Probability and Random Processes

ECS 315

Asst. Prof. Dr. Prapun Suksompong

prapun@siit.tu.ac.th

II. Events-Based Probability Theory

Office Hours:

Check Google Calendar on the course website.

Dr.Prapun's Office:

6th floor of Sirindhralai building, BKD

Probability and Random Processes

ECS 315

Asst. Prof. Dr. Prapun Suksompong

prapun@siit.tu.ac.th

5 Foundation of Probability Theory

Office Hours:

Check Google Calendar on the course website.

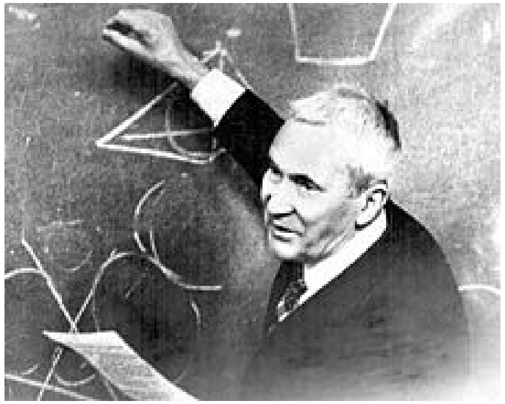
Dr.Prapun's Office:

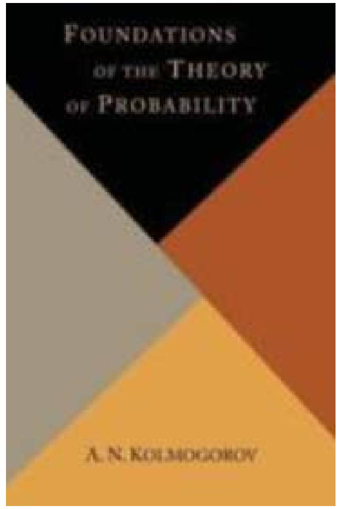
6th floor of Sirindhralai building, BKD

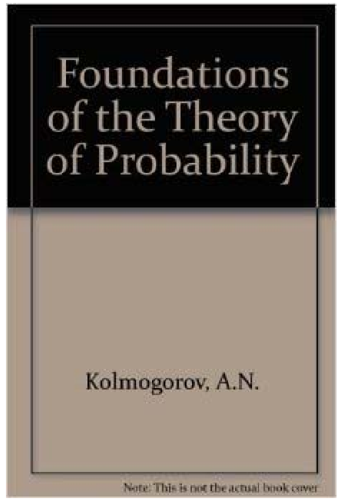
Kolmogorov

- Andrey Nikolaevich Kolmogorov
- Soviet Russian mathematician
- Advanced various scientific fields
  - **probability theory**
  - topology
  - classical mechanics
  - computational complexity.
- 1922: Constructed a **Fourier series** that diverges almost everywhere, gaining international recognition.
- **1933**: Published the book, **Foundations of the Theory of Probability**, laying the modern axiomatic foundations of probability theory and establishing his reputation as the world's leading living expert in this field.

This book is available at [https://archive.org/details/foundationsofthe00kolm]







