

Causal selection and causal decision theory

Philosophical Foundations of Explanation

Dean McHugh

Institute of Logic, Language and Computation
University of Amsterdam

January 16, 2024



- 1 Decision theory
 - Expected value
 - Newcomb's problem

- 2 Simpson's paradox

- 1 Decision theory
 - Expected value
 - Newcomb's problem
- 2 Simpson's paradox

Decision Theory

```
graph TD; A[Decision Theory] --> B[Descriptive]; A --> C[Normative]
```

Descriptive

How do people
actually make
decisions?

Normative

How should
people make
decisions?

Decision Theory

TODAY

Descriptive

How do people
actually make
decisions?

Normative

How should
people make
decisions?



Heads : + €12
Tails : - €10

Expected Utility

Let:

- O be a set of outcomes
- $u : O \rightarrow \mathbb{R}$ (utility)
- $P : O \rightarrow [0, 1]$ (probabilities) \leftarrow probability distribution
- $A \subseteq O$ (actions)

$$eu(a) = \sum_{o \in O} u(o) p(o|a)$$



Heads: +€12

Tails: -€10

$O = \{ \text{no flip, H, T, play, decline} \}$

$A = \{ \text{play, decline} \}$



Heads: +€12

Tails: -€10



$O = \{ \text{no flip, H, T, play, decline} \}$

$A = \{ \text{play, decline} \}$

$$eu(\text{decline}) = \underset{0}{(0 \cdot 1)} + (+€12 \cdot 0) + (-€10 \cdot 0)$$

$$eu(\text{play}) = \underset{+€1}{(0 \cdot 0)} + (+€12 \cdot .5) + (-€10 \cdot .5)$$



Heads: +€12

Tails: -€10

$O = \{\text{no flip, H, T, play, decline}\}$

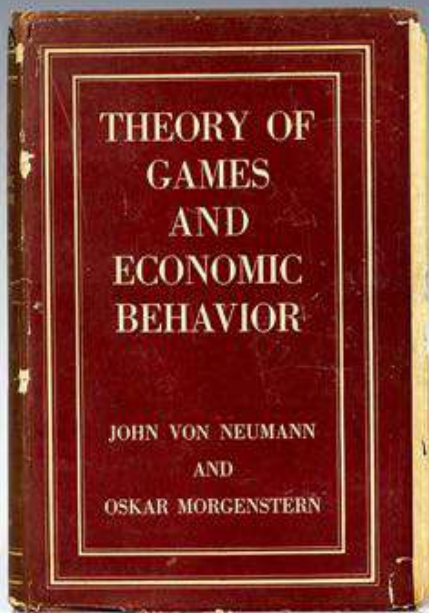
$A = \{\text{play, decline}\}$

$$eu(\text{decline}) = (0 \cdot 1) + (+€12 \cdot 0) + (-€10 \cdot 0)$$

$$eu(\text{play}) = (0 \cdot 0) + (+€12 \cdot .5) + (-€10 \cdot .5)$$

+ €1

Recommendation: PLAY!



The *expected utility* of an action is the sum, over each possible outcome, of the outcome given the action, times the utility of the outcome:

$$EU(a) = \sum_{o \in O} P(o | a)U(o)$$

The *expected utility* of an action is the sum, over each possible outcome, of the outcome given the action, times the utility of the outcome:

$$EU(a) = \sum_{o \in O} P(o | a)U(o)$$

Example (Coinflip)

I offer you a bet.

- If the coin lands tails you lose €1.
- If the coin lands heads you win €2.

Do you take the bet?

The Von Neumann–Morgenstern Utility Theorem

Suppose you have to rank lotteries, determining for each pair of lotteries A and B whether you would rather (i) play A over B (ii) play B over A , or (iii) are indifferent between them.

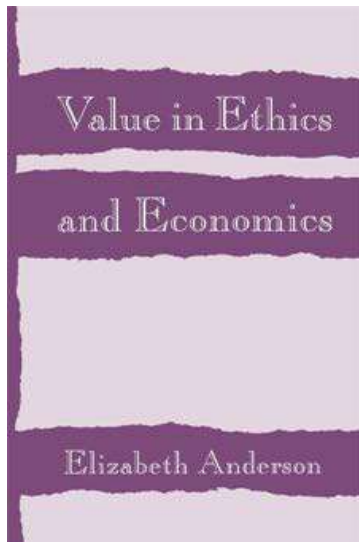
Theorem

Assume you satisfy some weak assumptions, called Completeness, Transitivity, Continuity and Independence of Irrelevant Alternatives.

