Instructors' Guide for

Translating Science into Policy The Role of Decision Science An Educational Module

Prepared by

Paul Brest Former Dean and Professor Emeritus (active) Stanford Law School

For

Committee on Preparing the Next Generation of Policy Makers for Science-Based Decisions Committee on Science, Technology, and Law

The National Academies of SCIENCES • ENGINEERING • MEDICINE

Revised April 20, 2017

Copyright © 2017 by Paul Brest. This module may be used subject to the terms of the <u>Creative Commons</u> <u>Attribution-NonCommercial 4.0 International Public License</u>. By using or further adapting the educational module, you agree to comply with the terms of the license. The educational module is solely the product of the author and others that have added modifications and is not necessarily endorsed or adopted by the National Academy of Sciences, Engineering, and Medicine or the sponsors of this activity.

Contents

Core Knowledge and Competencies	1
Calibrating the Text and Problems to Class Hours	1
Questionnaires	2
Problems	2
Slides	2
Evaluation of Student Work	2
II. Utility	3
III. Decision Making Where Risk or Uncertainty is Not a Major Factor	7
IV. Decision Making Under Risk and Uncertainty	8
V. The Intuitive Empiricist's Perspective and Errors1	15

Because this is a mini-textbook and the text is intended to be self-explanatory, most of the teaching notes do not focus on the text as such, but on the problems.

CORE KNOWLEDGE AND COMPETENCIES

Part II. The Concept of Utility

- The concept of utility
- Efforts to measure utility
- Systematic ways that people mispredict and subvert their future utility

Part III. Decision Making Where Risk or Uncertainty Is Not a Major Factor

- Axioms of subjective utility
- Subjective linear model—a simple and useful tool for multiattribute decision making
- Cost-benefit analysis

Part IV. Decision Making under Risk and Uncertainty

- Introduction to Bayesian statistics
- Expected utility under risk
- Decision trees
- Risk attitudes
- Prospect theory

Part V. The Intuitive Empiricist's Perspective and Errors

- The psychology of judgment and decision making, including
 - The idea of "two systems" for empirical judgment and estimation
 - Availability, anchoring, base rate neglect, hindsight bias
- Overconfidence, confirmation bias, naive realism, motivated reasoning
- The "cultural cognition" theory of risk perception

CALIBRATING THE TEXT AND PROBLEMS TO CLASS HOURS

I have taught all of these materials in four 2-hour classes, comprising the Introduction and Part II, followed by Parts III–IV. One could teach them in four 90-minute classes by omitting some of Part II and a few of the problems.

QUESTIONNAIRES

The first group of **slides (slides 2-29)** includes questionnaires, which students should submit in advance of reading the materials for the relevant classes. The aggregate responses are useful for class discussion. In some cases, to illustrate how different ways of framing the same problem can lead to different outcomes, a questionnaire will have two versions, each to be given to half the class. (To make it administratively simple, rather than do random assignment, I just divide the class in half alphabetically, and send one version to the first half and the other to the second.)

You can either have students answer all of the questionnaires before the first class, or for the class before it will be discussed (on the realistic assumption that they won't read ahead beyond the next class).

PROBLEMS

There are two sorts of problems.

- Assignments to be done before class. Handed in and/or discussed in groups during class before opening a class-wide discussion. Slides marked in red. If you plan to hand out the problems after the students have read the relevant text and you prefer that they not read the problems in advance, you can cut them out of the readings
- 2. In-class exercises. Discussed in groups during class before opening a class-wide discussion. Slides marked in purple.

SLIDES

I vary from year to year in how much I use slides in my teaching. Many but not all of the slides are pretty closely related to the text.

EVALUATION OF STUDENT WORK

Students can be evaluated on the assignments (marked in red) to be handed in. In addition, one could ask students to write a paper addressing a problem, like the one on climate change beliefs, which could incorporate decision making under risk or uncertainty as well as biases and their remedies.

