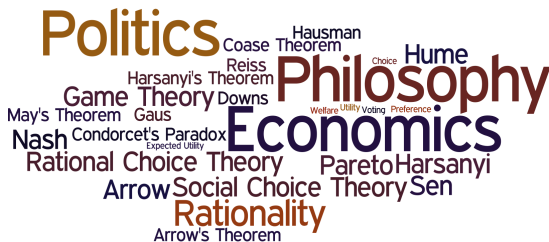
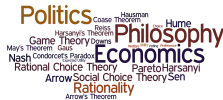


Methods in Philosophy, Politics and Economics: Individual and Group Decision Making

Eric Pacuit
University of Maryland
pacuit.org

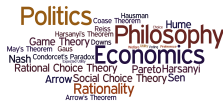


Two issues



- ✓ Utility is unique up to linear transformations
- ✓ Probabilities depends, in part, on the description of the problem

Three issues



- ✓ Utility is unique up to linear transformations
- ✓ Probabilities depends, in part, on the description of the problem
 - ▶ The probability of states are *independent* of the chosen act

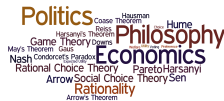
3 / 31

Dominance Reasoning



	w_1	w_2
A	1	3
B	2	4

Dominance Reasoning



Dominance reasoning is appropriate only when probability of outcome is *independent of choice*.

(A nasty nephew wants inheritance from his rich Aunt. The nephew wants the inheritance, but other things being equal, does not want to apologize. Does dominance give the nephew a reason to not apologize? *Whether or not the nephew is cut from the will may depend on whether or not he apologizes.*)



A



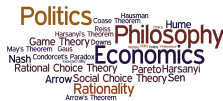
B

Choice:

one-box: choose box *B*

two-box: choose box *A* and *B*

Newcomb's Paradox



A very powerful being, who has been invariably accurate in his predictions about your behavior in the past, has already acted in the following way:

1. If he has predicted that you will open just box *B*, he has in addition put \$1,000,000 in box *B*
2. If he has predicted you will open both boxes, he has put nothing in box *B*.

What should you do?

R. Nozick. *Newcomb's Problem and Two Principles of Choice*. 1969.

	\$1 million in closed box	\$0 in closed box
one-box	\$1,000,000	\$0
two- box	\$1,001,000	\$1,000

	\$1 million in closed box	\$0 in closed box
one-box	\$1,000,000	\$0
two- box	\$1,001,000	\$1,000

act-state dependence: $P(s) \neq P(s \mid A)$

Newcomb's Paradox

Politics
Philosophy
Economics
Rationality
Arrow's Theorem
Arrow
Social Choice
Theory
Sen
Pareto
Harsanyi
Rational Choice
Theory
Condorcet's Paradox
Nash
May's Theorem
Gaus
Game Theory
Downs
Harsanyi's Theorem
Roths
Hume
Hausman
Coase
Theorem

	$B = 1M$	$B = 0$
1 Box	1M	0
2 Boxes	$1M + 1000$	1000



Newcomb's Paradox

	$B = 1M$	$B = 0$
1 Box	1M	0
2 Boxes	$1M + 1000$	1000

	$B = 1M$	$B = 0$
1 Box	h	$1 - h$
2 Boxes	$1 - h$	h

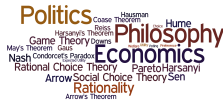


Newcomb's Paradox



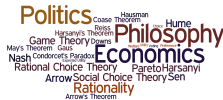
J. Collins. *Newcomb's Problem*. International Encyclopedia of Social and Behavioral Sciences, 1999.

Newcomb's Paradox



There is a conflict between maximizing your expected value (1-box choice) and dominance reasoning (2-box choice).

Newcomb's Paradox



There is a conflict between maximizing your expected value (1-box choice) and dominance reasoning (2-box choice).

