Rationality

Lecture 11

Eric Pacuit

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

November 11, 2010

Rationality: Two Themes

Rationality: Two Themes

Rationality is a matter of **reasons**:

- Whether a belief P is *rational* depends on the *reasons for holding* P
- Whether an act α is *rational* depends on the *reason for doing* α

Rationality: Two Themes

Rationality is a matter of **reasons**:

- Whether a belief P is *rational* depends on the *reasons for holding* P
- Whether an act α is *rational* depends on the *reason for doing* α

Rationality is a matter of *reliability*:

- A rational belief is one that is arrived at a through a process that reliably produces beliefs that are true.
- A act is rational if it is arrived at through a process that reliably achieves specified goals.

Rationality: Two Themes

"Neither theme alone exhausts our notion of rationality. Reasons without reliability seem emtpy, reliability without reasons seems blind. In tandem these make a powerful unit, but how exactly are they related and why?" (Nozick, pg. 64)

Instrumental Rationality

"The notion of instrumental rationality is a powerful and natural one...

Instrumental Rationality

"The notion of instrumental rationality is a powerful and natural one...Instrumental rationality is within the intersection of all theories of rationality (and perhaps nothing else is). In this sense, instrumental rationality is the default theory, the theory that all discussants of rationality can take for granted, whatever else they think.

Instrumental Rationality

"The notion of instrumental rationality is a powerful and natural one...Instrumental rationality is within the intersection of all theories of rationality (and perhaps nothing else is). In this sense, instrumental rationality is the default theory, the theory that all discussants of rationality can take for granted, whatever else they think. There is something more, I think. The instrumental theory of rationality does not seem to stand in need of justification, whereas every other theory does. Every other theory must produces reasons for holding that what it demarcates is indeed rationality. Instrumental rationality is the base state. The question is whether it is the *whole* of rationality." (Nozick, pg. 133)

Instrumental Rationality I

What does it mean to be *instrumentally rational*?

Instrumental Rationality I

What does it mean to be *instrumentally rational*?

Rationality as Effectiveness: Ann's action α is instrumentally rational iff Ann's α-ing is an *effective* way for Ann to achieve her goal, desire, end or taste G.

Instrumental Rationality I

What does it mean to be *instrumentally rational*?

Rationality as Effectiveness: Ann's action α is instrumentally rational iff Ann's α-ing is an *effective* way for Ann to achieve her goal, desire, end or taste G.

Too narrow:

Instrumental Rationality I

What does it mean to be *instrumentally rational*?

Rationality as Effectiveness: Ann's action α is instrumentally rational iff Ann's α-ing is an *effective* way for Ann to achieve her goal, desire, end or taste G.

Too narrow: Bob checks the forecast with on the local news, weather.com and the local newspaper.

Instrumental Rationality I

What does it mean to be *instrumentally rational*?

Rationality as Effectiveness: Ann's action α is instrumentally rational iff Ann's α-ing is an *effective* way for Ann to achieve her goal, desire, end or taste G.

Too narrow: Bob checks the forecast with on the local news, weather.com and the local newspaper. They all concur that it will be a gorgeous day. So, Bob leaves without an umbrella

Instrumental Rationality I

What does it mean to be *instrumentally rational*?

Rationality as Effectiveness: Ann's action α is instrumentally rational iff Ann's α-ing is an *effective* way for Ann to achieve her goal, desire, end or taste G.

Too narrow: Bob checks the forecast with on the local news, weather.com and the local newspaper. They all concur that it will be a gorgeous day. So, Bob leaves without an umbrella and gets soaked in a freak rainstorm.

Instrumental Rationality I

What does it mean to be *instrumentally rational*?

Rationality as Effectiveness: Ann's action α is instrumentally rational iff Ann's α-ing is an *effective* way for Ann to achieve her goal, desire, end or taste G.