II – UTILITY

If I were going to teach only portions of Part II, my priorities would be:

- Utility ≠ "Happiness," The inherent subjectivity of ends, Process as (Dis) Utility, The limits of revealed preferences: Individuals' mispredictions of their utility
- The diminishing marginal utility of wealth and other things (We return to this in the discussion of risk.)
- The Dubious Relationship between Daily Experiences and One's General Sense of Well-being
- Global Comparisons of Well-being
- National Measure of Utility

Slide 7: Utility

This unit is all about making good decisions. What is a good decision? One that maximizes utility. But what is utility? Economists tend to think of it in terms of consumer surplus and wealth. But does it also include happiness? Well-being? Freedom? Self-esteem? Accomplishment? Or some combination of these?

Slide 8: Why do mountaineering?

Mountaineering cannot readily be understood as an experience that yields any sort of sensory pleasure. You're constantly stressed that one wrong move will send you plummeting. You're mostly lonely since your climbing partners are 50 feet in front and behind. You're close to frostbitten. So why do it...

Class discussion:

- What analogous experiences have you had?
- What kind of utility does a mountaineer get from climbing a mountain?

Slide 9: Subjectivity of Utility

Just because I wouldn't do mountaineering doesn't mean that it isn't utility-enhancing for some individuals. Which is exactly the point. There are many components of well-being. Many of those components are nontangible. And those components are different for everyone.

Class discussion:

• Examples of you or people you know getting utility from activities that others might find "dis-utile."

Slides 10-13:

- The Decision Process Matters
- Choice Overload
- Pulling the Plug
- Ultimatum/Dictator Game

Problem: The Ultimatum and Dictator Games

The recipient's utility in the Ultimatum Game includes more than money—perhaps, dignity, pride, or a sense of fairness or how she will be perceived by others. The last two of these may be aspects of the Dictator's utility as well.

Ernst Ferh and Klaus Schmidt write in *Theories of Fairness and Reciprocity: Evidence and Economic Applications*:¹

The self-interest model has been very successful in explaining individual behavior on competitive markets, but it is unambiguously refuted in many situations in which individuals interact strategically. The experimental evidence on, e.g., ultimatum games, dictator games, gift exchange games, and public good games, demonstrates unambiguously that many people are not only maximizing their own material payoffs, but that they are also concerned about social comparisons, fairness, and the desire to reciprocate.

Gary Bolton and colleagues write in *Dictator game giving: Rules of fairness versus acts of kindness*:²

Our procedure suggests that dictator giving arises from a concern for fair distribution on the part of dictators. This is not to say that dictators give in order to improve the welfare of others. In our procedure, concerns for a fair distribution originate from personal and social rules that effectively constrain self-interested behavior—although within these constraints dictators do behave in a self-interested manner (they act first to secure what they consider to be their own fair share). What purpose these rules ultimately serve, whether it be to improve others' welfare or otherwise, is not clear from the data examined here.

Slides 14-15:

- Adaptation and Errors in Affective Forecasting
- Hedonic Adaptation

¹ Ernst Fehr and Klaus M. Schmidt, *Theories of Fairness and Reciprocity – Evidence and Economic Applications* (2001).

² Gary E. Bolton, Elena Katok, and Rami Zwick, *Dictator game giving: Rules of fairness versus acts of kindness*, 27 International Journal of Game Theory 2: 269–299 (1998).

Problem: Planning for the End of Life

Slide 16: Problem: Planning for the End of Life

An advance care directive is a legal document that a patient signs to instruct doctors not to sustain life-sustaining measures if the patient is no longer competent and suffering from an incurable disease whereby treatment would only prolong discomfort. Of course, the validity of an advance care directive depends on the individual's ability to make a decision now for a time in the future.

It would be ideal for counselors to help clients or patients to imagine themselves in the incapacitated position they are planning for. But how can they do this? Perhaps through statistics or stories about people changing their minds in such situations?

Would it be a good practice for lawyers, doctors, or other counselors to go beyond ensuring an individual is competent when providing advice about an advance health care directive? What else should they discuss?

