Recall

- Formal view
  - argument, conclusion, premise, soundness, validity
  - pitfalls of language (vagueness, equivocation, etc.)
  - valid argument forms (modus ponens, etc.)
  - informal fallacies (straw man, false dilemma, etc.)
  - analogy (formal, informal uses, etc.)
- Psychological view
  - human performance on statistical reasoning tasks

Interpretation of statistical data

- People are very good at seeing patterns

Mary-in-a-tree

9-11 A Demon's-eye View

God signed the tsunami

Priming

- Priming helps make this kind of arbitrary pattern recognition worse (it's probably at work in the images earlier too)
- Priming is being given some (possibly vague) characterization of a pattern you might find.
- E.g. Banditos by the Refreshments:

Hot hand

- Performance in basketball:
  - "Success breeds success" and "Failure breeds failure".
  - Not a real phenomenon
  - Chances of making a shot are statistically independent
  - Real issue: explanations for hot hand
  - Masked by extra attention: but, not true for free-throws
  - Players know: but, not predictable by the players themselves

Explanation for seeing 'hot hand'

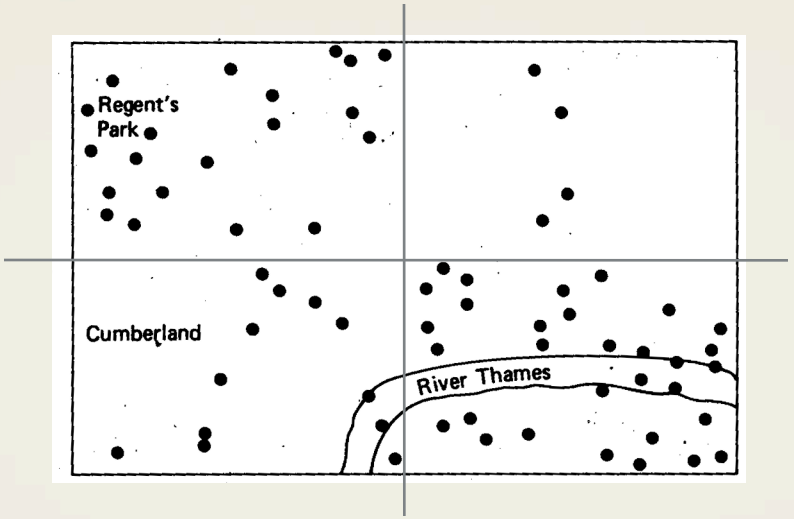
- Our preconceptions bias our interpretation of the data we are given.
  - Look for confirming evidence only
  - Ignore disconfirming evidence
- We are bad at knowing when a sequence of events is truly random.
  - Heads and tails experiment
  - The mistake isn't in seeing streaks, but it's interpreting them *as* streaks (i.e., predictable)

Representativeness

- Kahneman and Tversky (Nobel Prize) suggest chronic misperception of random events is an example of ‘representativeness’
- Errors in which people assume that the local features of an instance are the same as the global features that define a category
  - i.e. we expect a small series of events to be "As random" as a long series
- Related to, but more general than stereotyping

Misreading statistics

- Because we assume data is 'intuitively representative', we often misread statistics (V-1 missile)



- Can, in *hindsight*, find statistically significant results
- We need to verify, on an *independent* set of data, the hypotheses we generate given those statistically significant results

Generating explanations

- Exacerbating this misreading is our ability to generate seemingly reasonable explanations *post hoc*.
- Corpus callosum damage (split brain)

shovel

chicken head

Generating Explanations (cont.)

- We identify 'purpose' (i.e. intentional patterns) behind random or accidental events
- Interpretation is often based on ‘running hypotheses’, so accidental patterns are taken to support such hypotheses (conf. bias)
- The circle is quite tight -- and may be vicious