I learned probability theory from

Eugene Dynkin

Philip Protter

Gennady Samorodnitsky

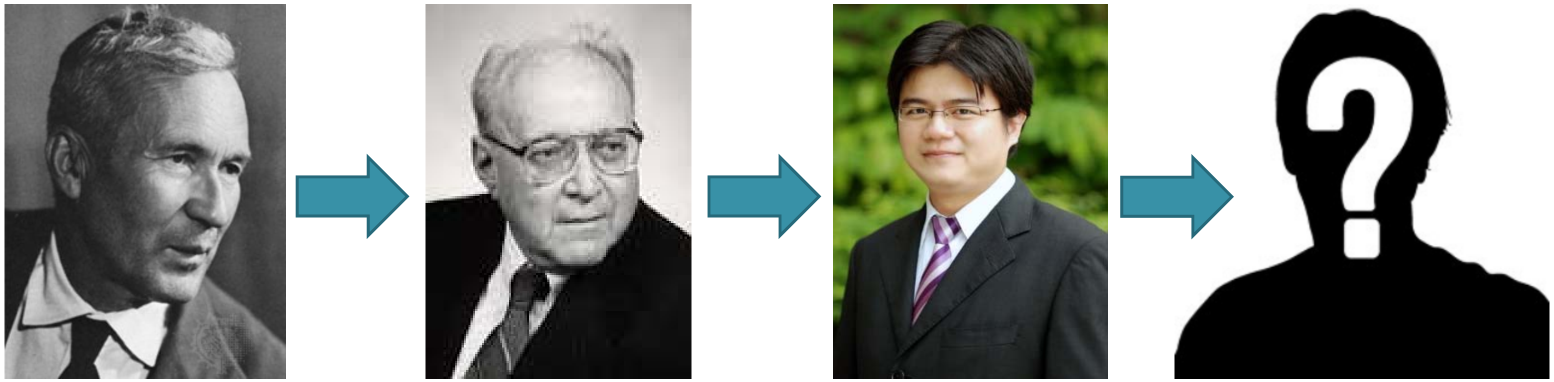
Terrence Fine

Xing Guo

Toby Berger

Rick Durrett

Not too far from Kolmogorov



You can be

the 4th-generation

probability theorists

Probability and Random Processes

ECS 315

Asst. Prof. Dr. Prapun Suksompong

prapun@siit.tu.ac.th

Event-Based Properties

Probability space

- Mathematically, to talk about probability, we refer to **probability space**.
- Probability space has three components
  1. Sample space Ω
  2. Collection of events
     - Example: All subsets of Ω . (Assume Ω is finite.)
  3. Probability Measure
     - A real-valued set function

[Definition 5.1]

Kolmogorov's Axioms for Probability

Abstractly, a **probability measure** is a function that assigns real numbers to events, which satisfies the following assumptions:

P1 Nonnegativity: For any event A,

$$P(A) \geq 0.$$

This is called the probability of the event A.

P2 Unit normalization:

$$P(\Omega) = 1$$

P3 Countable Additivity: If $A_1, A_2, ..$, is a (countably-infinite) sequence of **disjoint** events, then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

Additivity

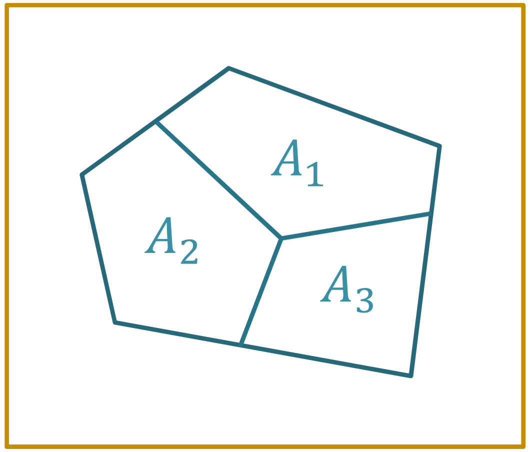
- **Assumption**: $A_1, A_2, \ldots$ are **disjoint** events.

[5.1 P3]
- Countable Additivity: $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

[5.4]
- Finite Additivity: $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$
  - The formula is quite intuitive when you visualize these events in a Venn diagram and think of their **probabilities** as **areas**.
  - Example:

The "area" of the sample space is 1.

[5.1 P2]



$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3)$

[5.6] Steps to find probability of an event that is defined by outcomes

1. Identify the sample space Ω and the probability $P(\{\omega\})$ for each outcome ω.
2. Identify all the outcomes inside the event under consideration.
3. When the event is countable,
   its probability can be found by adding the probability $P(\{\omega\})$ of the outcomes from the previous step.

$$P(\{a_1, a_2, \ldots, a_n\}) = \sum_{i=1}^{n} P(\{a_i\})$$

$$P(\{a_1, a_2, \ldots\}) = \sum_{i=1}^{\infty} P(\{a_i\})$$

See [Example 5.7].

Steps to Find Event-Based Probability

Step 1 Let n be the number of events' names used in the question.

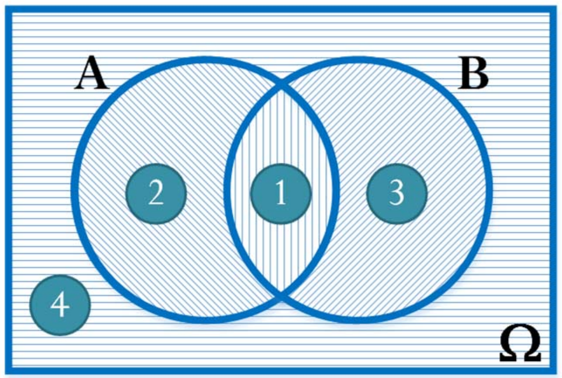
- For example, if the question only talks about A and B, then $n = 2$.

Step 2 Partition the sample space (Ω) into 2^n parts where each part is an intersection of the events or their complements.