Here are some resources on the subject:

- http://coalitionccc.org/2015/05/talking-with-young-adults-about-advance-careplanning/
- http://coalitionccc.org/wp-content/uploads/2014/02/Finding-Your-Way-English.pdf
- http://www.nap.edu/read/18748/chapter/5#158
- <u>http://theconversationproject.org/wp-</u> <u>content/uploads/2015/11/TCP_StarterKit_Final.pdf</u>
- https://mydirectives.com
- <u>http://www.huffingtonpost.com/jeff-zucker/the-selfie-that-isnt-self_b_8908674.html</u>
- https://www.youtube.com/watch?v=uo_Kjfw6kJk

Slides 17 and 18: Present bias

Video: The Marshmallow Experiment

https://www.youtube.com/watch?v=QX_oy9614HQ

This is a replication of Walter Mischel's famous experiment on the difficulties of deferring gratification. <u>https://en.wikipedia.org/wiki/Stanford marshmallow experiment</u>. Of particular interest is a follow-up student that "found unexpected correlations between the results of the marshmallow test and the success of the children many years later." The first follow-up study, in 1988, showed that "preschool children who delayed gratification longer in the self-imposed delay paradigm, were described more than 10 years later by their parents as adolescents who were significantly more competent." The Wikipedia entry describes subsequent follow-up studies. Some interesting questions, which I have not researched, is whether one can teach children to defer gratification, and if so, whether this has long-term consequences. Perhaps there are confounding genetic variables.

For a broader discussion of self-control, see Roy F. Baumeister and John Tierney, *Willpower: Rediscovering the Greatest Human Strength* (2012).

Slide 19: Impatience & Present Bias Slide 20: Overoptimism/Planning Fallacy

The Dubious Relationship between Daily Experiences and One's General Sense of Well-being

Slides 21-24:

- Two Concepts of Well-being (All Joy and No Fun)
- Utility: Past Present, and Future
- Momentary vs. Recollected Experience

Slide 25: Mountaineering Revisited

Problem: Measures of Utility

Slide 26: Measuring Utility

To preempt totally cynical responses, before the discussion, emphasize the last sentence: "Consider only what measures you deem to be most appropriate from a public policy perspective, and *not* what measures will help the mayor get re-elected." Some of my own answers below:

- *Childhood obesity*. Reduced obesity, diabetes and other health problems, reduced health care costs for the population, improved morale.
- *Homelessness.* Reduced homelessness and physical and mental disease and drug abuse among homeless individuals; improved sanitation of streets; residents and business customers feel safer and more comfortable
- *Elderly.* Improvements in their happiness and productivity and less demand on caregivers, as reported by the elderly and caregivers.
- *Congestion pricing.* Reduced congestion, pollution, and frustration and anger by drivers. Increased use of public transportation.
- *Cigarette smoking.* Reduced exposure to second-hand smoke, indicated by reduced complaints and reduced acute respiratory problems.

Problem: Easy cases for the liberal state?

Slide 27: Problem: Role of the Liberal State?

- 1. Expand services for people suffering from mental illness.
- 2. Teach children resilience skills.
- 3. Ban advertising to children.
- 4. *Reduce commute times*.

5. Strengthen social capital.

The first three interventions are on behalf of people who cannot legally consent and who therefore would seem to be appropriate objects of paternalism. In at least the second and possibly the third examples, however, parents may believe that these decisions should lie with them. But the third and fourth are beyond individuals' capacity to affect. All of them impose costs on someone. As for the fifth, is building social capital a government function or one properly left to civil society? Also, there is reason to believe that "bonding" social capital may compete with and undermine "bridging" social capital. For a general discussion of social capital, see https://en.wikipedia.org/wiki/Social_capital.

III – DECISION MAKING WHERE RISK OR UNCERTAINTY IS NOT A MAJOR FACTOR

Subjective Linear Model Problem

Slide 31: Subjective Linear Model Problem

The subjective linear model is a very simple tool and, based on feedback from former students as well as my counseling students and friends, a very powerful one for simplifying decisions.

Having students do this simple exercise makes them go through the example in the text and brings home the value of the tool. They often ask "is it OK to change the variables to make the decision come out the way I want it to?" to which my answer is: "Absolutely, if it's a personal decision. It's just a decision aid." Because I want students to go through the steps themselves; I don't give them the ready-made tool on this website in advance. http://www.somethingpop.com.

Students often have difficulty putting weights on the attributes, and so do I.

The subjective linear model is a good tool for making collective decisions. Though I haven't done this, it would be interesting to ask small groups of students to do this.

