

Is implies ought: Are there ethical cognitive biases?

What are cognitive biases?



cognitive biases are empirically discoverable and predictable deviations from normative standards of reasoning, observable in all definable populations with remarkable impartiality



categories



- hot cognition versus cold cognition
 - hot examples: Egocentric Bias, Avoidance of cognitive dissonance
 - cold examples: framing effect, anchoring, probability neglect

categories



- individual versus group
 - group: Bandwagon Effect, Group Polarization or Risky Effect
- capacities, e.g. decision-making, social, behavioural, memory, motivation.

the conjunction fallacy



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.

framing effect



Imagine that you face the following pair of concurrent decisions. First examine both decisions, then indicate the options you prefer.

framing effect



Decision (i) Choose between:

A. a sure gain of \$240

B. 25% chance to gain \$1000 and 75% chance to gain nothing

Decision (ii) Choose between:

C. a sure loss of \$750

D. 75% chance to lose \$1000 and 25% chance to lose nothing

framing effect



Decision (i) Choose between:

A. a sure gain of \$240

B. 25% chance to gain \$1000 and 75% chance to gain nothing

(84% & 16%)

Decision (ii) Choose between:

C. a sure loss of \$750

D. 75% chance to lose \$1000 and 25% chance to lose nothing

(13% & 87%)

anchoring



Question 1.

Was Gandhi older or younger than 9 when he died?

How old was Gandhi when he died?

anchoring



Question 2.

Was Gandhi older or younger than 140
when he died?

How old was Gandhi when he died?

anchoring



Average answers:

Question 1. 50 yrs.

Question 2. 67 yrs.

The Monty Hall Problem

Monty asks you to choose between three boxes. One box contains a valuable prize, the other two boxes do not.

The Monty Hall Problem



Box A

Box B

Box C

The Monty Hall Problem

After you choose Box A, Monty reveals Box C as empty, and then asks you if you would like to switch your choice. Of the remaining two Box A and Box B, do you switch your choice?

Do you switch from A to B?



Box A

Box B

Should you switch from A to B?



Box A

Box B

Yes, you should you switch from A
to B



Box A

33%

Box B

50%

monkeys



Why?



Early explanation: heuristics

Why?



Early explanation: heuristics

Current explanation: “Dual Process Theory”

Simply put: How phenomena or tasks (outcome) might occur in two different ways (process).

Why?



“Dual Process Theory”

Daniel Kahneman: ‘Fast and Slow’

Joshua Greene: ‘manual and automatic’

Why not moral biases?



Why not moral biases?



- There ARE many already identified biases that have moral consequences both direct and indirect. (empathy gap, bandwagon effect, ostrich effect, etc.)

Why not moral biases?



- There ARE many already identified biases that have moral consequences both direct and indirect. (empathy gap, bandwagon effect, ostrich effect, etc.)
- however, a bias requires *both*:
 - a empirical (observable), predictable, and systematic error in judgement or action
 - an established and accepted *norm* or rational optimal outcome by which a judgement is considered biased.

why not moral biases?



It would seem that philosophically and psychologically, moral failings have traditionally been accounted as either pathological (psycopathy) or stemming from individual character flaws (*akrasia*, etc.)

why not moral biases?



To make a claim for a moral bias, we would expect similar requirements:

- observable
- predictable
- deficient comparable to an accepted and realizable standard
- empirically and psychologically accountable (dual process?)

