Rationality

Lecture 12

Eric Pacuit

Center for Logic and Philosophy of Science
Tilburg University
ai.stanford.edu/∼epacuit
e.j.pacuit@uvt.nl

November 18, 2010

Instrumental Rationality

Instrumental Rationality: Ann's action α is instrumentally rational iff Ann chooses α because she soundly believes it is the best prospect to achieve her goals, desires, tastes, etc.

Cardinal Utility Theory

$x \succ y \succ z$ is represented by both $(3, 2, 1)$ and $(1000, 999, 1)$, so cannot say y is "closer" to x than to z.

Cardinal Utility Theory

$x \succ y \succ z$ is represented by both $(3, 2, 1)$ and $(1000, 999, 1)$, so cannot say y is "closer" to x than to z.

Key idea: Ordinal preferences over *lotteries* allows us to infer a cardinal scale (with some additional axioms).

John von Neumann and Oskar Morgenstern. *The Theory of Games and Economic Behavior*. Princeton University Press, 1944.

Axioms of Cardinal Utility

Suppose that X is a set of outcomes and consider **lotteries over** X (i.e., probability distributions over X)

Axioms of Cardinal Utility

Suppose that X is a set of outcomes and consider **lotteries over** X (i.e., probability distributions over X)

A **compound lottery** is $\alpha L + (1-\alpha)L'$ meaning "play lottery L with probability α and L' with probability $1-\alpha$"

Axioms of Cardinal Utility

Suppose that X is a set of outcomes and consider **lotteries over** X (i.e., probability distributions over X)

A **compound lottery** is $\alpha L + (1-\alpha)L'$ meaning "play lottery L with probability α and L' with probability $1-\alpha$"

Running example: Suppose Ann prefers pizza (p) over taco (t) over yogurt (y) ($p \succ t \succ y$) and consider the different lotteries where the prizes are p, t and y.

Cardinal Utility Theory: Continuity

Continuity: for all options x, y and z if $x \succeq y \succeq z$, there is some lottery L with probability p of getting x and $(1 - p)$ of getting z such that the agent is indifferent between L and y.

Cardinal Utility Theory: Continuity

Continuity: for all options x, y and z if $x \succeq y \succeq z$, there is some lottery L with probability p of getting x and $(1 - p)$ of getting z such that the agent is indifferent between L and y.

Suppose Ann has t.

Cardinal Utility Theory: Continuity

Continuity: for all options x, y and z if $x \succeq y \succeq z$, there is some lottery L with probability p of getting x and $(1 - p)$ of getting z such that the agent is indifferent between L and y.

Suppose Ann has t.

Consider the lottery $L = 0.99$ get y and 0.01 get p

Cardinal Utility Theory: Continuity

Continuity: for all options x, y and z if $x \succeq y \succeq z$, there is some lottery L with probability p of getting x and $(1 - p)$ of getting z such that the agent is indifferent between L and y.

Suppose Ann has t.

Consider the lottery $L = 0.99$ get y and 0.01 get p
Would Ann trade t for L?

Cardinal Utility Theory: Continuity

Continuity: for all options x, y and z if $x \succeq y \succeq z$, there is some lottery L with probability p of getting x and $(1-p)$ of getting z such that the agent is indifferent between L and y.

Suppose Ann has t.

Consider the lottery $L = 0.99$ get y and 0.01 get p
Would Ann trade t for L?

Consider the lottery $L' = 0.99$ get p and 0.01 get y

Cardinal Utility Theory: Continuity

Continuity: for all options x, y and z if $x \succeq y \succeq z$, there is some lottery L with probability p of getting x and $(1 - p)$ of getting z such that the agent is indifferent between L and y.

Suppose Ann has t.

Consider the lottery $L = 0.99$ get y and 0.01 get p
Would Ann trade t for L?

Consider the lottery $L' = 0.99$ get p and 0.01 get y
Would Ann trade t for L'?

Cardinal Utility Theory: Continuity

Continuity: for all options x, y and z if $x \succeq y \succeq z$, there is some lottery L with probability p of getting x and $(1-p)$ of getting z such that the agent is indifferent between L and y.

Suppose Ann has t.

Consider the lottery $L = 0.99$ get y and 0.01 get p
Would Ann trade t for L?

Consider the lottery $L' = 0.99$ get p and 0.01 get y
Would Ann trade t for L'?

Continuity says that there is must be some lottery where Ann is indifferent between keeping t and playing the lottery.

Cardinal Utility Theory: Better Prizes

Better Prizes: suppose L_1 is a lottery over (w, x) and L_2 is over (y, z) suppose that L_1 and L_2 have the same probability over prizes. The lotteries each have an equal prize in one position they have unequal prizes in the other position then if L_1 is the lottery with the better prize then $L_1 \succ L_2$; if neither lottery has a better prize then $L_1 \approx L_2$.

Cardinal Utility Theory: Better Prizes

Better Prizes: suppose L_1 is a lottery over (w, x) and L_2 is over (y, z) suppose that L_1 and L_2 have the same probability over prizes. The lotteries each have an equal prize in one position they have unequal prizes in the other position then if L_1 is the lottery with the better prize then $L_1 \succ L_2$; if neither lottery has a better prize then $L_1 \approx L_2$.

Lottery 1 (L_1) is 0.6 chance for p and 0.4 chance for y

Lottery 2 (L_2) is 0.6 chance for t and 0.4 chance for y

Cardinal Utility Theory: Better Prizes

Better Prizes: suppose L_1 is a lottery over (w, x) and L_2 is over (y, z) suppose that L_1 and L_2 have the same probability over prizes. The lotteries each have an equal prize in one position they have unequal prizes in the other position then if L_1 is the lottery with the better prize then $L_1 \succ L_2$; if neither lottery has a better prize then $L_1 \approx L_2$.