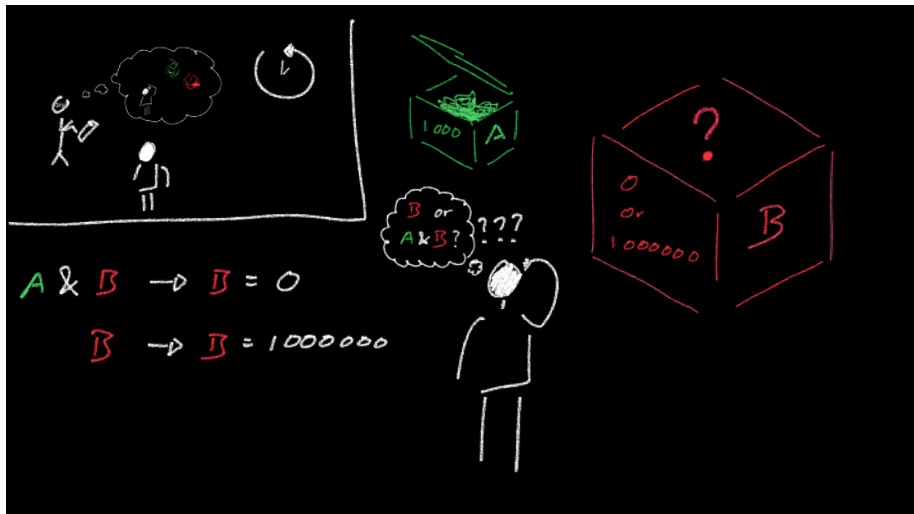
Then for any lotteries A and B , A is preferred to B just in case the expected utility of A is greater than the expected utility of B .



- We value things in different ways
- The reason why we value our loved ones is different from the reason why we value work
- Our values are not linear, that is, not always comparable



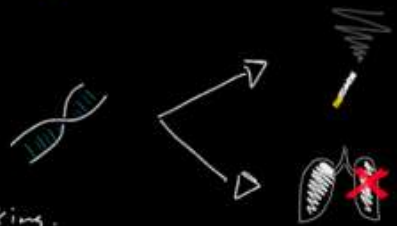
- 1 Decision theory
 - Expected value
 - Newcomb's problem
- 2 Simpson's paradox



Doctors :



Smoking lobby :



Suppose you enjoy smoking.



Alex Bellos's Adventures in Numberland

Newcomb's problem divides philosophers. Which side are you on?

Newcomb's problem has split the world of philosophy into two opposing camps. Two philosophers explain - then take the test yourself

UPDATE: Read [the poll result here](#).



The original post was viewed more than 200,000 times, with well over a thousand comments. We tallied 31,854 votes before we closed submissions. And the results are:

- I choose box B: **53.5 per cent**
- I choose both boxes: **46.5 per cent**

This is very close - an almost Brexit-like split down the middle of the voting public. And like Brexit, some families were deeply divided. Mine was anyway: I chose box B, while my wife chose both boxes. Arguing about it just entrenched our positions.

In the **only other mass survey on Newcomb's Problem**, the results were similar: 55 per cent chose box B, and 45 per cent both boxes.

So many excellent comments were made on the original piece that I will not attempt to summarise them. I do, however, want to add some background. The question is named after William A Newcomb, an American physicist who thought it up when thinking about another famous problem, the **prisoner's dilemma**. But it was only when the American philosopher Robert Nozick wrote about the puzzle in 1970 that it became well known.

1. Prodrome

One or two days before a migraine, you might notice subtle changes that warn of an upcoming migraine, including constipation, mood changes from depression to euphoria, food cravings, neck stiffness, increased thirst and urination or frequent yawning.

Evidential decision theory:

$$EU(a) = \sum_{o \in O} U(o)P(o | a)$$

Causal decision theory:

$$EU(a) = \sum_{o \in O} U(o)P(a > o)$$

Predestination as a Newcomb's Problem

Resnik (1987) argues that Calvinists who believe in divine predestination effectively face a real-life Newcomb Problem.

They believe their actions on earth have no causal influence over whether they go to heaven or hell, but doing good deeds give them evidence that they are going to heaven.

Evidential decision theory: Do good deeds!