Too narrow: Bob checks the forecast with on the local news, weather.com and the local newspaper. They all concur that it will be a gorgeous day. So, Bob leaves without an umbrella and gets soaked in a freak rainstorm.

Too broad:

Instrumental Rationality I

What does it mean to be *instrumentally rational*?

Rationality as Effectiveness: Ann's action α is instrumentally rational iff Ann's α-ing is an *effective* way for Ann to achieve her goal, desire, end or taste G.

Too narrow: Bob checks the forecast with on the local news, weather.com and the local newspaper. They all concur that it will be a gorgeous day. So, Bob leaves without an umbrella and gets soaked in a freak rainstorm.

Too broad: Charles never checks weather reports, but does consult her Ouiji board.

Instrumental Rationality I

What does it mean to be *instrumentally rational*?

Rationality as Effectiveness: Ann's action α is instrumentally rational iff Ann's α-ing is an *effective* way for Ann to achieve her goal, desire, end or taste G.

Too narrow: Bob checks the forecast with on the local news, weather.com and the local newspaper. They all concur that it will be a gorgeous day. So, Bob leaves without an umbrella and gets soaked in a freak rainstorm.

Too broad: Charles never checks weather reports, but does consult her Ouiji board. On the day that Bob got soaked, Charles' Ouiji board told him to take an umbrella, so he stayed dry.

Instrumental Rationality I

What does it mean to be *instrumentally rational*?

Rationality as Effectiveness: Ann's action α is instrumentally rational iff Ann's α-ing is an *effective* way for Ann to achieve her goal, desire, end or taste G.

Too narrow: Bob checks the forecast with on the local news, weather.com and the local newspaper. They all concur that it will be a gorgeous day. So, Bob leaves without an umbrella and gets soaked in a freak rainstorm.

Too broad: Charles never checks weather reports, but does consult her Ouiji board. On the day that Bob got soaked, Charles' Ouiji board told him to take an umbrella, so he stayed dry.

We need to take the agent's beliefs into account

Instrumental Rationality II

Subjective Rationality: Ann's action α is instrumentally rational iff when she chooses α: (1) her choice was based on her beliefs (B) and (2) *if* B were true beliefs, then α would be an effective way to achieve her goals, desires, tastes, etc.

Instrumental Rationality II

Subjective Rationality: Ann's action α is instrumentally rational iff when she chooses α: (1) her choice was based on her beliefs (B) and (2) *if* B were true beliefs, then α would be an effective way to achieve her goals, desires, tastes, etc.

Is Bob instrumentally rational according to the above definition?

Instrumental Rationality II

Subjective Rationality: Ann's action α is instrumentally rational iff when she chooses α: (1) her choice was based on her beliefs (B) and (2) *if* B were true beliefs, then α would be an effective way to achieve her goals, desires, tastes, etc.

Is Bob instrumentally rational according to the above definition?

Is Charles action deemed *irrational* according to the above definition?

Instrumental Rationality II

Subjective Rationality: Ann's action α is instrumentally rational iff when she chooses α: (1) her choice was based on her beliefs (B) and (2) *if* B were true beliefs, then α would be an effective way to achieve her goals, desires, tastes, etc.

Is Bob instrumentally rational according to the above definition?

Is Charles action deemed *irrational* according to the above definition?

What constraints should be placed on reasonable beliefs that underlie a rational choice?

Instrumental Rationality III

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.

Instrumental Rationality III

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.

Instrumental Rationality III

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.

Instrumental Rationality III

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.

Can goals be irrational?

Can goals be irrational?

Hume: Our reason cannot tell us what to desire, so *no desire can ever be against reason*

'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger...

Can goals be irrational?

Hume: Our reason cannot tell us what to desire, so *no desire can ever be against reason*

'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger...

Does this mean that "anything goes"?

Can goals be irrational?

Hume: Our reason cannot tell us what to desire, so *no desire can ever be against reason*

'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger...

Does this mean that "anything goes"?