Lottery 1 (L_1) is 0.6 chance for p and 0.4 chance for y

Lottery 2 (L_2) is 0.6 chance for t and 0.4 chance for y

Since Ann prefers p to t, this axiom says that Ann prefers L_1 to L_2

Cardinal Utility Theory: Better Chances

Better Chances: Suppose L_1 and L_2 are two lotteries which have the same prizes, then if L_1 offers a better chance of the better prize, then $L_1 \succ L_2$

Cardinal Utility Theory: Better Chances

Better Chances: Suppose L_1 and L_2 are two lotteries which have the same prizes, then if L_1 offers a better chance of the better prize, then $L_1 \succ L_2$

Lottery 1 (L_1) is 0.7 chance for p and 0.3 chance for y

Lottery 2 (L_2) is 0.6 chance for p and 0.4 chance for y

Cardinal Utility Theory: Better Chances

Better Chances: Suppose L_1 and L_2 are two lotteries which have the same prizes, then if L_1 offers a better chance of the better prize, then $L_1 \succ L_2$

Lottery 1 (L_1) is 0.7 chance for p and 0.3 chance for y

Lottery 2 (L_2) is 0.6 chance for p and 0.4 chance for y

This axioms states that Ann must prefer L_1 to L_2

Cardinal Utility Theory: Reduction of Compound Lotteries

Reduction of Compound Lotteries: If the prize of a lottery is another lottery, then this can be reduced to a simple lottery over prizes.

Cardinal Utility Theory: Reduction of Compound Lotteries

Reduction of Compound Lotteries: If the prize of a lottery is another lottery, then this can be reduced to a simple lottery over prizes.

This eliminates utility from the thrill of gambling and so the only ultimate concern is the prizes.

Cardinal Utility Theory

Von Neumann-Morgenstern Theorem. If an agent satisfies the previous axioms, then the agents ordinal utility function can be turned into cardinal utility function.

Cardinal Utility Theory

Von Neumann-Morgenstern Theorem. If an agent satisfies the previous axioms, then the agents ordinal utility function can be turned into cardinal utility function.

- Utility is unique only *up to linear transformations*. So, it still does not make sense to add two different agents cardinal utility functions.

Cardinal Utility Theory

Von Neumann-Morgenstern Theorem. If an agent satisfies the previous axioms, then the agents ordinal utility function can be turned into cardinal utility function.

- Utility is unique only *up to linear transformations*. So, it still does not make sense to add two different agents cardinal utility functions.
- Issue with continuity: $1\text{EUR} \succ 1 \text{ cent } \succ \text{death}$, but who would accept a lottery which is p for 1EUR and $(1 - p)$ for death??

Cardinal Utility Theory

Von Neumann-Morgenstern Theorem. If an agent satisfies the previous axioms, then the agents ordinal utility function can be turned into cardinal utility function.

- Utility is unique only *up to linear transformations*. So, it still does not make sense to add two different agents cardinal utility functions.
- Issue with continuity: $1\text{EUR} \succ 1\text{ cent } \succ \text{death}$, but who would accept a lottery which is p for 1EUR and $(1-p)$ for death??
- Deep issues about how to identify correct descriptions of the outcomes and options.

Issue with Better Prizes

Suppose you have a kitten, which you plan to give away to either Ann or Bob. Ann and Bob both want the kitten very much. Both are deserving, and both would care for the kitten. You are sure that giving the kitten to Ann (x) is at least as good as giving the kitten to Bob (y) (so $x \succeq y$). But you think that would be unfair to Bob. You decide to flip a fair coin: if the coin lands heads, you will give the kitten to Bob, and if it lands tails, you will give the kitten to Ann. (J. Drier, "Morality and Decision Theory" in [HR])

Issue with Better Prizes

Suppose you have a kitten, which you plan to give away to either Ann or Bob. Ann and Bob both want the kitten very much. Both are deserving, and both would care for the kitten. You are sure that giving the kitten to Ann (x) is at least as good as giving the kitten to Bob (y) (so $x \succeq y$). But you think that would be unfair to Bob. You decide to flip a fair coin: if the coin lands heads, you will give the kitten to Bob, and if it lands tails, you will give the kitten to Ann. (J. Drier, "Morality and Decision Theory" in [HR])

Why does this contradict better prizes?

Issue with Better Prizes

Suppose you have a kitten, which you plan to give away to either Ann or Bob. Ann and Bob both want the kitten very much. Both are deserving, and both would care for the kitten. You are sure that giving the kitten to Ann (x) is at least as good as giving the kitten to Bob (y) (so $x \succeq y$). But you think that would be unfair to Bob. You decide to flip a fair coin: if the coin lands heads, you will give the kitten to Bob, and if it lands tails, you will give the kitten to Ann. (J. Drier, "Morality and Decision Theory" in [HR])

Why does this contradict better prizes? consider the lottery which is x for sure (L_1) and the lottery which is 0.5 for y and 0.5 for x (L_2).