- For example, when we have two events, the sample space can be partitioned into 4 parts:

  ❶ $A \cap B$,
  ❷ $A \cap B^c$,
  ❸ $A^c \cap B$, and
  ❹ $A^c \cap B^c$

  as shown in the Venn diagram.



Step 3 Let p_i be the probability of the i^{th} part.

Steps to Find Event-Based Probability

Step 4 Turn the given information into equation(s) of the p_i.

- For example, if you are given that $P(A \cup B) = 0.3$, we see that $A \cup B$ cover parts ❶, ❷, and ❸. Therefore, by finite additivity, the corresponding equation is $p_1 + p_2 + p_3 = 0.3$.
- It is easier to work with expression involving intersection than the one with union.
  - Use de Morgan law [2.5] and complement rule [5.15]
  - For example, suppose we are given that $P(A \cup B^c) = 0.3$.
    - By the complement rule, $P\left((A \cup B^c)^c\right) = 1 - 0.3 = 0.7$.
    - By de Morgan law, $(A \cup B^c)^c = A^c \cap B$.
    - Therefore, the provided information is equivalent to $P(A^c \cap B) = 0.7$.
    - The corresponding equation is $p_3 = 0.7$.
- Don't forget that we always have an extra piece of information: $P(\Omega) = 1$.
  - With two events, this means $p_1 + p_2 + p_3 + p_4 = 1$.

Steps to Find Event-Based Probability

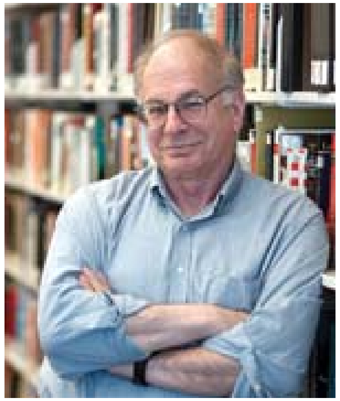
Step 5 Solve for the values of the p_i.

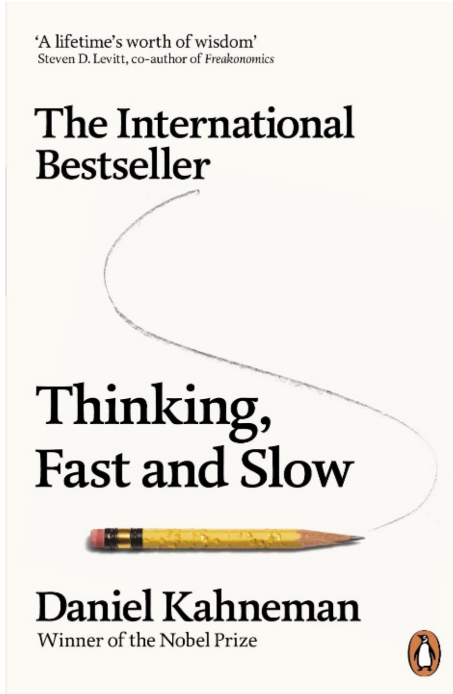
- Note that there are n unknowns; so we will need n equations to solve for the values of the p_i.
- If we don't have enough equations, you may be overlooking some given piece(s) of information or it is possible that you don't need to know the values of all the p_i to find the final answer(s).

Step 6 The probability of any event can be found by adding the probabilities of the corresponding parts.

Daniel Kahneman

- Daniel Kahneman
- Israeli-American **psychologist**
- 2002 **Nobel** laureate
  - In **Economics**
- Hebrew University, Jerusalem, Israel.
- Professor emeritus of psychology and public affairs at **Princeton** University's Woodrow Wilson School.
- With Amos **Tversky**, Kahneman studied and clarified the kinds of misperceptions of randomness that fuel many of the common fallacies.





K&T: Q1

Tversky, A., & Kahneman, D. (1982). Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under Uncertainty: Heuristics and Biases* (pp. 84-98). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511809477.007

6. Judgments of and by representativeness

Amos Tversky and Daniel Kahneman

Several years ago, we presented an analysis of judgment under uncertainty that related subjective probabilities and intuitive predictions to expectations and impressions about representativeness. Two distinct hypotheses incorporated this concept: (i) people expect samples to be highly similar to their parent population and also to represent the randomness of the sampling process (Tversky & Kahneman, 1971, **2**; 1974, **1**); (ii) people often rely on representativeness as a heuristic for judgment and prediction (Kahneman & Tversky, 1972b, **3**; 1973, **4**).

