Two Types of Thinking:
Intuitions, Ethics & Bounded Rationality

Peter Danielson
Univ. of British Columbia
COGS 300.002
Mar 10 2015
2 Kinds of Questions about Rationality: Prescriptive & Descriptive

• How ought we to decide?
  – Can we immunize ourselves from System 1’s defect?

• Should our artificial agents be designed to use rational choice or human prospect theory (mix of system 1 and 2)?

Are real (human) agents “designed” using prospect theory rather than utility theory?

• Today: Evidence & Psychological theory
An Historical aside: Simple Experiments Can Change the Scientific World

• Simon 1957: Bounded rationality
  – A footnote to be explained away for economists
• Tversky & Kahneman – late 70’s – early 80’s
  – Simple experiments (at UBC’s Psych Dept.) refute descriptive theory of rational choice
Philip E. Tetlock began systematically quizzing 284 political experts — most of whom were political science Ph.D.’s — on dozens of basic questions, like whether a country would go to war, leave NATO or change its boundaries or a political leader would remain in office. His book “Expert Political Judgment: How Good Is It? How Can We Know?” won the A.P.S.A.’s prize for the best book published on government, politics or international affairs. (2005) Professor Tetlock’s main finding? Chimps randomly throwing darts at the possible outcomes would have done almost as well as the experts.
Intuitions as Object of Study

- “We studied intuitions -- thought and preferences that come to mind quickly and without much reflection (p. 449).”
Kahneman’s Model

Perception
- Fast
- Parallel
- Automatic
- Effortless
- Associative
- Slow-learning

Intuition
- System 1

Reasoning
- System 2
- Slow
- Serial
- Controlled
- Effortful
- Rule-governed
- Flexible

Percepts
- Current stimulation
- Stimulus-bound

Conceptual representations
- Past, Present and Future
- Can be evoked by language

Figure 1.
Perceptual Q1

• “Which of the three figures is larger on the screen?"
  A. Top
  B. Middle
  C. Bottom
  D. All same size
  E. Cannot be determined
Perceptual Q1

• “Which of the three figures is larger on the screen?”

A. Top
B. Middle
C. Bottom
D. All same size
E. Cannot be determined
System 1: Perceptual?
Bateson: Watching and Cooperation
Choose the best answer.

What is/are (a) plausible reason(s) students at ivy league institutions incorrectly answer this riddle?: "A ball and bat cost $1.10 together. The bat costs $1.00 more than the ball. What does the ball cost?"

a) They are poorly skilled at basic arithmetic
b) They are highly skilled at basic arithmetic
c) There intuitive confidence neglects attention to the effortful system of operation.
d) b & c
e) $0.10 (this is a joke)
Rate Quiz Question 2

A. Excellent
B. Very Good
C. Good
D. Acceptable
E. Poor
Prospect theory differs from Bernoulli's concept of utility in that:

a) outcomes are states, as opposed to changes
b) it is concerned with short-term outcomes, not long-term
c) it's more useful for prescriptive purposes than for descriptive purposes
Prospect theory differs from Bernoulli's concept of utility in that:

a) outcomes are states, as opposed to changes
b) it is concerned with short-term outcomes, not long-term
c) it's more useful for prescriptive purposes than for descriptive purposes
Problem 1 – The Asian Disease

Imagine that the United States is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

<see handout for details on programs>

Which of the two programs would you favor?
Problem 1 – The Asian Disease

Imagine that the United States is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

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

Which of the two programs would you favor?
Imagine that the United States is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:

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

Which of the two programs would you favor?
Data on Problem 1

- If Program A is adopted, 200 people will be saved. [TK: 72% N = 152]
- If Program B is adopted, there is a 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved. [TK: 28%]
- If Program C is adopted, 400 people will die. [TK: 22% N = 155]
- If Program D is adopted, there is a 1/3 probability that nobody will die, and 2/3 probability that 600 people will die. [TK: 78%]
Prospect Theory’s Value Function

The value function is defined on gains and losses and is characterized by four features:

1. it is concave in the domain of gains, favoring risk aversion;
2. it is convex in the domain of losses, favoring risk seeking;
3. Most important, the function is sharply kinked at the reference point, and loss-averse – steeper for losses than for gains by a factor of about 2–2.5 (Kahneman, Knetsch, & Thaler, 1991; Tversky & Kahneman, 1992). (4) Several studies suggest that the functions in the two domains are fairly well approximated by power functions with similar exponents, both less than unity.