Causal decision theory: Your fate is sealed, sin all you want!

XOR Blackmail

An agent has been alerted to a rumor that her house has a terrible termite infestation, which would cost her \$1,000,000 in damages. She does not know whether this rumor is true. A greedy and accurate predictor with a strong reputation for honesty has learned whether or not it is true, and drafts a letter:

“I know whether or not you have termites, and I have sent you this letter iff exactly one of the following is true: (i) the rumor is false, and you are going to pay me \$1,000 upon receiving this letter; or (ii) the rumor is true, and you will not pay me upon receiving this letter.”

The predictor then predicts what the agent would do upon receiving the letter, and sends the agent the letter iff exactly one of (i) or (ii) is true. Thus, the claim made by the letter is true. Assume the agent receives the letter.

Should she pay up?

The Twin Prisoner's Dilemma

An agent has the option of cooperating with or defecting against her psychological twin, who is in a separate room, faces the same options, and has the same state of knowledge. Both rank the outcomes the same way, where “I defect and she cooperates” $>$ “We both cooperate” $>$ “We both defect” $>$ “I cooperate and she defects.”

Should the agent defect or cooperate?

- 1 Decision theory
 - Expected value
 - Newcomb's problem

- 2 Simpson's paradox

Simpson (1951) observed that a trend present in every subgroup may be reversed in the entire population.

	Population 1	Population 2
Drug	✓ (100%)	✓XXX (25%)
No drug	✓✓✓X (25%)	X (0%)

Table: The segregated data

Simpson (1951) observed that a trend present in every subgroup may be reversed in the entire population.

	Population 1	Population 2
Drug	✓ (100%)	✓XXX (25%)
No drug	✓✓✓X (25%)	X (0%)

Table: The segregated data

	Population 1 + 2
Drug	✓✓XXX (40%)
No drug	✓✓✓XX (60%)

Table: The aggregated data

Example taken from
<https://www.youtube.com/watch?v=ebEkn-BiW5k>

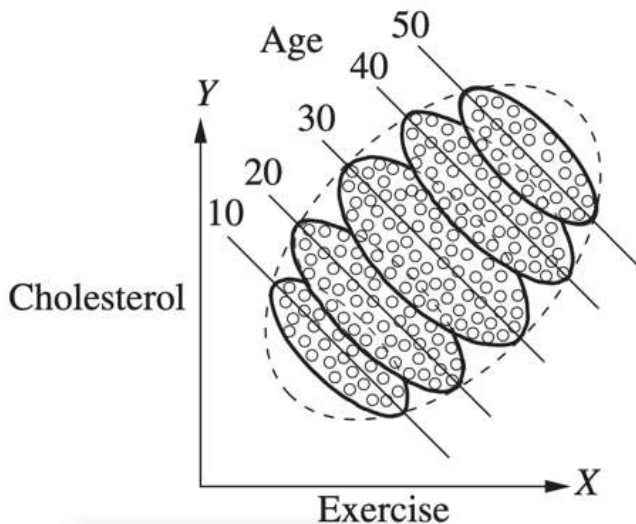


Figure: Segregated data

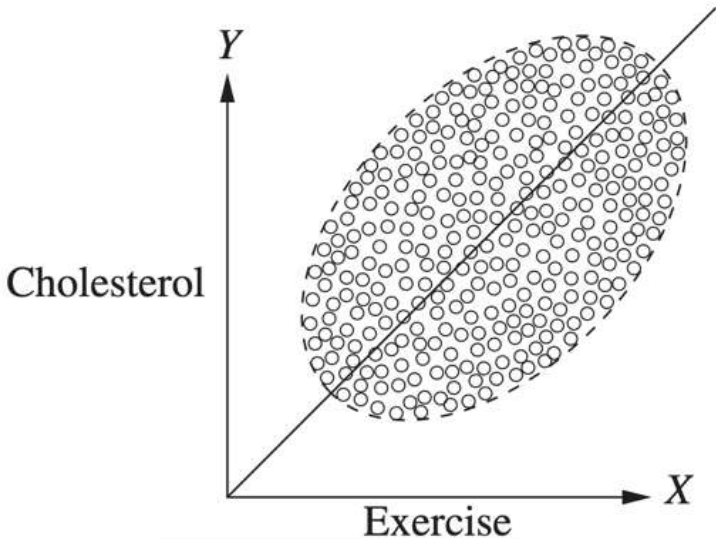


Figure: Aggregate data

Table 1.1 Results of a study into a new drug, with gender being taken into account