The first hypothesis was advanced to explain the common belief that chance processes are self-correcting, the exaggerated faith in the stability of results observed in small samples, the gambler's fallacy, and related biases in judgments of randomness. We proposed that the lay conception of chance incorporates a belief in the law of small numbers, according to which even small samples are highly representative of their parent populations (Tversky & Kahneman, 1971, **2**). A similar hypothesis could also explain the common tendency to exaggerate the consistency and the predictive value of personality traits (Mischel, 1979) and to overestimate the correlations between similar variables (see Chap. 15) and behaviors (Shweder & D'Andrade, 1980). People appear to believe in a hologram-like model of personality in which any fragment of behavior represents the actor's true character (Kahneman & Tversky, 1973, **4**).

The hypothesis that people expect samples to be highly representative of their parent populations is conceptually independent of the second hypothesis, that people often use the representativeness heuristic to make predictions and judge probabilities. That is, people often evaluate the probability of an uncertain event or a sample "by the degree to which it is

This work was supported by the Office of Naval Research under Contract N00014-79-C-0077 to Stanford University.

Two brief personality sketches were constructed. Each participant encountered one of these sketches in the within-subjects treatment and the other in a between-subjects treatment. In the former, the personality sketch was followed by eight possible outcomes, including a representative outcome, an unrepresentative outcome, and the conjunction of the two. In the between-subjects treatment the list of outcomes included either the two critical single outcomes or their conjunction. The within-subjects forms of the two problems are shown here. The numbers in parentheses are the mean ranks assigned to the various outcomes by the subjects who received this form.

Bill is 34 years old. He is intelligent, but unimaginative, compulsive, and generally lifeless. In school, he was strong in mathematics but weak in social studies and humanities.

Please rank order the following statements by their probability, using 1 for the *most* probable and 8 for the least probable.

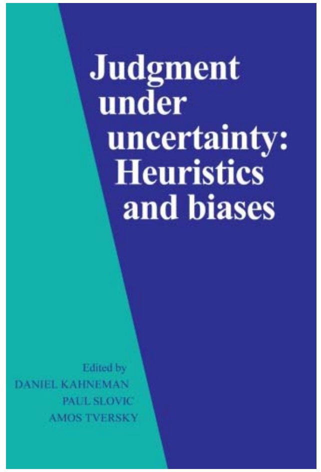
(4.1) Bill is a physician who plays poker for a hobby.
(4.8) Bill is an architect.
(1.1) Bill is an accountant. (*A*)
(6.2) Bill plays jazz for a hobby. (*J*)
(5.7) Bill surfs for a hobby.
(5.3) Bill is a reporter.
(3.6) Bill is an accountant who plays jazz for a hobby. (*A* & *J*)
(5.4) Bill climbs mountains for a hobby.

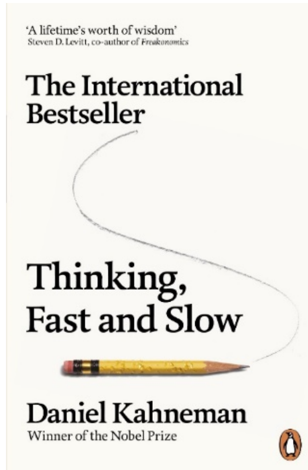
Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Please rank the following statements by their probability, using 1 for the most probable and 8 for the least probable.

(5.2) Linda is a teacher in elementary school.
(3.3) Linda works in a bookstore and takes Yoga classes.
(2.1) Linda is active in the feminist movement. (*F*)
(3.1) Linda is a psychiatric social worker.
(5.4) Linda is a member of the League of Women Voters.
(6.2) Linda is a bank teller. (*T*)
(6.4) Linda is an insurance salesperson.
(4.1) Linda is a bank teller and is active in the feminist movement. (*T* & *F*)

As the reader has probably guessed, the description of Bill was constructed to be representative of an accountant (*A*) and unrepresentative of a person who plays jazz for a hobby (*J*). The description of Linda was constructed to be representative of an active feminist (*F*) and unrepresentative of a bank teller (*T*). In accord with psychological principles of similarity (Tversky, 1977) we expected that the compound targets, an accountant who plays jazz for a hobby (*A* & *J*) and a bank teller who is active in the feminist movement (*T* & *F*), would fall between the respective simple targets. To test this prediction, we asked a group of 88





K&T: Q1

Imagine a **woman** named **Linda**, **31** years old, [outspoken = given to expressing yourself freely or insistently] **single**, **outspoken**, and very **bright**. In college she majored in **philosophy**. While a student she was deeply concerned with **discrimination** and **social justice** and participated in **antinuclear** [protest] **demonstrations**.