Issue with Better Prizes

Suppose you have a kitten, which you plan to give away to either Ann or Bob. Ann and Bob both want the kitten very much. Both are deserving, and both would care for the kitten. You are sure that giving the kitten to Ann (x) is at least as good as giving the kitten to Bob (y) (so $x \succeq y$). But you think that would be unfair to Bob. You decide to flip a fair coin: if the coin lands heads, you will give the kitten to Bob, and if it lands tails, you will give the kitten to Ann. (J. Drier, "Morality and Decision Theory" in [HR])

Why does this contradict better prizes? consider the lottery which is x for sure (L_1) and the lottery which is 0.5 for y and 0.5 for x (L_2). Better prizes implies $L_1 \succeq L_2$

Issue with Better Prizes

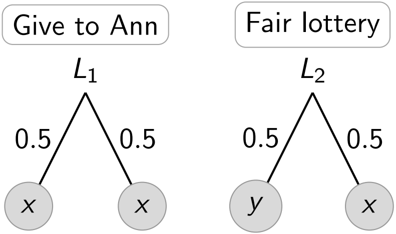
Suppose you have a kitten, which you plan to give away to either Ann or Bob. Ann and Bob both want the kitten very much. Both are deserving, and both would care for the kitten. You are sure that giving the kitten to Ann (x) is at least as good as giving the kitten to Bob (y) (so $x \succeq y$). But you think that would be unfair to Bob. You decide to flip a fair coin: if the coin lands heads, you will give the kitten to Bob, and if it lands tails, you will give the kitten to Ann. (J. Drier, "Morality and Decision Theory" in [HR])

Why does this contradict better prizes? consider the lottery which is x for sure (L_1) and the lottery which is 0.5 for y and 0.5 for x (L_2). Better prizes implies $L_1 \succeq L_2$ but a person concerned with fairness may have $L_2 \succeq L_1$.

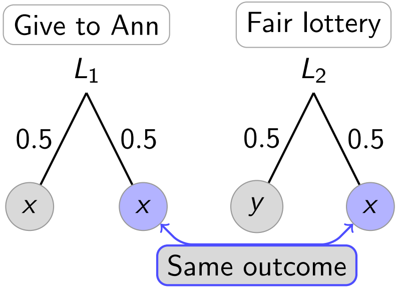
Issue with Better Prizes

Suppose you have a kitten, which you plan to give away to either Ann or Bob. Ann and Bob both want the kitten very much. Both are deserving, and both would care for the kitten. You are sure that giving the kitten to Ann (x) is at least as good as giving the kitten to Bob (y) (so $x \succeq y$). But you think that would be unfair to Bob. You decide to flip a fair coin: if the coin lands heads, you will give the kitten to Bob, and if it lands tails, you will give the kitten to Ann. (J. Drier, "Morality and Decision Theory" in [HR])

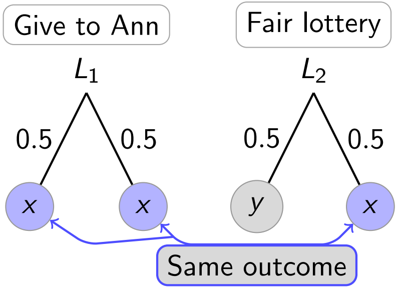
Why does this contradict better prizes? consider the lottery which is x for sure (L_1) and the lottery which is 0.5 for y and 0.5 for x (L_2). Better prizes implies $L_1 \succeq L_2$ but a person concerned with fairness may have $L_2 \succeq L_1$. *But if fairness is important then that should be part of the description of the outcome!*



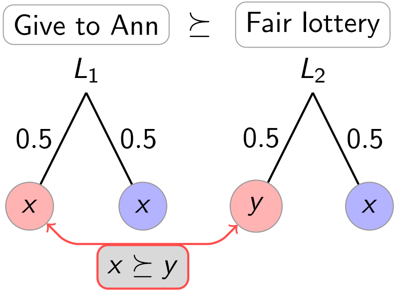
- x is the outcome "Ann gets the kitten"
- y is the outcome "Bob gets the kitten"



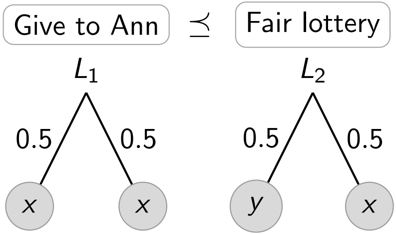
- x is the outcome "Ann gets the kitten"
- y is the outcome "Bob gets the kitten"



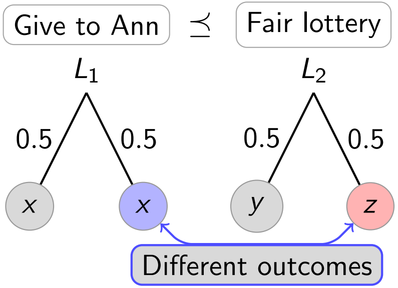
- x is the outcome "Ann gets the kitten"
- y is the outcome "Bob gets the kitten"



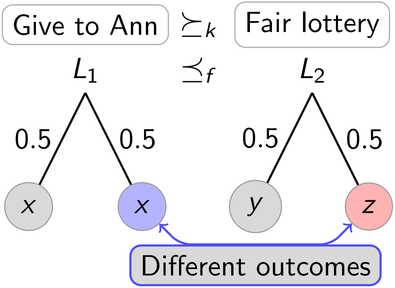
- x is the outcome "Ann gets the kitten"
- y is the outcome "Bob gets the kitten"



- x is the outcome "Ann gets the kitten, *in a fair way*"
- y is the outcome "Bob gets the kitten"



- x is the outcome "Ann gets the kitten"
- z is the outcome "Ann gets the outcome, *fairly*
- y is the outcome "Bob gets the kitten, *fairly*"



If all the agent cares about is who gets the kitten, then $L_1 \succeq L_2$

If all the agent cares about is being fair, then $L_1 \preceq L_2$

Allais Paradox, Again

	Options	Red (1)	White (89)	Blue (10)
S_1	A	$1M$	$1M$	$1M$
	B	0	$1M$	$5M$
S_2	C	$1M$	0	$1M$
	D	0	0	$5M$

Allais Paradox, Again

	Options	Red (1)	White (89)	Blue (10)
S_1	A	$1M$	$1M$	$1M$
	B	0	$1M$	$5M$
S_2	C	$1M$	0	$1M$
	D	0	0	$5M$

same prize configurations and the same chance of winning the prizes implies one will have the same preferences.