‘Normal’ confabulation

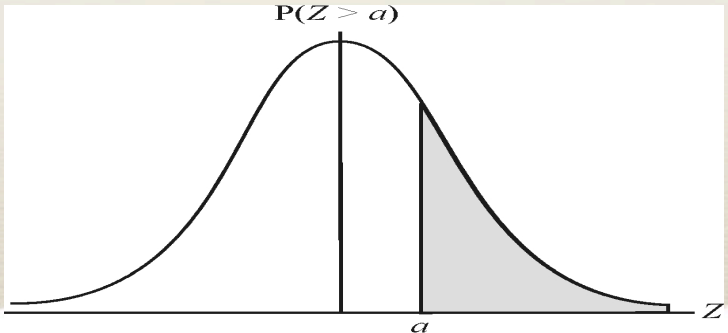
- A study by Marshall & Hanssen showed subjects a 42 second film that was comprised of a man pushing a pram, pulling its protective net down and then walking off while a woman appeared out of a house. Statements were taken immediately and one week later from all of the subjects. The results showed firstly, that police noted and remembered more detail, but were prone to special types of error (especially where actions were concerned). Secondly, that Police officers remembered more correct details about dress and appearance, and recalled twice as many incorrect facts (i.e. details that were not there and events that never happened) as the citizens. Thirdly, that the civilians merely reported what was there (with actions) without imputation of suspicious intent. Fourthly, that the police saw more than actually happened as 1 in 5 said they saw the man put his hand into the pram and take the baby out and many said that they saw the woman running towards the man with a look of worry on her face. Fifthly, that while police were better at remembering details initially (up to one minute) they made more errors one week later. In contrast, civilians were found to remember less initially but forget less over time. Therefore, the study assumed that time (among other factors) definitely interferes with the memory of a police officer. Finally, the study found that under longer exposure times the policemen recalled more than civilians immediately after the incident.

Confabulation

- So why is confabulation a problem?
  - It's false
  - It's easy
  - It goes unnoticed (try to catch yourself)
  - It's convincing
    - *It gives 'reasonable' explanations for non-existent phenomena*

Statistical regression

- The regression effect: the tendency of statistical data to take on its average
  - Very tall parents have tall children, but not very tall children
  - A company's disastrous years tend to be followed by more profitable ones
  - People who do extremely well in high school tend to do well in college but not as well (?)



Statistical regression (cont.)

- People tend to be insufficiently conservative when making predictions
- Predicting GPAs based on previous GPAs or on their sense of humour
- The better the basis of our prediction, the less regressive we need to be
  - people don't follow this rule at all

Statistical regression (cont.)

- Regression *fallacy*: don't recognize statistical regression as the best explanation and instead generate superfluous or complicated causal theories (i.e., confabulate)
  - Sports Illustrated jinx
  - Punishment vs. reward for behaviour

Morals of the stat. story...

- From the clustering illusion and regression fallacy:
  - We are apt to misinterpret data that we are presented with
  - We see patterns that are not really there
- Together, CI and RF can combine to produce all kinds of strange beliefs
- Also, we have a tendency to generate bad explanations to account for those 'patterns'
- Once such explanations are taken seriously, they inform how subsequent data is interpreted!

Necessary and sufficient evidence

- Necessary evidence: evidence which *must exist* in order for the belief to be true
- Sufficient evidence: *in itself*, makes the belief likely to be true
- Necessary evidence might not be sufficient, if some more necessary evidence is needed (and vice versa, if there is other sufficient evidence)
- Nevertheless, the mere existence of evidence is often cited as a *reason* that the belief is true (but it isn't!)

Base rate

- Suppose 90 percent of people who have a GRE over 700 complete their graduate training
- You might think that GRE scores provide a good way of determining whether or not someone will complete their graduate training
- However, if 90 percent of people complete their graduate training it regardless of their GRE score, that conclusion would be incorrect
- You have to consider the *base rate* (i.e. the likelihood of success regardless; e.g. 90 percent)

Statistical evidence

- Need to compare the probability of success with a high GRE a/(a+b) with the probability of success a given a low GRE c/(c+d)

	Succeed	*Doesn't succeed*
high GRE	a	b
low GRE	c	d

Statistical evidence (cont.)

- People tend to rely more than they should on confirmatory evidence, cells *a* and *d*.
  - they look for (some) necessary but not sufficient evidence
  - this can clearly lead us astray when assessing evidence for a hypothesis...

Example 1

- Does bumpiness help detect redness?
- Looking just at 'a' (and 'b'): yes since 80% of the red balls are bumpy
- But, it doesn't help determine redness at all: success 40/(40+10)=.8, failure: 45/(45+5)=.9. (Overall 40/(40+45)=.42)

High base rate of bumpy surface!

	bumpy surface	*smooth surface*
Red	40(a)	10(b)
Not red	45(c)	5(d)

Example 2

- Does texture help detect colour?
- Looking just at *a* and *d* we might say yes because as many red balls are bumpy as non-red balls are smooth, so maybe texture determines redness
- But, bumpy gives red: 8/(8+32)=.2, bumpy gives not-red: 2/(2+8)=.2.