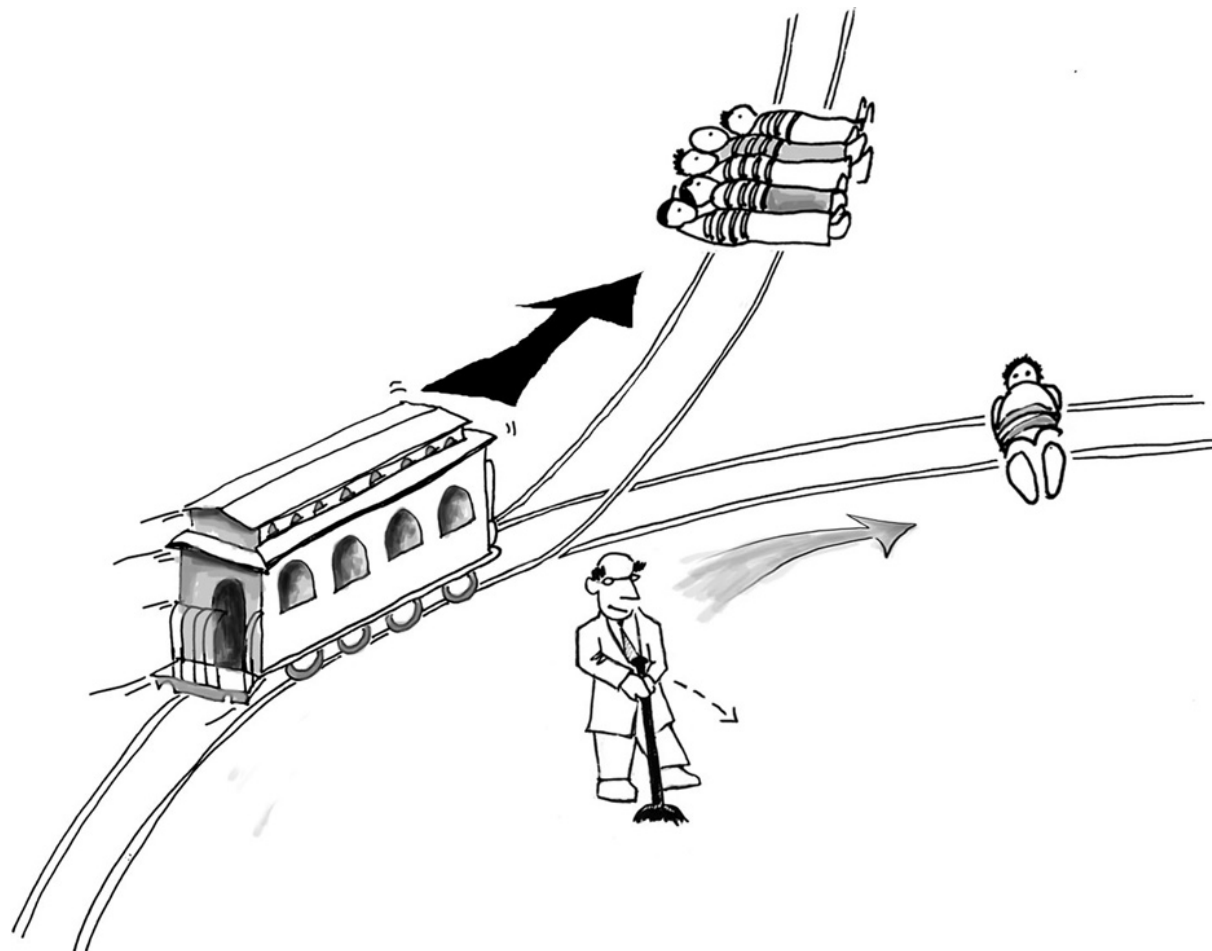


The Trolley Problem

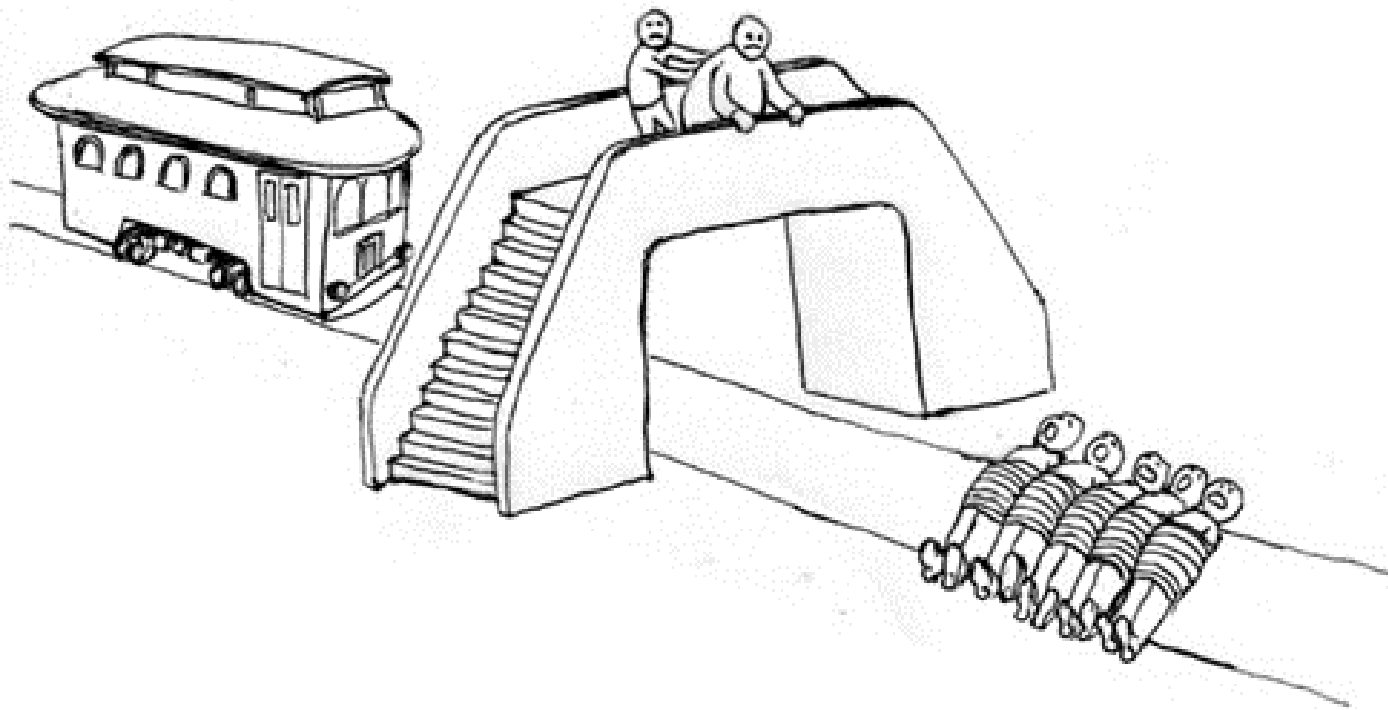


Joshua D. Greene, "The secret joke of Kant's soul", in *Moral Psychology*, 2008, Vol. 3: *The Neuroscience of Morality*, W. Sinnott-Armstrong, Ed., (Cambridge, MA: MIT Press)

The Traditional Problem



Bridge aka 'fat man' version



let's make a list



let's make a list



Historical Examples:

- Obedience to authority (Arendt, Milgram, Stanford Prison Experiment, Abu Ghraib).

let's make a list



Historical Examples:

- Obedience to authority (Arendt, Milgram, Stanford Prison Experiment, Abu Ghraib).
- bystander effect (Kitty Genovese effect)

let's make a list



Modern Examples:

- Singer's utilitarian example

The bias of physical distance



let's make a list



Modern Examples:

- Singer's utilitarian example
- Failure of folk psychology

attribution of intentionality bias



Knobe, J. (2003). "Intentional Action and Side Effects in Ordinary Language." *Analysis*, 63, 190-193.

experiment 1



The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but it will also harm the environment.'

The chairman of the board answered, 'I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program.'

They started the new program. Sure enough, the environment was harmed.

questions



- on a scale of 0 to 6, how much blame does the chairman deserve?
- did the chairman *intentionally* harm the environment?

experiment 2



The vice-president of a company went to the chairman of the board and said, 'We are thinking of starting a new program. It will help us increase profits, but it will also help the environment.'

The chairman of the board answered, 'I don't care at all about harming the environment. I just want to make as much profit as I can. Let's start the new program.'