- K&T presented this description to a group of 88 subjects and asked them to
  rank the eight statements (shown on the next slide) on a scale of 1 to 8 according to their probability, with
  - **1** representing the **most probable** and
  - **8** representing the **least probable**.

[Mlodinow, The Drunkard's Walk: How Randomness Rules Our Lives, 2008, p. 22-26]
[Kahneman, Slovic, and Tversky, Judgment under Uncertainty: Heuristics and Biases, 1982 p. 90–98.]
[https://youarenotsosmart.com/2016/06/16/yanss-077-the-conjunction-fallacy/]

K&T: Q1… Remarks

- The audiences who heard this description in the 1980s always laughed because they immediately knew that Linda had attended the University of California at Berkeley, which was famous at the time for its radical, politically engaged students.
- The League of Women Voters is no longer as prominent as it was, and the idea of a feminist "movement" sounds quaint, a testimonial to the change in the status of women over the last thirty years.
- Even in the Facebook era, however, it is still easy to guess the almost perfect consensus of judgments: Linda is a very good fit for an active feminist, a fairly good fit for someone who works in a bookstore and takes yoga classes—and a very poor fit for a bank teller or an insurance salesperson.

[feminist = of or relating to or advocating equal rights for women]

K&T: Q1 - Results

- Here are the results - from **most to least probable**

	Statement	*Average Probability Rank*
Most probable	Linda is active in the feminist movement.	2.1
	Linda is a psychiatric social worker.	3.1
	Linda works in a bookstore and takes yoga classes.	3.3
	Linda is a bank teller and is active in the feminist movement.	4.1
	Linda is a teacher in an elementary school.	5.2
	Linda is a member of the League of Women Voters.	5.4
	Linda is a bank teller.	6.2
Least probable	Linda is an insurance salesperson.	6.4

K&T: Q1 – Results (2)

- At first glance there may appear to be nothing unusual in these results: the description was in fact designed to be
  - representative of an active feminist and
  - unrepresentative of a bank teller or an insurance salesperson.

Most probable ↓ Least likely

Statement	*Average Probability Rank*	
Linda is active in the feminist movement.	2.1	←
Linda is a psychiatric social worker.	3.1	
Linda works in a bookstore and takes yoga classes.	3.3	
Linda is a bank teller and is active in the feminist movement.	4.1	←
Linda is a teacher in an elementary school.	5.2	
Linda is a member of the League of Women Voters.	5.4	
Linda is a bank teller.	6.2	←
Linda is an insurance salesperson.	6.4	

K&T: Q1 – Results (3)

- Let's focus on just three of the possibilities and their average ranks.
- This is the order in which **85 percent** of the respondents ranked the three possibilities:

	Statement	*Average Probability Rank*
More likely	Linda is active in the feminist movement.	2.1
	Linda is a bank teller and is active in the feminist movement.	4.1
Less likely	Linda is a bank teller.	6.2

- If nothing about this looks strange, then K&T have fooled you

K&T: Q1 - Contradiction

The probability that two events will both occur can never be greater than the probability that each will occur individually!

Statement	*Average Probability Rank*
Linda is active in the feminist movement.	2.1
Linda is a bank teller and is active in the feminist movement.	4.1
Linda is a bank teller.	6.2