High base rate of red!

	bumpy surface	*smooth surface*
Red	8(a)	32(b)
Not red	2(c)	8(d)

Confirmation bias

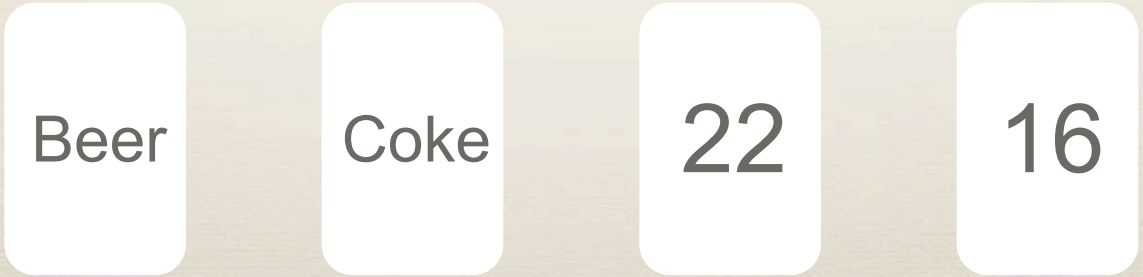
- We seem to enjoy ‘yes’ more than ‘no’ although information content is identical
- We prefer information easiest to deal with cognitively
- We prefer to consider information that is framed positively rather than negatively
- Result: tendency to review only part of the relevant data
  - Subjects emphasize cell a (or d)
  - Penn & Teller AA example

Confirmation Bias: Wason Task

- True for active information gathering too (seek confirmatory evicence)
- Abstract: Vowel → even number

10-30%

- Social: Alcohol → over 21



90%

Bias can be overcome with experience

Confirmation bias example

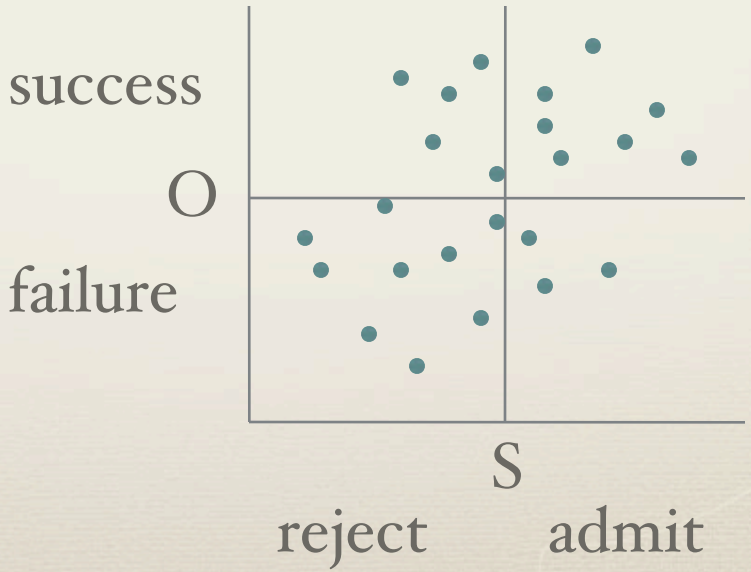
- Interviewers were asked to determine extroversion or introversion of subjects.
- Often asked questions that they expected to confirm their hypothesis.
- There are two problems:
  - 1) sometime the question can only allow information consistent with the hypothesis to be elicited (e.g., what would you do if you wanted to liven things up at a party?); and
  - 2) confirmation may seem to be the case if a positive response to the question is likely whether or not the hypothesis is true (e.g., do you sometimes feel that it is hard for you to really let yourself go at a party?).
- Memory search falls prey to the same problems:
  - Such searches may only allow eliciting consistent evidence
  - Such searches may encourage positive responses

Confirmation bias (cont.)

- *Sheer number* of confirmatory cases for hypothesis can affect a decision
- People find *more* information *more convincing* information
  - Subjects claim that East and West Germany are both more similar and more different than Sri Lanka and Nepal
  - Explanation: people know of more similarities and more differences between the former than the latter. Whether differences or similarities are recalled depends on the hypothesis at hand.