Slide 32: Expected Utility Theory Slide 33: Completeness Slide 34: Dominance Slide 35: Invariance Slide 36: Invariance Violation Slide 37: Independence from Irrelevant Alternatives Slide 38: Transitivity Slide 39: Intransitive Joan Slide 40: Cost-Benefit Analysis (CBA)

The CBA ratio shown on the bottom is useful in comparing the effectiveness of alternative interventions. But the simple subtraction shown on the top is an appropriate analysis for a single intervention. I'm grateful to Ricky Revesz for this point.

Cost-Benefit Analysis Problem

Slide 41: Cost-Benefit Analysis Problem

The problem doesn't require students to go through an entire cost-benefit analysis but only asks:

What further information about (1) the effectiveness of the program and (2) the program's costs and benefits for various stakeholders do you need to properly advise the commissioner? With respect to (1) focus questions involving evaluating the program's effectiveness. With respect to (2) begin by identifying all relevant costs and benefits.

If you and your students haven't studied program evaluation—and why should you have?—you might leave out the first question, or just mention that a before/after evaluation doesn't account for exogenous variables or bias in selecting the participants and that having some sort of comparison or control group is helpful.

- *Costs* include: payment to organization, transactions costs of negotiating agreement, evaluation administration? Should they include the costs of other people's jobs taken by parolees? Under the CBA discussion in the text, these are just transfer costs.
- *Benefits*: reduced beds, benefits to prisoners and families, to society from reduced crimes.

The example is a simplified version of a fairly popular kind of "social impact bond." See, for example, <u>http://www.thirdsectorcap.org/portfolio-item/massachusetts-juvenile-justice-pfs-initiative/</u>.

IV. DECISION MAKING UNDER RISK AND UNCERTAINTY

Bayesian Statistics

This is not essential to the rest of the course (except for a little note on base rate neglect in Section V, which you can omit). I teach it because it's new and interesting for many students.

Taxi Problem

Slides 45-52: Bayes' Theorem – Taxi Problem, and so forth

The sources of Bayesian priors:

- Base rates from frequentist statistics
 Population of taxis
- Nonquantified/nonquantifiable beliefs

- Population of black swans
- Belief that, as of 2003, Iraq did or did not possess weapons of mass destruction

P(H) often just reflects the subjective strength of our prior belief in H. When we are called upon to revise a prior belief in terms of new evidence, the task is not formulaic as much as intuitively combining the prior belief with the current data.

Problem: Does the Traveler Have Malaria

Slides 54-55: Does the traveler have malaria?

Problem: Expected Value of the Lotteries

Slide 56: Problem: What is the expected value of lotteries A and B?

A: 0.2 X \$1000 = \$200 + .08 X \$0 = \$0 = \$200

B: .06 X \$90 = \$54 .04 X \$0 = \$0 = \$54

Note that the probabilities for each branch must sum to 1.0.

Slide 57: Decision Trees Slides 58-65: Swine Flu

Note that in CBA, "costs" are the costs of taking action, and the costs of inaction are best thought of as negative benefits. This is counterintuitive for many people.

Problem: The Looming Storm

Slides 66-69: The Looming Storm

The key question facing the city is whether to evacuate, and so we're evaluating evacuating against not evacuating. The two slides show two different ways of representing the problem as a decision tree:

• The first treats the status quo of doing nothing as incurring \$0 costs and gaining \$0 benefits. Thus, any benefits of evacuating are positive.

• The second reflects the language that the mayor would use to describe the costs of inaction, that is, "inaction will cost lives."

The numbers come out the right way regardless of whether we frame the status quo has having no costs or benefits (and evacuation as providing a net benefit) or a loss (and evacuation as a lesser loss).

Attitudes toward Risk

Slides 70-72: Auction, Risk Preferences, Expected Value and Expected Preferences

Epidemic Problem

Slide 73: Epidemic Problem

The parentheses show the answers from various experiments. You might replace with the students' answers.

Slide 74: Framing a Medical Procedure

A little in-class exercise that suggests how the epidemic problem might come up in more quotidian contexts.

Slides 75-76: Prospect Theory

The (prescriptive or normative) essence of rational economic decision making is that decisions should be considered in terms of their effect on an individual's total wealth. Prospect theory is a generalization based on empirical or descriptive observations of behavior. It has three important characteristics.