K&T: Q2

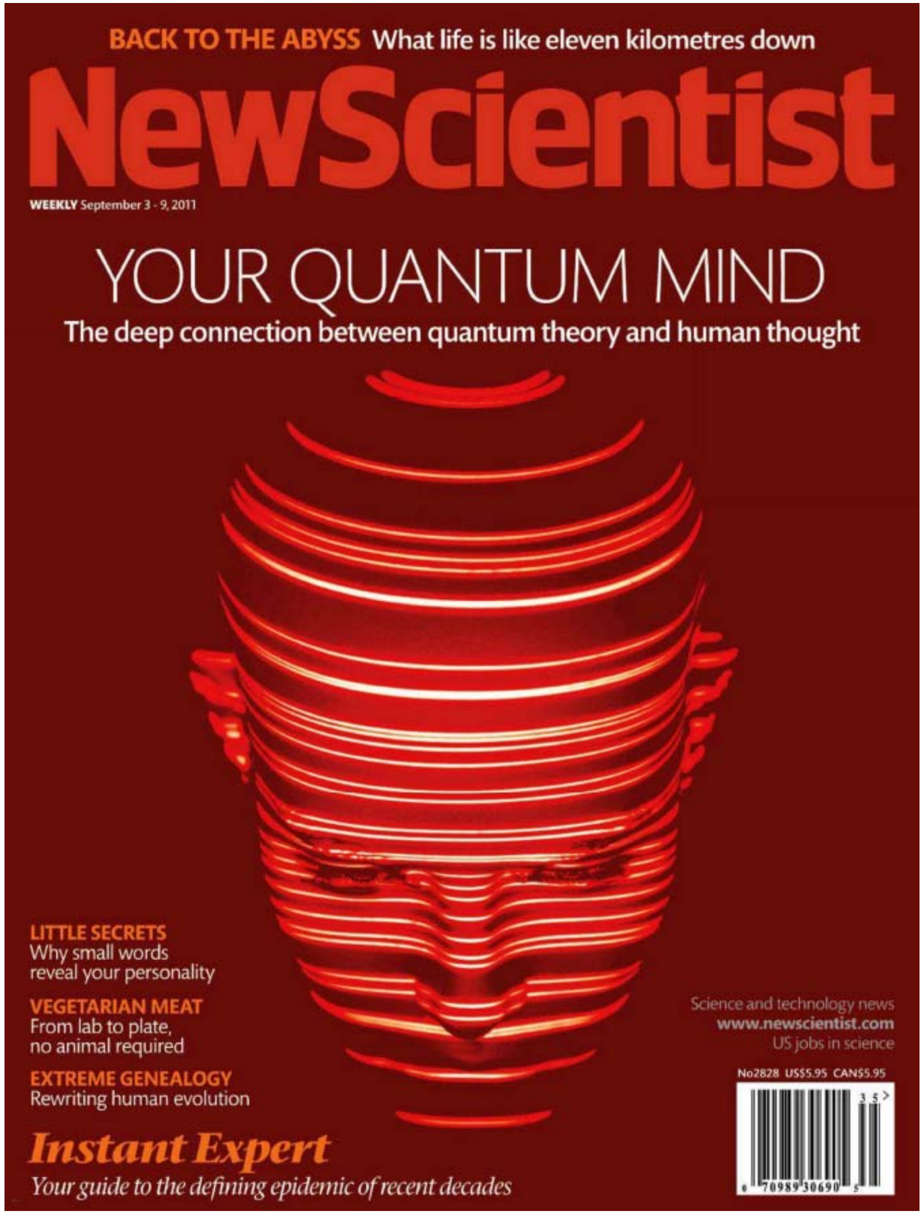
- K&T were not surprised by the result because they had given their subjects a **large** (eight) **number of** possibilities, and the connections among the three scenarios could easily have gotten lost in the shuffle.
- So they presented the description of Linda to another group, but this time they presented **only three possibilities**:
  - Linda is active in the feminist movement.
  - Linda is a bank teller and is active in the feminist movement.
  - Linda is a bank teller.
- Is it now obvious that the middle one is the least likely?

K&T: Q2 - Results

- To their surprise, **87 percent** of the subjects in this trial also **incorrectly** ranked the probability that "Linda is a bank teller and is active in the feminist movement" higher than the probability that "Linda is a bank teller".
- If the **details** we are given **fit our mental picture** of something, then the more details in a scenario, the more real it seems and hence the **more probable** we consider it to be
  - even though any act of adding less-than-certain details to a conjecture makes the conjecture less probable.
- Even **highly trained doctors** make this error when analyzing symptoms.
  - 91 percent of the doctors fall prey to the same bias.

[Amos Tversky and Daniel Kahneman, "Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," Psychological Review 90, no. 4 (October 1983): 293–315.]

Related Topic



- Page 34-37
- Tversky and Shafir @ Princeton University

Quantum minds

The fuzziness and weird logic of the way particles behave applies surprisingly well to how humans think. Mark Buchanan finds the "you" in quantum

THE quantum world defies the rules of ordinary logic. Particles routinely occupy two or more places at the same time and don't even have well-defined properties until they are measured. It's all strange, yet true – quantum theory is the most accurate scientific theory ever tested and its mathematics is perfectly suited to the weirdness of the atomic world.

Yet that mathematics actually stands on its own, quite independent of the theory. Indeed, much of it was invented well before quantum theory even existed, notably by German mathematician David Hilbert. Now, it's beginning to look as if it might apply to a lot more than just quantum physics, and quite possibly even to the way people think.