What the Predictor did yesterday is *probabilistically dependent* on the choice today, but *causally independent* of today's choice.

$$V(A) = \sum_w V(w) \cdot P_A(w)$$

(the expected value of act A is a probability weighted average of the values of the ways w in which A might turn out to be true)

$$V(A) = \sum_w V(w) \cdot P_A(w)$$

(the expected value of act A is a probability weighted average of the values of the ways w in which A might turn out to be true)

EDT: $P_A(w) := P(w \mid A)$ (Probability of w given A is chosen)

CDT: $P_A(w) = P(A \Box \rightarrow w)$ (Probability of *if A were chosen then w would be true*)

Suppose 99% confidence in predictors reliability.

B_1 : one-box (open box B)

B_2 : two-box choice (open both A and B)

N : receive nothing

K : receive \$1,000

M : receive \$1,000,000

L : receive \$1,001,000

Suppose 99% confidence in predictors reliability.

B_1 : one-box (open box B)

B_2 : two-box choice (open both A and B)

N : receive nothing

K : receive \$1,000

M : receive \$1,000,000

L : receive \$1,001,000

$$V(B_1) = V(M)P(M \mid B_1) + V(N)P(N \mid B_1)$$

Suppose 99% confidence in predictors reliability.

B_1 : one-box (open box B)

B_2 : two-box choice (open both A and B)

N : receive nothing

K : receive \$1,000

M : receive \$1,000,000

L : receive \$1,001,000

$$V(B_1) = V(M)P(M \mid B_1) + V(N)P(N \mid B_1) = 1000000 \cdot 0.99 + 0 \cdot 0.01$$

Suppose 99% confidence in predictors reliability.

B_1 : one-box (open box B)

B_2 : two-box choice (open both A and B)

N : receive nothing

K : receive \$1,000

M : receive \$1,000,000

L : receive \$1,001,000

$$V(B_1) = V(M)P(M \mid B_1) + V(N)P(N \mid B_1) = 1000000 \cdot 0.99 + 0 \cdot 0.01 = 990,000$$

Suppose 99% confidence in predictors reliability.

B_1 : one-box (open box B)

B_2 : two-box choice (open both A and B)

N : receive nothing

K : receive \$1,000

M : receive \$1,000,000

L : receive \$1,001,000

$$V(B_1) = V(M)P(M | B_1) + V(N)P(N | B_1) = 1000000 \cdot 0.99 + 0 \cdot 0.01 = 990,000$$

$$V(B_2) = V(L)P(L | B_2) + V(K)P(K | B_2)$$

Suppose 99% confidence in predictors reliability.

B_1 : one-box (open box B)

B_2 : two-box choice (open both A and B)

N : receive nothing

K : receive \$1,000

M : receive \$1,000,000

L : receive \$1,001,000

$$V(B_1) = V(M)P(M | B_1) + V(N)P(N | B_1) = 1000000 \cdot 0.99 + 0 \cdot 0.01 = 990,000$$

$$V(B_2) = V(L)P(L | B_2) + V(K)P(K | B_2) = 1001000 \cdot 0.01 + 1000 \cdot 0.99$$

Suppose 99% confidence in predictors reliability.

B_1 : one-box (open box B)

B_2 : two-box choice (open both A and B)

N : receive nothing

K : receive \$1,000

M : receive \$1,000,000

L : receive \$1,001,000

$$V(B_1) = V(M)P(M | B_1) + V(N)P(N | B_1) = 1000000 \cdot 0.99 + 0 \cdot 0.01 = 990,000$$

$$V(B_2) = V(L)P(L | B_2) + V(K)P(K | B_2) = 1001000 \cdot 0.01 + 1000 \cdot 0.99 = 11,000$$

Let μ be the assigned to the conditional $B_1 \square \rightarrow M$ (and $B_2 \square \rightarrow L$) (both conditionals are true iff the Predictor put \$1,000,000 in box B yesterday).