- The value function is defined over gains and losses with respect to some reference point. That is, individuals do not think of outcomes in terms of their overall wealth—as neoclassical economics holds—but rather in terms of gains and losses with respect to a reference point.
- 2. Individuals are loss averse. Notice that the slope of the curve for losses is steeper than the slope for gains. This means that people experience losses based on that reference point as greater than equivalent gains. Loss aversion is captured by the phrase that "losses loom larger than gains."
- 3. Both the gain and loss functions display diminishing sensitivity. Notice that they are not straight lines, but that the gain function is concave and the loss function is convex.

The fact that individuals' choices may be based on how the decision is framed violates the axiom of invariance.

Slides 77-85: Prospect Theory: WTP vs. WTA

What explains loss aversion?

- Sometimes emotional attachment to an object or the status quo, for example, right to hunt, good visibility.
- Discomfort in participating in market transactions, especially with respect to goods that are difficult to value (environment).
- Regret avoidance—that giving up an entitlement is more likely to cause future regret than not gaining the entitlement in the first place.

Highway Speed Limit Problem

Slides 86-90: Highway Speed Limit Problem

You can do this little exercise on loss aversion in class. Students tend to treat the existing speed limit as the status quo. On average, students are more willing to reduce from 70 MPH to 65 MPH to save 50 lives than to increase from 65 MPH to 70 MPH at a cost of 50 lives. They also try to avoid what Philip Tetlock has termed the "taboo trade-off" between lives and other values.³

Against Ebola Foundation Problem

Slides 91-104

I use this, together with the ClimateWorks Foundation problem, to teach about risk attitudes. Students try to modify the hypothetical to avoid difficult trade-offs. You'll have to push back.

The Benefit/Cost-Expected return slide (Slide 97) looks at CBA and ER from a foundation's perspective—the former applicable to service delivery programs, or in this case the protective suits; the latter to advocacy programs or, in this case, to a vaccine development program. See my little article <u>Risky Business</u>.

Identifiable Victim Bias slides. You can have a lively discussion by asking students to justify their choices. While the choice of the protective suits doesn't protect identifiable victims as such, it protects real and certain victims compared to developing the vaccine.

Slide 104: Foundation's Risk Attitude in Grant Making

Risk neutrality maximizes expected return. Should a foundation take into account that its founders/donors, board, and staff may be risk averse? Should it take account of reputational risks? Note that most private foundations rely on endowments established by their

³ See, e.g., <u>http://leeds-faculty.colorado.edu/mcgrawp/pdf/mcgraw.tetlock.2005.pdf.</u>

founders/donors and not on fundraising. Some students think that a foundation should be risk seeking, to counter the general trend toward risk aversion. But this also compromises expected return.

ClimateWorks Foundation Slides (Slides 105-120)

- The first four slides set out the problem
- The next several slides show mitigation (reducing global warming) and adaptation strategies, and ask how one might compare them in terms of expected return—very difficult, since mitigation works globally while adaptation is highly local.
- The Sudoku shows how ClimateWorks once (it no longer does) calculated which regions and industries to focus on with the ultimate goal of avoiding 32.2 gigatons of carbon by 2030.
- The next group of slides relate directly to the questions asked at the end of the ClimateWorks case:
 - The big question: Based on CWF's performance to date, if you were the president or Board of the Hewlett Foundation, would you have voted to renew the commitment of \$500 million for another 5 years?
 - Like cost-benefit analysis, expected return analysis (ER) assumes that one can assign numbers to the variables with a reasonable degree of confidence. With respect to the likelihood of success (LOS) in ER, economists distinguish between risk and uncertainty, where the former is specifiable (e.g., probability of 0.6) within some reasonable margin of error and the latter is not.
 - Can you assign numbers to the benefit and LOS in the ORV and climate strategies?
 - If not, is the analytic framework nonetheless useful?
 - What strategies does ClimateWorks Foundation (CWF) use to mitigate climate change?
 - The CWF "Sudoku" (Ex. 2) quantifies technically feasible emissions reductions by 2030 by country/region and sector.
 - If you were CWF, how would you decide where to concentrate your work?
 - How could the ER formula aid your decision process?
 - Page 3 of the CWF case describes some of successes, which it terms singles and doubles, and two major strikeouts: the failure to reach an international treaty in Copenhagen in 2009 and the U.S. Congress's failure to enact comprehensive energy and climate legislation in 2010.⁴