Human thinking, as many of us know, often fails to respect the principles of classical logic. We make systematic errors when reasoning with probabilities, for example. Physicist Diederik Aerts of the Free University of Brussels, Belgium, has shown that these errors actually make sense within a wider logic based on quantum mathematics. The same logic also seems to fit naturally with how people link concepts together, often on the basis of loose associations and blurred boundaries. That means search algorithms based on quantum logic could uncover meanings in masses of text more efficiently than classical algorithms.

It may sound preposterous to imagine that the mathematics of quantum theory has something to say about the nature of human thinking. This is not to say there is anything quantum going on in the brain, only that "quantum" mathematics really isn't owned by physics at all, and turns out to be better than classical mathematics in capturing the fuzzy and flexible ways that humans use ideas. than the one dictated by classical logic," says Aerts. "The mathematics of quantum theory turns out to describe this quite well."

It's a finding that has kicked off a burgeoning field known as "quantum interaction", which explores how quantum theory can be useful in areas having nothing to do with physics, ranging from human language and cognition to biology and economics. And it's already drawing researchers to major conferences.

One thing that distinguishes quantum from classical physics is how probabilities work. Suppose, for example, that you spray some particles towards a screen with two slits in it, and study the results on the wall behind (see diagram, page 36). Close slit B, and particles going through A will make a pattern behind it. Close A instead, and a similar pattern will form behind slit B. Keep both A and B open and the pattern you should get – ordinary physics and logic would suggest – should be the sum of these two component patterns.

But the quantum world doesn't obey. When electrons or photons in a beam pass through the two slits, they act as waves and produce an interference pattern on the wall. The pattern with A and B open just isn't the sum of the two patterns with either A or B open alone, but something entirely different – one that varies as light and dark stripes.

Such interference effects lie at the heart of many quantum phenomena, and find a natural description in Hilbert's mathematics. But the phenomenon may go well beyond physics, and one example of this is the violation of what logicians call the "sure thing" principle. This is the idea that if you prefer one action over another in one situation – coffee over tea in situation A, say, when it's before noon – and you prefer the same thing in the opposite situation – coffee over tea in situation B, when it's after noon – then you should have the same preference when you don't know the situation: that is, coffee over tea when you don't know what time it is.

Remarkably, people don't respect this rule. In the early 1990s, for example, ›

K&T: Q3

- Which is greater:
  - the number of **six-letter** English words having "**n**" as their **fifth letter** or
  - the number of **six-letter** English words **ending** in "**-ing**"?
- Most people choose the group of words ending in "ing". Why? Because words ending in "-ing" are easier to think of than generic six letter words having "n" as their fifth letter.
- Fact: The group of six-letter words having "n" as their fifth letter words includes all six-letter words ending in "-ing".
- Psychologists call this type of mistake the **availability bias**
  - In reconstructing the past, we give unwarranted importance to memories that are most vivid and hence most available for retrieval.

[Amos Tversky and Daniel Kahneman, "Availability: A Heuristic for Judging Frequency and Probability," Cognitive Psychology 5 (1973): 207–32.]

Misuse of probability in law

- It is not uncommon for experts in **DNA analysis** to testify at a criminal trial that a DNA sample taken from a crime scene matches that taken from a suspect.
- How certain are such matches?
- When DNA evidence was first introduced, a number of experts testified that false positives are impossible in DNA testing.
- Today DNA experts regularly testify that the odds of a random person's matching the crime sample are less than **1 in 1 million** or **1 in 1 billion**.
- In Oklahoma a court sentenced a man named Timothy Durham to prison even though **eleven witnesses** had placed him in another state at the time of the crime.

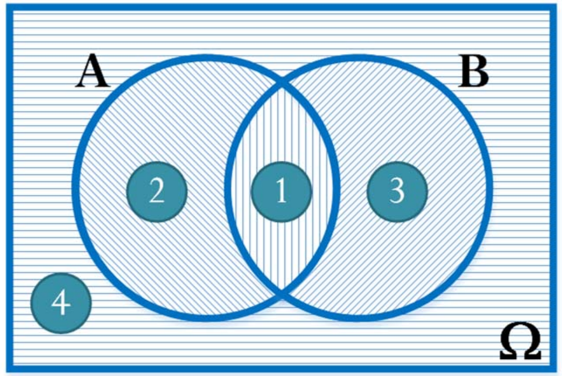
[Mlodinow, 2008, p 36-37]

Tips for Finding Event-Based Probability

- Don't forget that we always have an extra piece of information: $P(\Omega) = 1$.
- It is easier to work with expression involving intersection than the one with union.
  - Use de Morgan law [2.5] and complement rule [5.15]
  - For example, suppose we are given that $P(A \cup B^c) = 0.3$.
    - By the complement rule, $P\left((A \cup B^c)^c\right) = 1 - 0.3 = 0.7$.
    - By de Morgan law, $(A \cup B^c)^c = A^c \cap B$.
    - Therefore, the provided information is equivalent to $P(A^c \cap B) = 0.7$.