- constraints on how preferences "hang together"

Can goals be irrational?

Hume: Our reason cannot tell us what to desire, so *no desire can ever be against reason*

'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger...

Does this mean that "anything goes"?

- constraints on how preferences "hang together"
  - transitivity, completeness, etc.

Can goals be irrational?

Hume: Our reason cannot tell us what to desire, so *no desire can ever be against reason*

'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger...

Does this mean that "anything goes"?

- constraints on how preferences "hang together"
  - transitivity, completeness, etc.
  - "a person shows herself to lack "rational integration" if she has some desire for x, yet also desires not to desire x" (Nozick, pg. 139 - 151)

Can goals be irrational?

Hume: Our reason cannot tell us what to desire, so *no desire can ever be against reason*

'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger...

Does this mean that "anything goes"?

- constraints on how preferences "hang together"
  - transitivity, completeness, etc.
  - "a person shows herself to lack "rational integration" if she has some desire for x, yet also desires not to desire x" (Nozick, pg. 139 - 151)
- the ultimate goal is *happiness*, other desires are the manifestation of the pursuit of happiness or pleasure

Reasons for preference

Can you simply prefer x to y for no reason at all?

Reasons for preference

Can you simply prefer x to y for no reason at all?

If the person prefers x to y , either

1. the person is willing to switch to preferring y to x for a small gain, or

Reasons for preference

Can you simply prefer x to y for no reason at all?

If the person prefers x to y , either

1. the person is willing to switch to preferring y to x for a small gain, or
2. the person *has some reason* to prefer x to y, or

Reasons for preference

Can you simply prefer x to y for no reason at all?

If the person prefers x to y , either

1. the person is willing to switch to preferring y to x for a small gain, or
2. the person *has some reason* to prefer x to y, or
3. the person has some reason to prefer preferring x to y to not doing that.

Preferences, Desires, Goals

Preferences, Desires, Goals

The person's preferences and desires are in equilibrium (with her beliefs about their causes)

Preferences, Desires, Goals

The person's preferences and desires are in equilibrium (with her beliefs about their causes)

The person does not have desires that she knows are impossible to fulfill

Preferences, Desires, Goals

The person's preferences and desires are in equilibrium (with her beliefs about their causes)

The person does not have desires that she knows are impossible to fulfill

A person will not have a goal for which she knows that there is no feasible route, however long, for her current situation to the achievement of that goal.

Preferences, Desires, Goals

The person's preferences and desires are in equilibrium (with her beliefs about their causes)

The person does not have desires that she knows are impossible to fulfill

A person will not have a goal for which she knows that there is no feasible route, however long, for her current situation to the achievement of that goal.

Some goals are *stable* (recall Bratman on *plans*)

Preferences, Desires, Goals

The person's preferences and desires are in equilibrium (with her beliefs about their causes)

The person does not have desires that she knows are impossible to fulfill

A person will not have a goal for which she knows that there is no feasible route, however long, for her current situation to the achievement of that goal.

Some goals are *stable* (recall Bratman on *plans*)

R. Nozick. *"Rational Preferences"*. in The Nature of Rationality, pgs. 139 - 151.

Economic Rationality

Can we characterize *Homo Economicus* simply in terms of instrumental rationality?

Eg., Ann is eating ice cream.

Economic Rationality

Can we characterize *Homo Economicus* simply in terms of instrumental rationality?

Eg., Ann is eating ice cream.

Consumption Rationality: Ann's action α is "consumptively rational" only if it is an instance of the α-type — a general desire, value, or end of hers.

Economic Rationality

Can we characterize *Homo Economicus* simply in terms of instrumental rationality?

Eg., Ann is eating ice cream.

Consumption Rationality: Ann's action α is "consumptively rational" only if it is an instance of the α-type — a general desire, value, or end of hers.

Economic Rationality Ann's action α is economically rational only if it is (a) instrumentally rational or (b) consumptively rational.

