buy one and received an average answer of $170. When they probed the students who had won a ticket for the lowest amount they would sell, they received an average of about $2,400. This showed that students who had won the tickets placed a value on the same tickets roughly fourteen times as high as those who had not won the tickets.

[show page 1 again]

So we have a choice of what to conclude from this kind of observed behavior.

- People are irrational.
- Rationality is irrelevant – what matters is psychological satisfaction.

Let's anyway continue with utility theory as a prescription for how people **should** behave. Consider the context of "decisions under uncertainty", that is random outcomes.

If we assign utilities to all possible outcomes, and if we know how our available choices affect the probabilities of the various outcomes, then we can make the choice that maximizes expected utility (MEU; note Wikipedia discusses this as *Expected utility hypothesis*).

First, a few uncontroversial comments. If your wealth is \$10K and somehow have an opportunity to bet (win or lose) \$5K with a 51% chance of winning, most people would not take that bet. But if the chance of winning were 99% then most people *would* take that bet. MEU theory explains this by saying that you would take the bet if the probability p of winning satisfies

$$pU(15K) + (1-p)U(5K) > U(10K).$$

So according to MEU theory

being risk-averse for large amounts is rational
being risk-averse for small amounts in not rational.

So, for instance,

buying insurance to offset possible large losses is rational
buying insurance to offset possible small losses is not rational