They started the new program. Sure enough, the environment was helped.

questions



- on a scale of 0 to 6, how much praise does the chairman deserve?
- did the chairman *intentionally* help the environment?

results



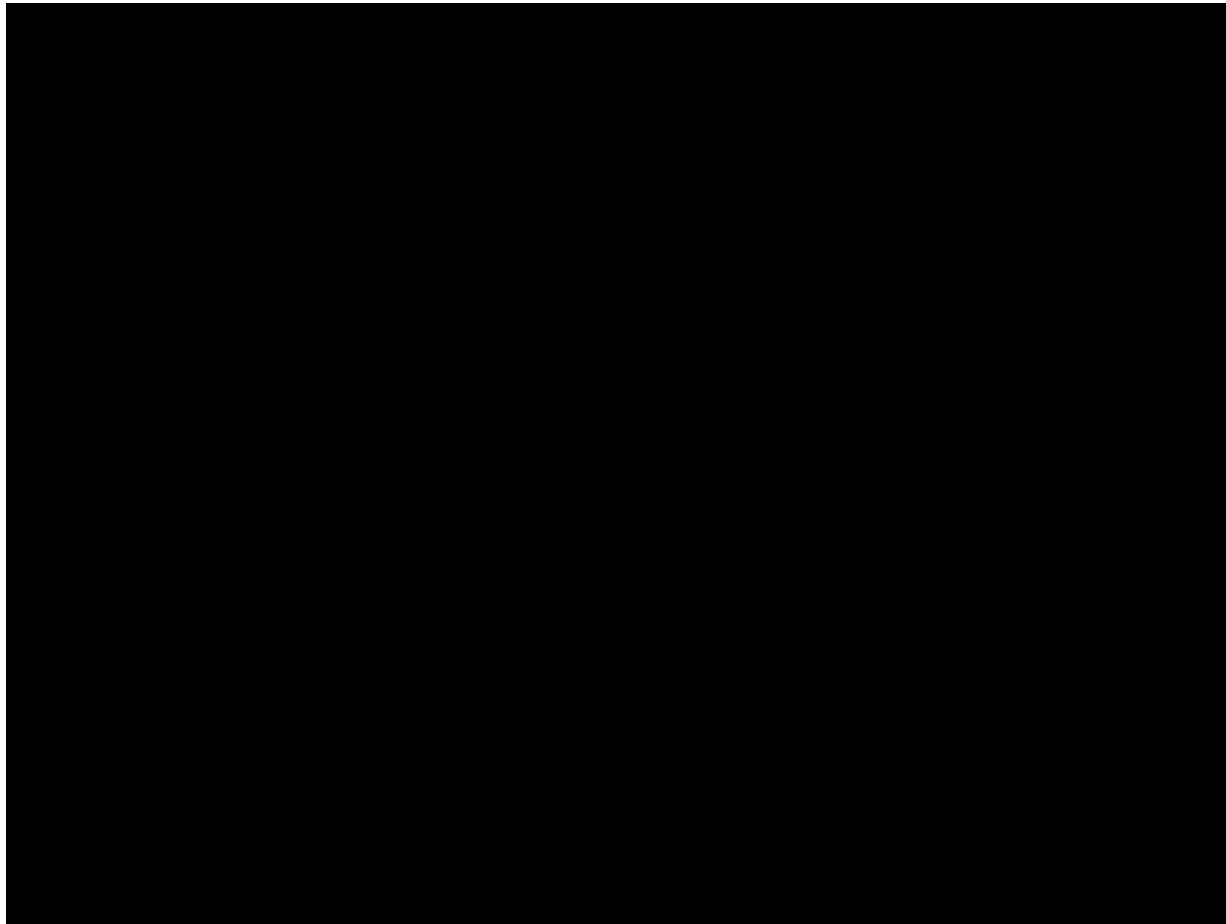
Harm

- blame
M=4.8
- 77%
intended

Help

- praise
M=1.4
- 70%
unintended

More Monkeys



mitigating bias



- healthcare challenges
- abhorrence to harm
- compassion 'fatigue'

thanks!



Brendan is bleier@ualberta.ca