Allais Paradox, Again

	Options	Red (1)	White (89)	Blue (10)
S_1	A	$1M$	$1M$	$1M$
	B	0	$1M$	$5M$
S_2	C	$1M$	0	$1M$
	D	0	0	$5M$

same prize configurations and the same chance of winning the prizes implies one will have the same preferences.

In S_1, many people would choose A over B ($A \succeq B$).

Allais Paradox, Again

	Options	Red (1)	White (89)	Blue (10)
S_1	A	$1M$	$1M$	$1M$
	B	0	$1M$	$5M$
S_2	C	$1M$	0	$1M$
	D	0	0	$5M$

same prize configurations and the same chance of winning the prizes implies one will have the same preferences.

In S_1, many people would choose A over B ($A \succeq B$). *But, according to the axioms, this cannot be because of the white ball*.

Allais Paradox, Again

	Options	Red (1)	White (89)	Blue (10)
S_1	A	$1M$	$1M$	$1M$
	B	0	$1M$	$5M$
S_2	C	$1M$	0	$1M$
	D	0	0	$5M$

same prize configurations and the same chance of winning the prizes implies one will have the same preferences.

In S_1, many people would choose A over B ($A \succeq B$). *But, according to the axioms, this cannot be because of the white ball*. So, your preferences in S_2 should be C over D ($C \succeq D$),

Allais Paradox, Again

	Options	Red (1)	White (89)	Blue (10)
S_1	A	$1M$	$1M$	$1M$
	B	0	$1M$	$5M$
S_2	C	$1M$	0	$1M$
	D	0	0	$5M$

same prize configurations and the same chance of winning the prizes implies one will have the same preferences.

In S_1, many people would choose A over B ($A \succeq B$). *But, according to the axioms, this cannot be because of the white ball*. So, your preferences in S_2 should be C over D ($C \succeq D$), but many people prefer D over C.

Allais Paradox

We should not conclude either

Allais Paradox

We should not conclude either

(a) The axioms of cardinal utility fail to adequately capture our understanding of rational choice, or

Allais Paradox

We should not conclude either

(a) The axioms of cardinal utility fail to adequately capture our understanding of rational choice, or

(b) those who choose A in S_1 and D is L_2 are irrational.

Allais Paradox

We should not conclude either

(a) The axioms of cardinal utility fail to adequately capture our understanding of rational choice, or

(b) those who choose A in S_1 and D is L_2 are irrational.

Rather, peoples utility functions (*their rankings over outcomes*) are often far more complicated than the monetary bets would indicate....

Instrumental Rationality

Is decision theory a formalization of instrumental rationality?

Instrumental Rationality

Is decision theory a formalization of instrumental rationality?

If "goals" and "preferences" are the same thing, then decision theory is simply a formal version of instrumental rationality.

Instrumental Rationality

Is decision theory a formalization of instrumental rationality?

If "goals" and "preferences" are the same thing, then decision theory is simply a formal version of instrumental rationality.

Decision theory gives the agent some way to determine what is the "best" option, but in general this need not be the option that leads to the highest satisfaction of one's goals.

Ultimatum Game

There is a good (say an amount of money) to be divided between two players.

Ultimatum Game

There is a good (say an amount of money) to be divided between two players. In order for either player to get the money, both players must agree to the division.

Ultimatum Game

There is a good (say an amount of money) to be divided between two players. In order for either player to get the money, both players must agree to the division. One player is selected by the experimenter to go first and is given all the money (call her the "Proposer"): the Proposer gives and ultimatum of the form "I get x percent and you get y percent — take it or leave it!".

Ultimatum Game

There is a good (say an amount of money) to be divided between two players. In order for either player to get the money, both players must agree to the division. One player is selected by the experimenter to go first and is given all the money (call her the "Proposer"): the Proposer gives and ultimatum of the form "I get x percent and you get y percent — take it or leave it!". No negotiation is allowed ($x + y$ must not exceed 100%).

Ultimatum Game

There is a good (say an amount of money) to be divided between two players. In order for either player to get the money, both players must agree to the division. One player is selected by the experimenter to go first and is given all the money (call her the "Proposer"): the Proposer gives and ultimatum of the form "I get x percent and you get y percent — take it or leave it!". No negotiation is allowed ($x + y$ must not exceed 100%). The second player is the Disposer: she either accepts or rejects the offer. If the Disposer rejects, then both players get 0 otherwise they get the proposed division.

Ultimatum Game

There is a good (say an amount of money) to be divided between two players. In order for either player to get the money, both players must agree to the division. One player is selected by the experimenter to go first and is given all the money (call her the "Proposer"): the Proposer gives and ultimatum of the form "I get x percent and you get y percent — take it or leave it!". No negotiation is allowed ($x + y$ must not exceed 100%). The second player is the Disposer: she either accepts or rejects the offer. If the Disposer rejects, then both players get 0 otherwise they get the proposed division.

Suppose the players meet only once. It would seem that the Proposer should propose 99% for herself and 1% for the Disposer. And if the Disposer is instrumentally rational, then she should accept the offer.

Ultimatum Game

But this is not what happens in experiments: if the Disposer is offered 1%, 10% or even 20%, the Disposer very often rejects. Furthermore, the proposer tends demand only around 60%.

Ultimatum Game

But this is not what happens in experiments: if the Disposer is offered 1%, 10% or even 20%, the Disposer very often rejects. Furthermore, the proposer tends demand only around 60%.

A typical explanation is that the players' utility functions are not simply about getting funds to best advance their goals, but about acting according to some norms of fair play.

Ultimatum Game

But this is not what happens in experiments: if the Disposer is offered 1%, 10% or even 20%, the Disposer very often rejects. Furthermore, the proposer tends demand only around 60%.