	Drug	No drug
Men	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
Women	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

suppose we knew an additional fact: Estrogen has a negative effect on recovery, so women are less likely to recover than men, regardless of the drug. In addition, as we can see from the data, women are significantly more likely to take the drug than men are. So, the reason the drug appears to be harmful overall is that, if we select a drug user at random, that person is more likely to be a woman and hence less likely to recover than a random person who does not take the drug. Put differently, being a woman is a common cause of both drug taking and failure to recover. Therefore, to assess the effectiveness, we need to compare subjects of the same gender, thereby ensuring that any difference in recovery rates between those who take the drug and those who do not is not ascribable to estrogen. This means we should consult the segregated data, which shows us unequivocally that the drug is helpful.

(Pearl, Glymour, and Jewell 2016, p. 3)

Table 1.2 Results of a study into a new drug, with posttreatment blood pressure taken into account

	No drug	Drug
Low BP	81 out of 87 recovered (93%)	234 out of 270 recovered (87%)
High BP	192 out of 263 recovered (73%)	55 out of 80 recovered (69%)
Combined data	273 out of 350 recovered (78%)	289 out of 350 recovered (83%)

“Now, would you recommend the drug to a patient? Once again, the answer follows from the way the data were generated. In the general population, the drug might improve recovery rates because of its effect on blood pressure. But in the subpopulations—the group of people whose posttreatment BP is high and the group whose posttreatment BP is low—we, of course, would not see that effect; we would only see the drug’s toxic effect.

(Pearl, Glymour, and Jewell 2016, pp. 4–5)

As in the gender example, the purpose of the experiment was to gauge the overall effect of treatment on rates of recovery. But in this example, since lowering blood pressure is one of the mechanisms by which treatment affects recovery, it makes no sense to separate the results based on blood pressure. (If we had recorded the patients' blood pressure before treatment, and if it were BP that had an effect on treatment, rather than the other way around, it would be a different story.) So we consult the results for the general population, we find that treatment increases the probability of recovery, and we decide that we should recommend treatment. Remarkably, though the numbers are the same in the gender and blood pressure examples, the correct result lies in the segregated data for the former and the aggregate data for the latter."

(Pearl, Glymour, and Jewell 2016, pp. 4–5)

Let $Y_{X=x}$ be the value that variable Y takes after an intervention to set $X = x$. That is, $Y_{X=x} = y$ in a model M and context u just in case $M, u \models [X \leftarrow x]Y = y$.

Let D , R and Age be binary variables taken from a hospital database, with $D = 1$ representing the patient taking a given drug, $R = 1$ representing recovery from illness, and $A = a$ representing the patient's age.

Fact

If, for every age a , we have $P(A = a) > 0$ and

$$P(R_{D=1} = 1 \mid A = a) > P(R_{D=0} = 1 \mid A = a)$$

then

$$P(R_{D=1} = 1) > P(R_{D=0} = 1).$$

Simpson's paradox shows that this does **not** hold if we replace $P(R_{D=d} = 1 \mid A = a)$ with $P(R = 1 \mid A = a, D = d)$.

Fact

If, for every age a , we have $P(A = a) > 0$ and




$$P(R_{D=1} = 1 \mid A = a) > P(R_{D=0} = 1 \mid A = a)$$

then

$$P(R_{D=1} = 1) > P(R_{D=0} = 1).$$

Proof.

$$\begin{aligned} P(R_{D=1} = 1) &= \sum_a P(R_{D=1} = 1, A = a) && \text{(Marginalise over age)} \\ &= \sum_a P(R_{D=1} = 1 \mid A = a)P(A = a) && \text{(Bayes' rule)} \\ &> \sum_a P(R_{D=0} = 1 \mid A = a)P(A = a) && \text{(Above)} \\ &= \sum_a P(R_{D=0} = 1, A = a) && \text{(Bayes' rule)} \\ &= P(R_{D=0} = 1) \end{aligned}$$

-  Pearl, Judea, Madelyn Glymour, and Nicholas P Jewell (2016). *Causal inference in statistics: A primer*. John Wiley & Sons.
-  Resnik, Michael D (1987). *Choices: An introduction to decision theory*. University of Minnesota Press.
-  Simpson, Edward H (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society: Series B (Methodological)* 13.2, pp. 238–241.