On June 20, 2014 Stiftung Mercator hosted a conference on Science, Policy, and Philanthropy, which included a panel with James Wilsdon, Professor of Science and Democracy at the University of Sussex, and Paul Brest, former president of the William and Flora Hewlett Foundation. Professor Wilsdon criticized Hewlett and other foundations for supporting the risky

⁴ Larry Kramer, the current president of the Hewlett Foundation, estimates that during its first 6 years, CWF achieved a reduction of 3-4 gigatons of CO_2 toward its goal of 11 GT.

and failed effort to achieve a global agreement to reduce greenhouse gas emissions at the United Nations Climate Change Conference in Copenhagen in 2009. There followed an e-mail exchange between them.

Brest to Wilsdon

We all knew that affecting the outcome in Copenhagen ... was a high risk strategy. I told the Hewlett Foundation's board that I thought the chances of success were 5-10%. The basic framework within which anyone pursue policy advocacy is expected return: (benefit X likelihood of success)/cost. While the likelihood of success was low, we thought the benefit of either outcome was tremendous. As the saying goes, hindsight vision is 20-20; but decisions can only be made based on the odds ex ante. So my question for you is, if either strategy had succeeded, against all odds, would you have cited the Design to Win effort as a positive example of strategic philanthropy? If not, is it because you would have plugged different numbers into the equation and, if so, what? Or do you think the strategies were poorly conceived or executed? What am I missing?

Wilsdon to Brest

You're right of course that these criticisms are being made *ex post*, and it's not my purpose to question the noble intentions of those who were involved in the planning and execution of 'Design to Win'. But as part of any honest evaluation of the work of foundations in this area (something you yourself called for in our session), it seems to me entirely legitimate to evaluate 7-8 years on why 'Design to Win' failed in its own terms, and the lessons that this might offer to future philanthropic interventions at the science-policy interface (whether on climate, GM crops, synthetic biology, nuclear power, geoengineering or other contested issues).

Brest to Wilsdon

Evaluation can only be done ex post and, viewed retrospectively, the efforts to improve climate policies at Copenhagen and in the U.S. Congress were acknowledged failures. But a criticism of a strategy—especially a purposively high risk/high return strategy—must look at the decision from the ex ante point of view. The alternative is the well-known phenomenon of hindsight bias. I'm sure that the strategies were far from perfect, but just what do you think was wrong with them?

Wilsdon to Brest

Personally, I find it pretty shocking to spend half a billion dollars⁵ [actually around \$18 million, but still a large number] on strategic choices that were, in your own words, "acknowledged failures", with at best a 5-10% chance of success. But I guess that's why I stick to science policy rather than writing books on 'smart philanthropy."

 What is the fundamental disagreement between Brest and Wilsdon regarding the failures?

⁵ Professor Wilsdon perhaps was attributing the Hewlett Foundation's total 5-year unrestricted commitment to ClimateWorks, which covered all of the organization's activities.

- If a strategy fails to achieve its objectives, how can you tell whether it was well- or poorly designed and implemented?
- Different philanthropists, like different investors, may have different tolerances for risk. Broadly speaking, what kinds of programs should highly risk-averse philanthropists fund, and what kind of programs might risk-tolerant philanthropists fund?
- Should a foundation be risk neutral, risk averse, or risk seeking with respect to its grants budget (not its endowment)?
- Should a foundation that is willing to take some risks with its grants budget, act like a financial manager, who seeks a portfolio of investments with a diverse range of (uncorrelated) risks? In other words, should it balance grants that are pretty risky with ones that are pretty sure to succeed? Why, or why not?
- Consider this hypothesis: A foundation with a \$100 million budget, that wishes to
 mitigate global warming through (risky) policy advocacy strategies, should be
 indifferent between (1) making a \$100 million grant to an organization it believes to
 be highly effective, and (2) making 10 \$10 million grants to 10 organizations it
 believes to be highly effective. Assume that, viewed in isolation, the ER of both
 strategies is the same.
- At the same panel discussion where Wilsdon criticized the risky strategies of a group
 of foundations that supported advocating for policies to mitigate climate change, he
 also argued that it was narrow-minded of those foundations to focus on policy
 advocacy to the exclusion of supporting (1) adaptation to the inevitable
 consequences of global warming, (2) the development of innovative energy-saving
 technologies, and (3) the development of innovative geoengineering technologies.
 What facts would affect your judgment on this issue?