A typical explanation is that the players' utility functions are not simply about getting funds to best advance their goals, but about acting according to some norms of fair play. But acting according to norms of fair play does not seem to be a goal: it is a principle to which a person wishes to conform.

Choice Processes and Outcomes

A. Sen. *Maximization and the Act of Choice*. Econometrica, Vol. 65, No. 4, 1997, 745 - 779.

"The formulation of maximizing behavior in economics has often parallels the modeling of maximization in physics an related disciplines.

Choice Processes and Outcomes

A. Sen. *Maximization and the Act of Choice*. Econometrica, Vol. 65, No. 4, 1997, 745 - 779.

"The formulation of maximizing behavior in economics has often parallels the modeling of maximization in physics an related disciplines. But maximizing *behavior* differs from nonvolitional maximization because of the fundamental relevance of the choice act, which has to be placed in a central position in analyzing maximizing behavior.

Choice Processes and Outcomes

A. Sen. *Maximization and the Act of Choice*. Econometrica, Vol. 65, No. 4, 1997, 745 - 779.

"The formulation of maximizing behavior in economics has often parallels the modeling of maximization in physics an related disciplines. But maximizing *behavior* differs from nonvolitional maximization because of the fundamental relevance of the choice act, which has to be placed in a central position in analyzing maximizing behavior. A person's preferences over *comprehensive* outcomes (including the choice process) have to be distinguished form the conditional preferences over *culmination* outcomes *given* the act of choice." (pg. 745)

Choice Functions

Suppose X is a set of options. And consider $B \subseteq X$ as a choice problem. A **choice function** is any function where $C(B) \subseteq B$. B is sometimes called a menu and $C(B)$ the set of "rational" or "desired" choices.

Choice Functions

Suppose X is a set of options. And consider $B \subseteq X$ as a choice problem. A **choice function** is any function where $C(B) \subseteq B$. B is sometimes called a menu and $C(B)$ the set of "rational" or "desired" choices.

A relation R on X **rationalizes a choice function** C if for all B $C(B) = \{x \in B \mid \text{for all } y \in B \quad xRy\}$. (i.e., the agent is chooses according to some preference ordering).

Choice Functions

Suppose X is a set of options. And consider $B \subseteq X$ as a choice problem. A **choice function** is any function where $C(B) \subseteq B$. B is sometimes called a menu and $C(B)$ the set of "rational" or "desired" choices.

A relation R on X **rationalizes a choice function** C if for all B $C(B) = \{x \in B \mid \text{for all } y \in B \quad xRy\}$. (i.e., the agent is chooses according to some preference ordering).

Sen's α: If $x \in C(A)$ and $B \subset A$ and $x \in B$ then $x \in C(B)$

Choice Functions

Suppose X is a set of options. And consider $B \subseteq X$ as a choice problem. A **choice function** is any function where $C(B) \subseteq B$. B is sometimes called a menu and $C(B)$ the set of "rational" or "desired" choices.

A relation R on X **rationalizes a choice function** C if for all B $C(B) = \{x \in B \mid \text{for all } y \in B \quad xRy\}$. (i.e., the agent is chooses according to some preference ordering).

Sen's α: If $x \in C(A)$ and $B \subset A$ and $x \in B$ then $x \in C(B)$

Sen's β: If $x, y \in C(A)$, $A \subset B$ and $y \in C(B)$ then $x \in C(B)$.

You arrive at a garden party and can readily identify the most comfortable chair. You would be delighted if an imperious host were to assign you that chair. However, if the matter is left to your own choice, you may refuse to rush to it.

You arrive at a garden party and can readily identify the most comfortable chair. You would be delighted if an imperious host were to assign you that chair. However, if the matter is left to your own choice, you may refuse to rush to it. You select a "less preferred" chair.

You arrive at a garden party and can readily identify the most comfortable chair. You would be delighted if an imperious host were to assign you that chair. However, if the matter is left to your own choice, you may refuse to rush to it. You select a "less preferred" chair. Are you still a maximizer?

You arrive at a garden party and can readily identify the most comfortable chair. You would be delighted if an imperious host were to assign you that chair. However, if the matter is left to your own choice, you may refuse to rush to it. You select a “less preferred” chair. Are you still a maximizer? Quite possibly you are, since your preference ranking for choice behavior may well be defined over “comprehensive outcomes”, including choice processes (in particular, who does the choosing) as well as the outcomes at culmination (the distribution of chairs).

You arrive at a garden party and can readily identify the most comfortable chair. You would be delighted if an imperious host were to assign you that chair. However, if the matter is left to your own choice, you may refuse to rush to it. You select a "less preferred" chair. Are you still a maximizer? Quite possibly you are, since your preference ranking for choice behavior may well be defined over "comprehensive outcomes", including choice processes (in particular, who does the choosing) as well as the outcomes at culmination (the distribution of chairs).

To take another example, you may prefer mangoes to apples, but refuse to pick the last mango from a fruit basket, and yet be very pleased if someone else were to "force" that last mango on you. "
(Sen, pg. 747)

Let $X = \{x, y, z\}$ and consider $B_1 = X$ and $B_2 = \{x, y\}$. Define

$$C(B_1) = C(\{x, y, z\}) = \{x\}$$
$$C(B_2) = C(\{x, y\}) = \{y\}$$

This choice function cannot be rationalized.

Framing effects

Logicophilia, a virulent virus, threatens 600 students at Tilburg University

[Adapted from Tversky and Kahneman (1981)]

Framing effects

Logicophilia, a virulent virus, threatens 600 students at Tilburg University

1. You must choose between two prevention programs, resulting in:
   - A: 200 participants will be saved for sure.
   - B: 33 % chance of saving all of them, otherwise no one will be saved.