B_1 : one-box (open box B)

B_2 : two-box choice (open both A and B)

N : receive nothing

K : receive \$1,000

M : receive \$1,000,000

L : receive \$1,001,000

Let μ be the assigned to the conditional $B_1 \square \rightarrow M$ (and $B_2 \square \rightarrow L$) (both conditionals are true iff the Predictor put \$1,000,000 in box B yesterday).

B_1 : one-box (open box B)

B_2 : two-box choice (open both A and B)

N : receive nothing

K : receive \$1,000

M : receive \$1,000,000

L : receive \$1,001,000

$$V(B_1) = V(M)P(B_1 \square \rightarrow M) + V(N)P(B_1 \square \rightarrow N)$$

Let μ be the assigned to the conditional $B_1 \Box \rightarrow M$ (and $B_2 \Box \rightarrow L$) (both conditionals are true iff the Predictor put \$1,000,000 in box B yesterday).

B_1 : one-box (open box B)

B_2 : two-box choice (open both A and B)

N : receive nothing

K : receive \$1,000

M : receive \$1,000,000

L : receive \$1,001,000

$$V(B_1) = V(M)P(B_1 \Box \rightarrow M) + V(N)P(B_1 \Box \rightarrow N) = 1000000 \cdot \mu + 0 \cdot (1 - \mu)$$

Let μ be the assigned to the conditional $B_1 \square \rightarrow M$ (and $B_2 \square \rightarrow L$) (both conditionals are true iff the Predictor put \$1,000,000 in box B yesterday).

B_1 : one-box (open box B)

B_2 : two-box choice (open both A and B)

N : receive nothing

K : receive \$1,000

M : receive \$1,000,000

L : receive \$1,001,000

$$V(B_1) = V(M)P(B_1 \square \rightarrow M) + V(N)P(B_1 \square \rightarrow N) = 1000000 \cdot \mu + 0 \cdot (1 - \mu) = 1000000\mu$$

Let μ be the assigned to the conditional $B_1 \square \rightarrow M$ (and $B_2 \square \rightarrow L$) (both conditionals are true iff the Predictor put \$1,000,000 in box B yesterday).

B_1 : one-box (open box B)

B_2 : two-box choice (open both A and B)

N : receive nothing

K : receive \$1,000

M : receive \$1,000,000

L : receive \$1,001,000

$$V(B_1) = V(M)P(B_1 \square \rightarrow M) + V(N)P(B_1 \square \rightarrow N) = 1000000 \cdot \mu + 0 \cdot (1 - \mu) = 1000000\mu$$

$$V(B_2) = V(L)P(B_2 \square \rightarrow L) + V(K)P(B_2 \square \rightarrow K)$$

Let μ be the assigned to the conditional $B_1 \square \rightarrow M$ (and $B_2 \square \rightarrow L$) (both conditionals are true iff the Predictor put \$1,000,000 in box B yesterday).

B_1 : one-box (open box B)

B_2 : two-box choice (open both A and B)

N : receive nothing

K : receive \$1,000

M : receive \$1,000,000

L : receive \$1,001,000

$$V(B_1) = V(M)P(B_1 \square \rightarrow M) + V(N)P(B_1 \square \rightarrow N) = 1000000 \cdot \mu + 0 \cdot (1 - \mu) = 1000000\mu$$

$$V(B_2) = V(L)P(B_2 \square \rightarrow L) + V(K)P(B_2 \square \rightarrow K) = 1001000 \cdot \mu + 1000 \cdot (1 - \mu)$$

Let μ be the assigned to the conditional $B_1 \square \rightarrow M$ (and $B_2 \square \rightarrow L$) (both conditionals are true iff the Predictor put \$1,000,000 in box B yesterday).