Tips for Finding Event-Based Probability

- Given n events, the sample space (Ω) can be partitioned into 2^n parts where each part is an intersection of the events or their complements.
- For example, when we have two events, the sample space can be partitioned into 4 parts:
  ❶ $A \cap B$,
  ❷ $A \cap B^c$,
  ❸ $A^c \cap B$, and
  ❹ $A^c \cap B^c$
  as shown in the Venn diagram.



- Any event can be written as a disjoint union of these parts. Therefore, if we can find the probabilities for these parts, then we can find the probability for any event by adding the probabilities of the corresponding parts.

Tips for Finding Event-Based Probability

- If your aim is simply to find one working method to solve a problem (not trying to find the smart way to solve it), then the steps on the next slide will be helpful.
- It turns the problem into solving system of linear equations.

Misuse of probability in law

- It is not uncommon for experts in **DNA analysis** to testify at a criminal trial that a DNA sample taken from a crime scene matches that taken from a suspect.
- How certain are such matches?
- When DNA evidence was first introduced, a number of experts testified that false positives are impossible in DNA testing.
- Today DNA experts regularly testify that the odds of a random person's matching the crime sample are less than **1 in 1 million** or **1 in 1 billion**.
- In Oklahoma, a court sentenced a man named Timothy Durham to more than **3,100 years** in prison even though **eleven witnesses** had placed him in another state at the time of the crime.

[Mlodinow, 2008, p 36-37]

Misuse of probability in law

State: Oklahoma

Charge: Rape, Robbery

Conviction: Rape, Robbery

Sentence: 3,200 Years

Incident Date: 05/31/91

Conviction Date: 03/13/93

Exoneration Date: 12/09/97

Served: 4 years

[https://www.innocenceproject.org/cases/timothy-durham/]

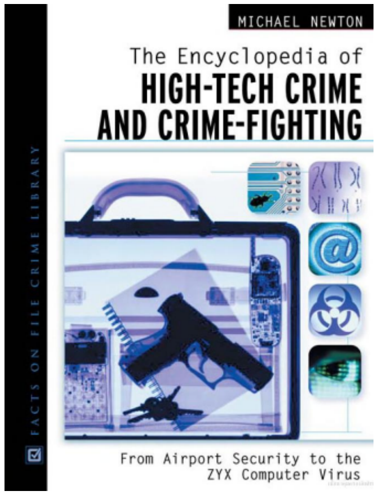
Lab Error (Human and Technical Errors)

- There is **another stat**istic that is often **not presented** to the jury, one having to do with the fact that **labs make errors**, for instance, in **collecting** or **handling** a sample, by accidentally **mixing** or **swapping** samples, or by **misinterpreting** or incorrectly reporting results.
- Each of these errors is rare but not nearly as rare as a random match.
- The Philadelphia City Crime Laboratory admitted that it had swapped the reference sample of the defendant and the victim in a rape case
- A testing firm called Cellmark Diagnostics admitted a similar error.

[Mlodinow, 2008, p 36-37]

Timothy Durham's case

- It turned out that in the initial analysis the lab had **failed to** completely **separate the DNA** of the rapist and that of the victim in the fluid they tested, and the combination of the victim's and the rapist's DNA produced a positive result when compared with Durham's.
- A later **retest** turned up the error, and Durham was released after spending nearly **four years** in prison.



[Mlodinow, 2008, p 36-37]

DNA-Match Error + Lab Error

- Estimates of the error rate due to human causes vary, but many experts put it at around 1 percent.
- Most jurors assume that given the two types of error—the **1 in 1 billion** accidental match and the **1 in 100 lab-error match**—the overall error rate must be somewhere in between, say 1 in 500 million, which is still, for most jurors, **beyond a reasonable doubt**.

[Mlodinow, 2008, p 36-37]

Wait!...

- Even if the DNA match error was extremely accurate + Lab error is very small,
- there is also another probability concept that should be taken into account.
- More about this later.
- Right now, back to notes for more properties of probability measure.