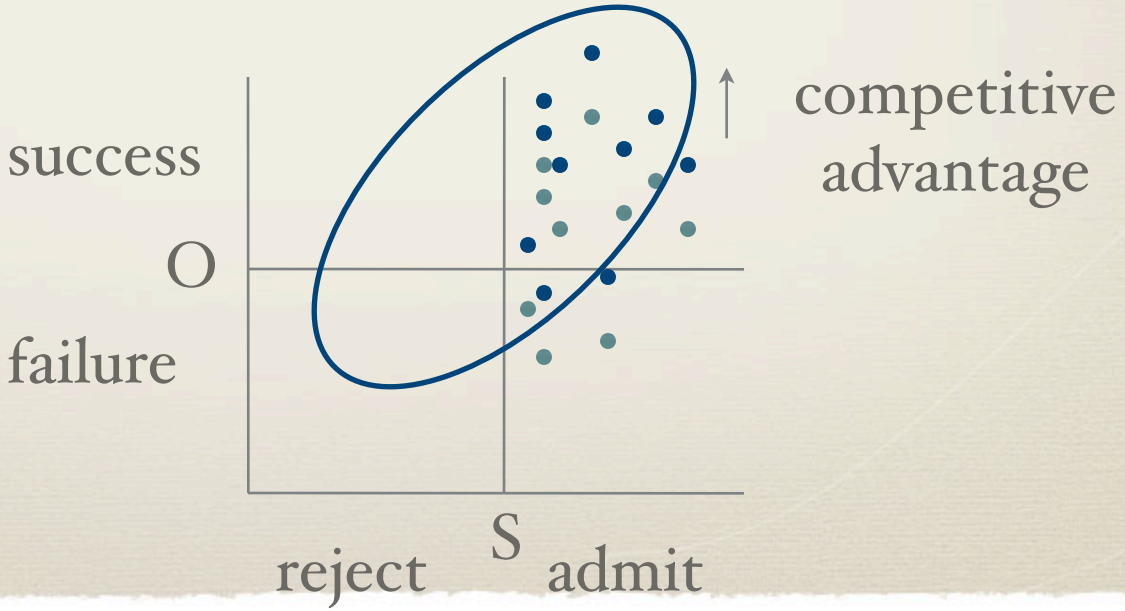
Missing data

- Want to know if the selection criterion predicts future performance
- Here, all relevant data is available



Missing data (cont.)

- Need to compare: success and success was predicted *versus* success and failure was predicted
- Often we can't carry out this comparison



Missing data (cont.)

- Problems:
  - when the base rate of success is high, you can often be misled into thinking that the selection criterion is effective
  - merely being accepted might give a competitive advantage
  - a selection criterion may seem extremely effective when in fact it is only marginally so
- Experts are only immune to this "Illusion of validity" when the performance of those they rejected is available to them

Missing data (cont.)

- In many domains, this information isn't available
  - Public policy (Afghanistan, privatizing utilities, health care)
    - No identical conditions
    - No controlled experiments
    - Bad policies can look good accidentally and vice versa
  - Some good indicators can be missed (e.g. GPA and SAT correlations seem lower than they are).
  - Lifestyle differences (only get info about those we associate with)

Self-fulfilling prophecies

- Special case of missing data: beliefs determine actions that make their consequences more likely (rumours of an insolvent bank)
- Seemingly-fulfilled prophecies are those that cause us to act in a way to that it is impossible for our expectations to be disconfirmed
  - thinking someone is unfriendly and thus avoiding them
  - baldness

Summary

- All of these different sources of error serve to provide us with incomplete or bad information that can lead to false conclusions or evaluations of belief.
- While it might be nearly impossible, to avoid all of these kinds of error, by knowing that they are present we can learn to make better decisions.
- The fact that careful evaluation of statements (and re-evaluation) can help us avoid some of these errors should make it clear how important and useful certain methods are for generating belief (e.g., statistical methods and the scientific method).

Websites

- 1) http://shazam.econ.ubc.ca/flip/index.html
  - Basically you select pennies or dimes, how many, and the page will return the result of flipping that many coins
- 2)http://www.ship.edu/%7Edeensl/mathdl/stats/Poker.html
  - See why so few poker hands are any good
- 3)http://stat-www.berkeley.edu/users/stark/Java/Venn.htm
  - A wonderful Java applet (opens in a separate window) that lets you fool around with Venn diagrams and probabilities
- 4)http://javaboutique.internet.com/BallDrop/
  - the "Bell Curve"

Question

- Question: What kind of evidence are people least likely to seek when evaluating a belief? Give one example of how this is a problem.