B_1 : one-box (open box B)

B_2 : two-box choice (open both A and B)

N : receive nothing

K : receive \$1,000

M : receive \$1,000,000

L : receive \$1,001,000

$$V(B_1) = V(M)P(B_1 \square \rightarrow M) + V(N)P(B_1 \square \rightarrow N) = 1000000 \cdot \mu + 0 \cdot (1 - \mu) = 1000000\mu$$

$$V(B_2) = V(L)P(B_2 \square \rightarrow L) + V(K)P(B_2 \square \rightarrow K) = 1001000 \cdot \mu + 1000 \cdot (1 - \mu) = 1000000\mu + 1000$$

Causal Decision Theory



A. Egan. *Some Counterexamples to Causal Decision Theory*. *Philosophical Review*, 116(1), pgs. 93 - 114, 2007.

The Psychopath Button: Paul is debating whether to press the 'kill all psychopaths' button. It would, he thinks, be much better to live in a world with no psychopaths.

The Psychopath Button: Paul is debating whether to press the 'kill all psychopaths' button. It would, he thinks, be much better to live in a world with no psychopaths. Unfortunately, Paul is quite confident that only a psychopath would press such a button.

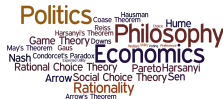
The Psychopath Button: Paul is debating whether to press the 'kill all psychopaths' button. It would, he thinks, be much better to live in a world with no psychopaths. Unfortunately, Paul is quite confident that only a psychopath would press such a button. Paul very strongly prefers living in a world with psychopaths to dying. Should Paul press the button?

(Set aside your theoretical commitments and put yourself in Paul's situation. Would you press the button? Would you take yourself to be irrational for not doing so?)

- ▶ The crucial distinction is between an act and a decision to perform the act.
- ▶ Before performing an act, an agent may assess the act in light of a decision to perform it. Information the decision carries may affect the act's expected utility and its ranking with respect to other acts.
- ▶ Decision makers should make *self-ratifying*, or *ratifiable*, decisions.

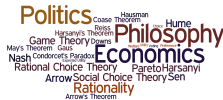
H. Gaifman. *Self-reference and the acyclicity of rational choice*. Annals of Pure and Applied Logic, 96, pgs. 117 - 140, 1999.

The Irrational Choice



Mr. Z offers Adam two boxes, each containing \$10. Adam can choose either S1: to take the leftmost box and get \$10, or S2: to take the two boxes and get \$20.

The Irrational Choice

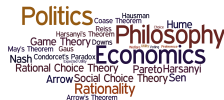


Mr. Z offers Adam two boxes, each containing \$10. Adam can choose either S1: to take the leftmost box and get \$10, or S2: to take the two boxes and get \$20. Before making his decision, Adam is informed by Mr. Z that if he acts irrationally, Mr. Z will give him a bonus of \$100.

(...to eliminate noise factors, assume that Adam believes that Mr. Z is serious, has the relevant knowledge, is a perfect reasoner and is completely trustworthy.)

Framing Effects

Schelling's Example



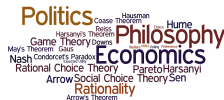
Suppose your tax depends on your income and how many kids you have.

- The “child deduction” might be, say, 1000 per child:

$$Tax(i, k) = Base(i) - [\max(k, 3) \cdot 1000]$$

Q1: Should the child deduction be larger for the rich than for the poor?

Schelling's Example



Instead of taking the “standard” household to be childless, we could lower the base tax for everyone (e.g., by 3000), and add a surcharge for households with less than 3 kids (e.g., 1000/2000/3000).

We could also let the surcharge depend on income.

$$Tax(i, k) = LowerBase(i) + [(3 - k) \cdot Surcharge(i)]$$

Q2: Should the childless poor pay as large a surcharge as the childless rich?

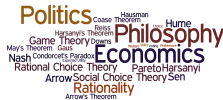
Schelling's Example



Q1: Should the child exemption be larger for the rich than for the poor?

Q2: Should the childless poor pay as large a surcharge as the childless rich?

Schelling's Example



Q1: Should the child exemption be larger for the rich than for the poor?