However, the value function is not expected to describe preferences for losses that are large relative to total assets, where ruin or near-ruin is a possible outcome. (p. 462, references omitted)
Q4: Normative Rational Choice requires:

Since A=C and B=D, normative rational choice requires: [corrected]

A. A is preferred to B and C is preferred to D
B. B is preferred to A and D is preferred to C
C. A is indifferent to B and C is indifferent to D
D. A is preferred to B if and only if C is preferred to D
E. None of the above
Thinking slow (with ethical math)

A. 200 saved
B. $\frac{1}{3} \times 600$ saved

Doesn’t need $\frac{1}{3} \times 600 = 200$

C. $(600 - 400)$ saved
D. $(\frac{2}{3} - \frac{1}{3}) \times 600$ saved

Only that $\frac{1}{3} \times 600 \times$ discount same B & D.
According to Kahneman, invariance is an essential aspect of rationality which:

A) is violated in demonstrations of framing effects such as the Asian disease problem
B) Is an aspect of the phenomenon known as doubt: a meta-cognitive appreciation of one’s ability to think incompatible thoughts about the same thing.
C) cannot be achieved by a finite mind
D) A & C
E) Is one of the three heuristics of judgment described by Tversky & Kahneman in 1974
 Quiz 5

According to Kahneman, invariance is an essential aspect of rationality which:

A) is violated in demonstrations of framing effects such as the Asian disease problem
B) is an aspect of the phenomenon known as doubt: a meta-cognitive appreciation of one’s ability to think incompatible thoughts about the same thing.
C) cannot be achieved by a finite mind
D) A & C
E) is one of the three heuristics of judgment described by Tversky & Kahneman in 1974
Descriptive Realism of Rational Choice Theory

“The impossibility of invariance raises significant doubts about the descriptive realism of rational-choice models (Tversky & Kahneman, 1986). Absent a system that reliably generates appropriate canonical representations, intuitive decisions will be shaped by the factors that determine the accessibility of different features of the situation. Highly accessible features will influence decisions, while features of low accessibility will be largely ignored. Unfortunately, there is no reason to believe that the most accessible features are also the most relevant to a good decision (p. 459).”
Canonical Representations

• What is a canonical representation of Problem 1?
• I.e. How ought we ethically, to decide it?
• 1 life saved: basic unit (of good)
• A = C = 200 saved
• B = D = 1/3 x 600 saved = ??
Persistence & Optical Illusions

On several occasions we presented both versions to the same respondents and discussed with them the inconsistent preferences evoked by the two frames. Many respondents expressed a wish to remain risk averse in the "lives saved" version and risk seeking in the "lives lost" version, although they also expressed a wish for their answers to be consistent. In the persistence of their appeal, framing effects resemble visual illusions more than computational errors.
Primbing System 2: Social Contract

- Veil of Ignorance
- Harsanyi: Equal probability
- Rawls: Worst off
- How to represent probabilistic outcome?
Q6:

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?

A) Linda is a bank teller.

B) Linda is a bank teller and is active in the feminist movement.
Q6: The Linda Problem

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?

A) Linda is a bank teller.
B) Linda is a bank teller and is active in the feminist movement.
The Linda Problem

• Substituting representativeness for probability
• Criticisms: Is the Linda question fair?
  – “And” and conjunction fallacy
  – Probability applied to single cases
Affect Heuristic

...is probably the most important development in the study of judgment heuristics in the last decades. There is compelling evidence for the proposition that every stimulus evokes an affective evaluation, which is not always conscious.

– How happy are your life in general?
– How many dates did you have last month?
Extreme Distortions

- “Evaluate attractiveness of purchasing new equipment for use in the event of a crash landing of an airliner.”
- “Saving 150 lives is diffusely good...whereas saving 98% of something is clearly very good...”
Cognitive vs. Affective Biases

The failure to identify the affect heuristic much earlier, as well as its enthusiastic acceptance in recent years, reflect significant changes in the general climate of psychological opinion. It is worth noting that in the early 1970’s the idea of purely cognitive biases appeared novel and distinctive, because the prevalence of motivated and emotional biases of judgment was taken for granted by the social psychologists of the time. There followed a period of intense emphasis on cognitive processes, in psychology generally and in the field of judgment in particular. It took another thirty years to achieve what now appears to be a more integrated view of the role of affect in intuitive judgment.
Nudge: Choice Architecture

Design & Policy Implications

D’oh, the donuts are wayy over there. Think I’ll have fruit.