Attribution vs. contribution. There is much talk in the foundation and nonprofit sector about attributing outcomes to a particular organization or funder, with an emphasis on appropriate modesty in making claims. But is trying to discern attribution actually important for reasons besides PR or self-congratulation, and what might they be? Can you propose useful definitions of attribution and contribution?

Diversification Slides (Slides 121-122)

We now return to Ebola problem, giving students an opportunity to spread the \$10 million between vaccine development and the protective suits. Ask students who chose the vaccine because of its higher expected return and who now diversify why they would sacrifice expected return. Some people make an implicit analogy to an investment portfolio, where diversification reduces risk. But why should a foundation's grants be risk-averse? If this year's grants have no impact at all, the foundation can try again with next year's grants. Moreover, it may be most efficient for our foundation to focus entirely on one strategy and another foundation to focus on another.

V. THE INTUITIVE EMPIRICIST'S PERSPECTIVE AND ERRORS

Two Systems Slides (Slides 130-133)

Two systems. This is where we introduce the two-systems view of decision making that pervades the book and which is one of the cornerstones of judgment and decision-making (JDM) research. Here, or later, one can say a few words about the origins and intellectual history of JDM research. While they didn't start on a blank slate, Amos Tversky and Daniel Kahneman are the founders. Tversky died in 1996, and Kahneman received the Nobel Prize in Economics in 2002, essentially for both of their work. The new field of behavioral economics grows out of the JDM research. Although the two-systems approach precedes fMRI (functional magnetic resonance imaging), the imaging work indicates that affective responses (System 1) and cognitive work (System 2) may take place in different parts of the brain.

The visible part of the Iceberg is System 2.

People's answers to the What's His Mood (Slide 134) and What's Her Attitude (Slide 134) slides are based on System 1.

WYSIATI (Slides 136-137) comes from Daniel Kahneman's very readable *Thinking Fast* and *Slow*.

Show the students **What's Unusual about this hand (Slide 138)** for just a couple of seconds and ask what they saw? Many don't notice the black 3 of hearts because they tend to see what they expect. The **Tamir Rice slide (Slide 139)** also is about expectations. You may or may not want to wade into the controversies around more recent police shootings.

Heuristics and Biases (Slide 140) Availability Slides (Slides 141-144) Anchoring Slides (Slides 145-157)

How it works

Adjustment process—cognitively effortful

Anchor may increase the *availability* of features that the anchor and the number to be determined—the target—hold in common, selectively *activating* information about the target that is consistent with the anchor:

- Real estate anchor through price. Think of good/bad features.
- Computer programmer. Think of similar jobs you know high/low.

Debiasing

Techniques that don't work: Warning people of the phenomenon, providing incentives for accurate estimates.

Techniques that may work: Perhaps drawing attention to the features of the target that differ from the anchor, or providing a different anchor.

Base Rate Neglect Slides (Slides 158-159) Hindsight Bias Slides (Slides 160-166)

Mark Kelman, David Fallas, and Hillary Folger identify three categories of hindsight bias:

Primary: Your <u>revised</u> (ex post) estimate of the ex ante probability of an event that has occurred is higher than your estimate of that event before outcome was known.

- *Rational.* If the event provides better information about ex ante probability, for example, that the coin was loaded or the deck was stacked.
- *Irrational.* If you think the probability is higher than it actually is based, say, on a sampling error—for example, it is irrational if you think that getting five heads in a row means that the probability of this coin's coming up heads was greater than 0.5.

Secondary: You mistakenly believe that your (higher) hindsight estimate of probability is the same as your (lower) foresight estimate, that is, that you "knew all along" that the actual outcome would occur, when in fact you didn't. People tend to forget their actual foresight estimates. Secondary hindsight bias may help you remember a particular probability, but it won't help you improve the process of prediction.