Q2: Should the childless poor pay as large a surcharge as the childless rich?

If you answered “No” to both, then you are not endorsing a coherent policy

As Kahneman puts the point...

“The difference between the tax owed by a childless family and by a family with two children can be described as a reduction or as an increase. If you want the poor to receive at least the same benefit as the rich for having children, then you must want the poor to pay at least the same penalty as the rich for being childless. ”

“The message about the nature of framing is stark: framing should not be viewed as an intervention that masks or distorts an underlying preference. At least in this instance...there is no underlying preference that is masked or distorted by the frame. Our preferences are about framed problems, and our moral intuitions are about descriptions, not substance.”

reference dependence: people derive utility from *gains and losses*, measured relative to some reference point, rather than from absolute levels of wealth. Leads to phenomena such as the **endowment effect**.

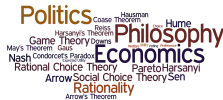
loss aversion: people are much more sensitive to losses—even small losses—than to gains of the same magnitude. Many people turn down a gamble ($-\$100 : \frac{1}{2}, \$110 : \frac{1}{2}$), but this is very hard to explain in classical utility theory (Rabin, 2000)

diminishing sensitivity: people tend to be risk averse over moderate probability gains (they typically prefer a certain gain of \$500 to a 50 percent chance of \$1,000) and *risk seeking* over losses (they prefer a 50 percent chance of loosing \$1000 to loosing \$500 for sure)

diminishing sensitivity: people tend to be risk averse over moderate probability gains (they typically prefer a certain gain of \$500 to a 50 percent chance of \$1,000) and *risk seeking* over losses (they prefer a 50 percent chance of loosing \$1000 to loosing \$500 for sure)

probability weighting: people tend to overweight the tails of a probability distribution (they tend to overweight extremely unlikely outcomes).

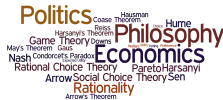
Readings On Prospect Theory



D. Kahneman and A. Tversky. *Prospect Theory: An Analysis of Decision under Risk*. *Econometrica*, Vol. 47, No. 2., pgs. . 263 - 292, 1979.

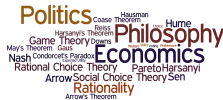
N. Barberis. *Thirty Years of Prospect Theory in Economics: A Review and Assessment*. *Journal of Economic Perspectives*, 27:1, pgs. 171 - 196, 2013.

Evaluating Rational Choice Axioms



What should we make of the patterns found by psychologists and behavioral economists? Are these descriptive issues relevant for decision theory or rational choice theory?

Evaluating Rational Choice Axioms

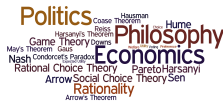


What should we make of the patterns found by psychologists and behavioral economists? Are these descriptive issues relevant for decision theory or rational choice theory?

Any apparent violation of an axiom of the theory can always be interpreted in three different ways:

1. the subjects' preferences *genuinely* violate the axioms of the theory;

Evaluating Rational Choice Axioms

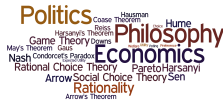


What should we make of the patterns found by psychologists and behavioral economists? Are these descriptive issues relevant for decision theory or rational choice theory?

Any apparent violation of an axiom of the theory can always be interpreted in three different ways:

1. the subjects' preferences *genuinely* violate the axioms of the theory;
2. the subjects' preferences have changed during the course of the experiment;

Evaluating Rational Choice Axioms

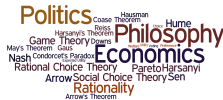


What should we make of the patterns found by psychologists and behavioral economists? Are these descriptive issues relevant for decision theory or rational choice theory?

Any apparent violation of an axiom of the theory can always be interpreted in three different ways:

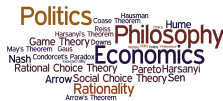
1. the subjects' preferences *genuinely* violate the axioms of the theory;
2. the subjects' preferences have changed during the course of the experiment;
3. the experimenter has overlooked a relevant feature of the context that affects the the subjects' preferences.