My cunning choice architecture will soon have Homer eating healthy.
Discussion Question

Explain the phenomenon that after a while you stop being as excited or happy about the new iphone you got with the concepts introduced in the video.

Anoosh
Rate Discussion Question 1

A. Excellent
B. Very Good
C. Good
D. Acceptable
E. Poor
In light of the findings discussed in the preceding section, it is useful to consider situations in which people will not neglect extension completely. Extension effects are expected, in the present model, if the individual (i) has information about the extension of the relevant set; (ii) is reminded of the relevance of extension; and (iii) is able to detect that her intuitive judgments neglect extension. These conditions are least likely to hold—and complete neglect most likely to be observed—when the judge evaluates a single object and when the extension of the set is not explicitly mentioned. At the other extreme, the conditions for a positive effect of extension are all satisfied in psychologists’ favorite research design: the within-subject factorial experiment, in which values of extension are crossed with the values of other variables in the design. As noted earlier, this design provides an obvious cue that the experimenter considers every manipulated variable relevant, and it enables participants to ensure that their judgments exhibit sensitivity to all these variables. The factorial design is therefore especially inappropriate for testing hypotheses about biases of neglect (Kahneman & Frederick, 2002).

In spite of these objections, within-subject factorial designs have been used in several experimental studies of extension neglect. Figure 10 illustrates the remarkably consistent additive extension effect that has emerged in these experiments (Schreiber & Kahneman, 2000). In each of the experiments, the extension variable has a slight but significant effect, and combines additively with other information. The additivity is noteworthy, because it is normatively inappropriate. For each panel of Figure 10, a compelling normative argument can be made for a quasi-multiplicative rule in which the lines should fan out.

The observed pattern is compatible with a process of anchoring and adjustment: the intuitive judgment provides an anchor, and small adjustments from that anchor are made to accommodate the role of extension.

![Pain intensity reported by two colonoscopy patients.](Image)
Experiencing vs. Remembering

- B experiences more pain
- A remembers experience as more painful
- A will be used in future choices
  - Is this a system 1 failure of rationality?

In light of the findings discussed in the preceding section, it is useful to consider situations in which people will not neglect extension completely. Extension effects are expected, in the present model, if the individual (i) has information about the extension of the relevant set; (ii) is reminded of the relevance of extension; and (iii) is able to detect that her intuitive judgments neglect extension. These conditions are least likely to hold—and complete neglect most likely to be observed—when the judge evaluates a single object and when the extension of the set is not explicitly mentioned. At the other extreme, the conditions for a positive effect of extension are all satisfied in psychologists' favorite research design: the within-subject factorial experiment, in which values of extension are crossed with the values of other variables in the design. As noted earlier, this design provides an obvious cue that the experimenter considers every manipulated variable relevant, and it enables participants to ensure that their judgments exhibit sensitivity to all these variables. The factorial design is therefore especially inappropriate for testing hypotheses about biases of neglect (Kahneman & Frederick, 2002).

In spite of these objections, within-subject factorial designs have been used in several experimental studies of extension neglect. Figure 10 illustrates the remarkably consistent additive extension effect that has emerged in these experiments (Schreiber & Kahneman, 2000). In each of the experiments, the extension variable has a slight but significant effect, and combines additively with other information. The additivity is noteworthy, because it is normatively inappropriate. For each panel of Figure 10, a compelling normative argument can be made for a quasi-multiplicative rule in which the lines should fan out.

The observed pattern is compatible with a process of anchoring and adjustment: the intuitive judgment provides an anchor, and small adjustments from that anchor are made to accommodate the role of extension.

4 The observed pattern is compatible with a process of anchoring and adjustment: the intuitive judgment provides an anchor, and small adjustments from that anchor are made to accommodate the role of extension.

Patient A

Patient B

Figure 9. Pain intensity reported by two colonoscopy patients.

Anderson (1996, p. 253) has described several other situations in which variables that should be combined multiplicatively are combined additively.
Framing & Anchoring: Experimental Design

• Within vs. Between individual design
  – Within individual is gold standard
  – But how present A’/B’ for intuitive judgement?

• Factorial designs
  – Expose each subject to one value of each variable
  – But reveals size of category space, blocking attribute substitution
Next: Neuroethics


• R. Saxe, “How we read each other’s minds.” TED talk, September, 2009. (video) Quiz questions OK


References