Tertiary: You mistakenly believe that a third party lacking outcome information made an unreasonably low ex ante estimate of the probability of the actual outcome, and impose blame or liability. This is relevant to legal liability—tort, criminal liability, prudent investor, search and seizures. But note that what may appear to be hindsight bias may be a jury's (unauthorized) imposition of absolute liability or nullification of exclusionary rule when cops found drugs.

The mechanisms by which hindsight bias operates are analogous to those underlying the availability heuristic and anchoring bias. More specifically: We develop a theory based on the actual outcome of an event and believe that we always held that theory. It is a natural (and useful) tendency for the brain to automatically incorporate known outcomes into existing knowledge, and make further inferences from that knowledge. Ignoring a known outcome is unnatural. But...

In many cases of disasters, responsible officials have been warned about risk—for example, airline security before 9/11. There are obvious hindsight-bias related problems here. A decision maker can't devote resources to mitigating every possible disaster. How can one hold an official accountable for poor judgment without falling prey to hindsight bias?

*Debiasing_*is difficult. Asking people to describe alternative possible outcomes doesn't work. It's cognitively difficult to construct an alternative scenario to the one that you know occurred. If ease of recall contributes to availability, this difficulty makes the alternative less available and therefore seem less probable.

Legal rules sometimes sidestep the problem through burdens of production and standards of proof; suppressing evidence that would exacerbate bias, for example, repairs; industry custom, standards of conduct; and bright lines and safe harbors.

Emotions Slides (Slides 167-172) Slide 173: Confirmation Biases

In the taxi problem, P(H), the prior probability, or the strength of our belief in H before obtaining event-specific data, was determined by the base rate of the event in question (the proportion of Blue taxis).

But many beliefs cannot be described in terms of base rates. We rely on our own experience. Or, for example, in believing that, as of 2003, Iraq did or did not possess weapons of mass destruction, we rely on the assertions of others—teachers, scientists, politicians, journalists—whom we may deem more or less accurate or trustworthy.

In these cases, P(H) is does not reflect a calculable probability, but simply the subjective strength of our prior belief in H. When we are called upon to revise a prior belief in terms of new evidence, the task is not formulaic as much as intuitively combining the prior belief with the current data. Confirmation bias can play a role here.

In developing a hypothesis, it may be appropriate to seek confirming data, but in testing it's important also to look for disconfirming evidence. Most of the instances of confirmation bias examined below involve testing.

Q. What is danger of hearing an estimate from even a source we deem reliable? [anchoring]

Why are people overconfident?

Availability: We pay attention to strength, or extremeness, of evidence vs. its weight.

- *Strength* size of the effect
- Weight significance or reliability (sample size or base rate)

We recruit evidence that favors our hypothesis.

We anchor on strength and adjust (insufficiently) for its weight.

The vividness of particular impressions trumps our knowledge of base rates. After interviewing individuals, participants were asked to predict whether, given the choice, the interviewees would take a free subscription to *Playboy* or *The New York Review of Books*. Even when told that 68% of all the people interviewed preferred *Playboy*, participants tended to ignore this base rate and premise their predictions on their own impressions from the interviews. They were only correct 50% of the time, but 72% were certain that their predictions were accurate.

People are generally overconfident about general-knowledge items of moderate or extreme difficulty, such as many of the trivia questions.

While knowledgeable subjects responding to easy questions tend to be *underconfident*, also a deviation from good calibration.

With the possible exception of authoritarian personalities, individual differences don't seem to correlate with overconfidence.

Overconfidence Dyad Exercise Slides (Slides 174-178) Confirmation Biases: Barriers to Revision (Slide 179) You can use any trivial pursuit-type item to be estimated—though some students will inevitably say that they had superior knowledge to their partner. Collect the dyads' differences. In the two classes I've taught this, there is some movement toward the mean in the first part, more movement toward the mean in the second part, but still divergence. The best strategy in conditions of ignorance is to adopt the mean of the two estimates. Many students resist this, but putting them in the position of an observer helps.

Naive Realism Slides (Slides 180-187) Debiasing Slides (Slides 188-189) Slides 190-191: Beliefs about climate change, Open carry problem