Aim of rational choice theory



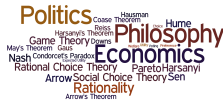
- ▶ Recommendation
- ▶ Prediction
- ▶ Explanation

Recommending Behavior



- On the one hand, the fact that many people have faulty reasoning about probabilities or deviate from EU theory does not mean that the theories are wrong (Hume's Law: *is* **does not** imply *can*). It could simply be that people are not naturally good at all kinds of reasoning, which is part of the reason why we study rational choice in the first place.

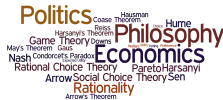
Recommending Behavior



- ▶ On the one hand, the fact that many people have faulty reasoning about probabilities or deviate from EU theory does not mean that the theories are wrong (Hume's Law: *is* **does not** imply *can*). It could simply be that people are not naturally good at all kinds of reasoning, which is part of the reason why we study rational choice in the first place.
- ▶ On the other hand, *ought* **does** imply *can* (from Kant), meaning that if we're going to say that people should follow EU theory, it needs to be possible that they actually do so.

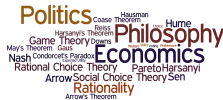
- 28 / 31

Explaining/Predicting Behavior



Stability: Individuals' preferences are stable over the period of the investigation.

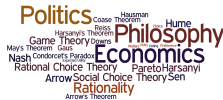
Explaining/Predicting Behavior



Stability: Individuals' preferences are stable over the period of the investigation.

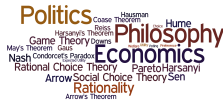
Invariance: Individuals' preferences are invariant to irrelevant changes in the context of making the decision.

Explaining/Predicting Behavior



“This shows, I think, that utility theory is a way to formalize and model rational action, but is not itself a complete **theory** of rational action.

Explaining/Predicting Behavior



“This shows, I think, that utility theory is a way to formalize and model rational action, but is not itself a complete **theory** of rational action. To employ utility theory presupposes that we know which are the relevant, and which are the irrelevant , features for evaluating states of affairs.

Explaining/Predicting Behavior



“This shows, I think, that utility theory is a way to formalize and model rational action, but is not itself a complete **theory** of rational action. To employ utility theory presupposes that we know which are the relevant, and which are the irrelevant , features for evaluating states of affairs. Unless we possess such a criteria we cannot distinguish framing effects from redescribing the world in such a way that we call attention to an important feature.

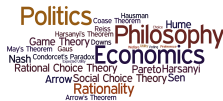
Explaining/Predicting Behavior



“This shows, I think, that utility theory is a way to formalize and model rational action, but is not itself a complete **theory** of rational action. To employ utility theory presupposes that we know which are the relevant, and which are the irrelevant, features for evaluating states of affairs. Unless we possess such a criteria we cannot distinguish framing effects from redescribing the world in such a way that we call attention to an important feature. However, only a value and/or moral theory can allow us to do that; utility theory does not imply any specific value or moral theory” (Gaus, p65).

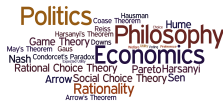
31 / 31

A Dilemma



Either stick to the “formal axioms” of completeness, transitivity, Independence, etc. and refuse to assume the principles of stability and invariance. But then rational choice theory will be useless for all explanatory and predictive purposes because people could have fully rational preferences that constantly change or are immensely context-dependent.

A Dilemma



Either stick to the “formal axioms” of completeness, transitivity, Independence, etc. and refuse to assume the principles of stability and invariance. But then rational choice theory will be useless for all explanatory and predictive purposes because people could have fully rational preferences that constantly change or are immensely context-dependent. Alternatively, an economists can assume stability and invariance but only at the expense of making rational-choice theory a substantive theory, a theory laden not just with values but with *the economist's* values.