[Adapted from Tversky and Kahneman (1981)]

Framing effects

Logicophilia, a virulent virus, threatens 600 students at Tilburg University

1. You must choose between two prevention programs, resulting in:
   - A: 200 participants will be saved for sure.
   - B: 33 % chance of saving all of them, otherwise no one will be saved.

   72 % of the participants choose A over B.

[Adapted from Tversky and Kahneman (1981)]

Framing effects

Logicophilia, a virulent virus, threatens 600 students at Tilburg University

2. You must choose between two prevention programs, resulting in:
   - A': 400 will not be saved, for sure.
   - B': 33 % chance of saving all of them, otherwise no one will be saved.

[Adapted from Tversky and Kahneman (1981)]

Framing effects

Logicophilia, a virulent virus, threatens 600 students at Tilburg University

2. You must choose between two prevention programs, resulting in:
   - A': 400 will not be saved, for sure.
   - B': 33 % chance of saving all of them, otherwise no one will be saved.

   78 % of the participants choose B' over A'.

[Adapted from Tversky and Kahneman (1981)]

Framing effects

Logicophilia, a virulent virus, threatens 600 students at Tilburg University

1. You must choose between two prevention programs, resulting in:
   - A: 200 participants will be saved for sure.
   - B: 33 % chance of saving all of them, otherwise no one will be saved.

   72 % of the participants choose A over B.
2. You must choose between two prevention programs, resulting in:
   - A': 400 will not be saved, for sure.
   - B': 33 % chance of saving all of them, otherwise no one will be saved.

   78 % of the participants choose B' over A'.

[Adapted from Tversky and Kahneman (1981)]

Framing effects

The Experiment:												
A:	0	$+$	200	for	sure.	**B:**	(33%	600)	$+$	(66%	0).
$\Rightarrow$ 72 % of the participants choose A over B.											
A':	600	-	400	for	sure.	**B':**	(33%	600)	$+$	(66%	0).
$\Rightarrow$ 78 % of the participants choose B' over A'.											

- Standard decision theory is extensional
  - Choosing A and $A \leftrightarrow B$ implies Choosing B.

Framing effects

The Experiment:	
A: 0 + 200 for sure.	**B:** (33% 600) + (66% 0).
⇒ 72 % of the participants choose A over B.	
A': 600 - 400 for sure.	**B':** (33% 600) + (66% 0).
⇒ 78 % of the participants choose B' over A'.	

- Standard decision theory is extensional
  - Choosing A and $A \leftrightarrow B$ implies Choosing B.

  Also true of many formalisms of beliefs:
  - "Believing" A and $\vdash A \leftrightarrow B$ implies "Believing" B.

Conclusions, I

- *Instrumental rationality* is a fundamental account of "rationality", but it is not necessarily the "whole of rationality"

Conclusions, I

- *Instrumental rationality* is a fundamental account of "rationality", but it is not necessarily the "whole of rationality"

- Utility is not a sort of "value", but simply a representation of one's ordering of options based on one's underlying values, ends and principles.

Conclusions, II

- If people are *really awful* and calculating probabilities, then it certainly does not help to understand their actions in terms of maximizing expected utility

Conclusions, II

- If people are *really awful* and calculating probabilities, then it certainly does not help to understand their actions in terms of maximizing expected utility (BUT, when mistakes are pointed out people tend to adjust their probabilities, and if the cases are described in terms of *frequencies*, then people are much better)

Conclusions, II

- If people are *really awful* and calculating probabilities, then it certainly does not help to understand their actions in terms of maximizing expected utility (BUT, when mistakes are pointed out people tend to adjust their probabilities, and if the cases are described in terms of *frequencies*, then people are much better)
- We need an account of which distinctions are relevant and which are not...what justifies a preference.

Conclusions, II

- If people are *really awful* and calculating probabilities, then it certainly does not help to understand their actions in terms of maximizing expected utility (BUT, when mistakes are pointed out people tend to adjust their probabilities, and if the cases are described in terms of *frequencies*, then people are much better)

- We need an account of which distinctions are relevant and which are not...what justifies a preference.

- Utility theory is a way to formalize and model rational action, but it is not itself a complete theory of rational action.

J. Pollock. *Rational Choice and Action Omnipotence*. The Philosophical Review, Vol. 111, No. 1 (2002), pgs. 1 - 23.

Three Issues

Three Issues

- What sort of *rational requirements* are imposed by decision theory?

J. Pollock. *How do you maximize Expectation Value?*. Nous, Vol. 17, No. 3 (1983), pgs. 409 - 421.

Three Issues

- What sort of *rational requirements* are imposed by decision theory?

J. Pollock. *How do you maximize Expectation Value?*. Nous, Vol. 17, No. 3 (1983), pgs. 409 - 421.

- How do rational choice-explanations *explain*?

J. Elster. *The Nature and Scope of Rational-Choice Explanation*. in *in Readings in the Philosophy of Social Science* (1985), pgs. 311 - 322.

Three Issues

- What sort of *rational requirements* are imposed by decision theory?

J. Pollock. *How do you maximize Expectation Value?*. Nous, Vol. 17, No. 3 (1983), pgs. 409 - 421.

- How do rational choice-explanations *explain*?

J. Elster. *The Nature and Scope of Rational-Choice Explanation*. in *in Readings in the Philosophy of Social Science* (1985), pgs. 311 - 322.

- Following Hume, there is a strict division between beliefs and desires (they may be entangled, but play very different roles in rational agency). Why should we maintain this division?

D. Lewis. *Desire as Belief*. Mind, 97, (1988), pgs. 323 - 332.

D. Lewis. *Desire as Belief II*. Mind, 105, (1996), pgs. 303 - 313.

Maximizing

A person should (rationally) perform an act iff the expectation value of his doing so is greater than the expectation value of his performing any alternative act.

Maximizing

A person should (rationally) perform an act iff the expectation value of his doing so is greater than the expectation value of his performing any alternative act.

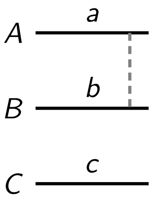
Maximizing

A person should (rationally) perform an act iff the expectation value of his doing so is greater than the expectation value of his performing any alternative act.

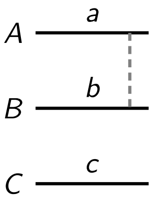
1. *A* is rationally obligatory iff *A* has a higher expectation value than any act incompatible with A
2. *A* is rationally permissible iff *A* is at least as great as that of any act incompatible with *A*

1. A is rationally obligatory iff A has a higher expectation value than any act incompatible with A
2. A is rationally permissible iff A is at least as great as that of any act incompatible with A

1. *A* is rationally obligatory iff *A* has a higher expectation value than any act incompatible with A
2. *A* is rationally permissible iff *A* is at least as great as that of any act incompatible with *A*

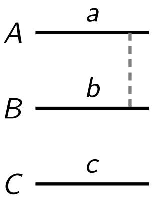


1. A is rationally obligatory iff A has a higher expectation value than any act incompatible with A
2. A is rationally permissible iff A is at least as great as that of any act incompatible with A



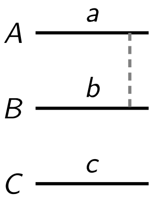
- Assume E_A, E_B and E_C are all positive

1. A is rationally obligatory iff A has a higher expectation value than any act incompatible with A
2. A is rationally permissible iff A is at least as great as that of any act incompatible with A



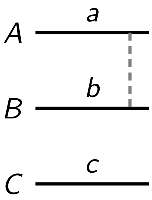
- Assume E_A, E_B and E_C are all positive
- Assume $E_A = E_B$, so by (2), both A and B are rationally permissible

1. A is rationally obligatory iff A has a higher expectation value than any act incompatible with A
2. A is rationally permissible iff A is at least as great as that of any act incompatible with A



- Assume E_A, E_B and E_C are all positive
- Assume $E_A = E_B$, so by (2), both A and B are rationally permissible
- $E_{BC} = E_B + E_C$ (acts B and C and independent)

1. A is rationally obligatory iff A has a higher expectation value than any act incompatible with A
2. A is rationally permissible iff A is at least as great as that of any act incompatible with A



- Assume E_A, E_B and E_C are all positive
- Assume $E_A = E_B$, so by (2), both A and B are rationally permissible
- $E_{BC} = E_B + E_C$ (acts B and C and independent)
- *$E_A < E_{BC}$, so by (1) one is rationally obligated to refrain from performing A.*

Maximizing Expectation Value

Consistency: If each member of a set of acts is rationally obligatory (at time t) then it must be possible (at t) to perform them all.

Maximizing Expectation Value

Consistency: If each member of a set of acts is rationally obligatory (at time t) then it must be possible (at t) to perform them all.

Maximizing Expectation Value

Consistency: If each member of a set of acts is rationally obligatory (at time t) then it must be possible (at t) to perform them all.

1. A is rationally obligatory iff A is prescribed by some maximal strategy which is rationally preferable to any maximal strategy which does not prescribe A.

Maximizing Expectation Value

Consistency: If each member of a set of acts is rationally obligatory (at time t) then it must be possible (at t) to perform them all.

1. A is rationally obligatory iff A is prescribed by some maximal strategy which is rationally preferable to any maximal strategy which does not prescribe A.

Maximizing Expectation Value

Consistency: If each member of a set of acts is rationally obligatory (at time t) then it must be possible (at t) to perform them all.

1. A is rationally obligatory iff A is prescribed by some maximal strategy which is rationally preferable to any maximal strategy which does not prescribe A.

2. One maximal strategy is rationally preferable to another iff the first has a higher expectation value than the second.

Maximizing Expectation Value

Consistency: If each member of a set of acts is rationally obligatory (at time t) then it must be possible (at t) to perform them all.

1. A is rationally obligatory iff A is prescribed by some maximal strategy which is rationally preferable to any maximal strategy which does not prescribe A.

2. A maximal strategy S is rationally preferable to another S^* iff there is a time t_0 such that for every time t later than t_0, $E(S_t) > E(S_t^*)$.

Riddles of Maximization

Ever Better Wine:
The wine slowly improves with age. More good news: You are immortal. Consequently, you are indifferent as to when you consume a particular good. When should you drink the wine?

Riddles of Maximization

Ever Better Wine:
The wine slowly improves with age. More good news: You are immortal. Consequently, you are indifferent as to when you consume a particular good. When should you drink the wine?

- Not now. The wine will be better later.

Riddles of Maximization

Ever Better Wine:
The wine slowly improves with age. More good news: You are immortal. Consequently, you are indifferent as to when you consume a particular good. When should you drink the wine?

- Not now. The wine will be better later.
- Not later. For at any given time it will be true that the wine will be even better if you waited longer

Riddles of Maximization

Ever Better Wine:
The wine slowly improves with age. More good news: You are immortal. Consequently, you are indifferent as to when you consume a particular good. When should you drink the wine?

- Not now. The wine will be better later.
- Not later. For at any given time it will be true that the wine will be even better if you waited longer
- But if you do not drink the wine now and do not drink it later, then you will not drink it at all!

Rational Choice Explanations

J. Elster. *The Nature and Scope of Rational-Choice Explanation*. in *in Readings in the Philosophy of Social Science* (1985), pgs. 311 - 322.

Ideally, a fully satisfactory rational-choice explanation of an action would have the following structure.

Rational Choice Explanations

J. Elster. *The Nature and Scope of Rational-Choice Explanation*. in *in Readings in the Philosophy of Social Science* (1985), pgs. 311 - 322.

Ideally, a fully satisfactory rational-choice explanation of an action would have the following structure.

Optimality Part: It would show that he action is the (unique) best way of satisfying the full set of the agent's desires, given the (uniquely) best beliefs that agent could form, relative to the (uniquely determined) optimal amount of evidence.

Rational Choice Explanations

J. Elster. *The Nature and Scope of Rational-Choice Explanation*. in *in Readings in the Philosophy of Social Science* (1985), pgs. 311 - 322.

Ideally, a fully satisfactory rational-choice explanation of an action would have the following structure.

Optimality Part: It would show that he action is the (unique) best way of satisfying the full set of the agent's desires, given the (uniquely) best beliefs that agent could form, relative to the (uniquely determined) optimal amount of evidence.

Causal Part: In addition the explanation would show that the action was caused (in the right way) by the desires and beliefs, and the beliefs caused (in the right way) by consideration of the evidence.

Desire As Belief Thesis

D. Lewis. *Desire as Belief*. Mind, 97, (1988), pgs. 323 - 332.

D. Lewis. *Desire as Belief II*. Mind, 105, (1996), pgs. 303 - 313.

Desire-As-Belief

Let P be a probability function (assigning elements of $[0, 1]$ to propositions) and U a utility function (assigning elements of $\mathbb{R}$ to propositions).

Desire-As-Belief

Let P be a probability function (assigning elements of $[0, 1]$ to propositions) and U a utility function (assigning elements of $\mathbb{R}$ to propositions).

Let $P_p(\cdot) = P(\cdot \mid p)$ be the probability obtained by Bayesian updating upon learning p; and $U_p(\cdot)$ the utility function of the agent upon learning p.

Desire-As-Belief

Let P be a probability function (assigning elements of $[0, 1]$ to propositions) and U a utility function (assigning elements of $\mathbb{R}$ to propositions).

Let $P_p(\cdot) = P(\cdot \mid p)$ be the probability obtained by Bayesian updating upon learning p; and $U_p(\cdot)$ the utility function of the agent upon learning p.

For each proposition p, there is a corresponding proposition p° expressing that p is *desirable*.

Desire-As-Belief

Let P be a probability function (assigning elements of $[0, 1]$ to propositions) and U a utility function (assigning elements of $\mathbb{R}$ to propositions).

Let $P_p(\cdot) = P(\cdot \mid p)$ be the probability obtained by Bayesian updating upon learning p; and $U_p(\cdot)$ the utility function of the agent upon learning p.

For each proposition p, there is a corresponding proposition p° expressing that p is *desirable*.

For all utility functions U and probability functions P:

(1) Desire-as-Belief Thesis: For any p, $U(p) = P(p^\circ)$
(2) Invariance Thesis: For any p, $U_p(p) = U(p)$

Desire-As-Belief

(1) Desire-as-Belief Thesis: For any p, $U(p) = P(p^\circ)$
(2) Invariance Thesis: For any p, $U_p(p) = U(p)$

Lewis: (1) and (2) conflict with each other.

Desire-As-Belief

(1) Desire-as-Belief Thesis: For any p, $U(p) = P(p^{\circ})$

(2) Invariance Thesis: For any p, $U_p(p) = U(p)$

Lewis: (1) and (2) conflict with each other.

For any p, $P_p(p^{\circ}) = U_p(p)$

(1) applied to U_p and P_p

Desire-As-Belief

(1) Desire-as-Belief Thesis: For any p, $U(p) = P(p^\circ)$
(2) Invariance Thesis: For any p, $U_p(p) = U(p)$

Lewis: (1) and (2) conflict with each other.

For any p, $P_p(p^\circ) = U_p(p) = U(p)$

(2) applied to U_p

Desire-As-Belief

(1) Desire-as-Belief Thesis: For any p, $U(p) = P(p^\circ)$

(2) Invariance Thesis: For any p, $U_p(p) = U(p)$

Lewis: (1) and (2) conflict with each other.

For any p, $P_p(p^\circ) = U_p(p) = U(p) = P(p^\circ)$

(1) applied to U and P

Desire-As-Belief

(1) Desire-as-Belief Thesis: For any p, $U(p) = P(p^\circ)$
(2) Invariance Thesis: For any p, $U_p(p) = U(p)$

Lewis: (1) and (2) conflict with each other.

For any p, $P_p(p^\circ) = U_p(p) = U(p) = P(p^\circ)$

So, for all p, $P(p^\circ \mid p) = P(p^\circ)$.

Desire-As-Belief

(1) Desire-as-Belief Thesis: For any p, $U(p) = P(p^\circ)$
(2) Invariance Thesis: For any p, $U_p(p) = U(p)$

Lewis: (1) and (2) conflict with each other.

For any p, $P_p(p^\circ) = U_p(p) = U(p) = P(p^\circ)$

So, for all p, $P(p^\circ \mid p) = P(p^\circ)$.

This fails for many probability measures P and if not, let $q = \neg(p \wedge p^\circ)$, then (assuming $P_p(p^\circ) = P(p^\circ)$) $0 = P_q(p^\circ \mid p) \neq P_q(p^\circ) > 0$.

Analyzing the Argument

R. Bradley and C. List. *Desire-as-belief revisited*. Analysis, 69(1), pgs. 31 - 37, 2009.

A. Hájek and P. Pettit. *Desire Beyond Belief*. Australasian Journal of Philosophy, 82(1), pgs. 77 - 92, 2004.

H. Árlo-Costa, J. Collins and I. Levi. *Desire-as-Belief Implies Opinionation or Indifference*. Analysis, 55, pgs. 2 - 5, 1995.

J. Collins. *Desire, Belief and Expectation*. Mind, 100, pgs. 333 - 342, 1997.

Next: Game Theory