What are preferences?

What are preferences?

Preferring or choosing x is different that "liking" x or "having a taste for x": one can prefer x to y but *dislike* both options

What are preferences?

Preferring or choosing x is different that "liking" x or "having a taste for x": one can prefer x to y but *dislike* both options

In utility theory, preferences are always understood as comparative: "preference" is more like "bigger" than "big"

What are preferences?

Preferring or choosing x is different that "liking" x or "having a taste for x": one can prefer x to y but *dislike* both options

In utility theory, preferences are always understood as comparative: "preference" is more like "bigger" than "big"

Revealed Preferences: Ann is said to have a preference for x over y iff Ann chooses x over y where choice is conceived of as overt behavior.

What are preferences?

Preferring or choosing x is different that "liking" x or "having a taste for x": one can prefer x to y but *dislike* both options

In utility theory, preferences are always understood as comparative: "preference" is more like "bigger" than "big"

Revealed Preferences: Ann is said to have a preference for x over y iff Ann chooses x over y where choice is conceived of as overt behavior.

Deliberative Preferences: A person deliberates and (ideally) ranks all the possible "outcomes"

What are preferences?

Preferring or choosing x is different that "liking" x or "having a taste for x": one can prefer x to y but *dislike* both options

In utility theory, preferences are always understood as comparative: "preference" is more like "bigger" than "big"

Revealed Preferences: Ann is said to have a preference for x over y iff Ann chooses x over y where choice is conceived of as overt behavior.

Deliberative Preferences: A person deliberates and (ideally) ranks all the possible "outcomes"

Are preferences over *outcomes* or *options*?

Preliminaries: Orderings

An ordering is a *relation* R on a set X: a subset of the set of pairs of elements from X: $R \subseteq X \times X$

Write aRb iff $(a, b) \in R$

Preliminaries: Orderings

An ordering is a *relation* R on a set X: a subset of the set of pairs of elements from X: $R \subseteq X \times X$

Write aRb iff $(a, b) \in R$

Properties of orderings:

- Reflexivity: for all $a \in X$, aRa
- Transitivity: for all $a, b, c \in X$, aRb and bRc then aRc
- Symmetry: for all $a, b \in X$, aRb implies bRa
- Asymmtery: for all $a, b \in X$, aRb implies not-bRa
- Completeness: for all $a, b \in X$, aRb or bRa (or $a = b$)

Preliminaries: Orderings

Let X be the set of outcomes (or options) and $\succeq$ an ordering ($\succeq \subseteq X \times X$).

Preliminaries: Orderings

Let X be the set of outcomes (or options) and $\succeq$ an ordering ($\succeq \subseteq X \times X$).

Given two outcomes $x, y \in X$, there are four possibilities:

Preliminaries: Orderings

Let X be the set of outcomes (or options) and $\succeq$ an ordering ($\succeq \subseteq X \times X$).

Given two outcomes $x, y \in X$, there are four possibilities:

1. $x \succeq y$ and $y \not\succeq x$: The agent *strictly prefers* x to y ($x \succ y$)

Preliminaries: Orderings

Let X be the set of outcomes (or options) and $\succeq$ an ordering ($\succeq \subseteq X \times X$).

Given two outcomes $x, y \in X$, there are four possibilities:

1. $x \succeq y$ and $y \not\succeq x$: The agent *strictly prefers* x to y ($x \succ y$)
2. $y \succeq x$ and $x \not\succeq y$: The agent *strictly prefers* y to x ($y \succ x$)

Preliminaries: Orderings

Let X be the set of outcomes (or options) and $\succeq$ an ordering ($\succeq \subseteq X \times X$).

Given two outcomes $x, y \in X$, there are four possibilities:

1. $x \succeq y$ and $y \not\succeq x$: The agent *strictly prefers* x to y ($x \succ y$)
2. $y \succeq x$ and $x \not\succeq y$: The agent *strictly prefers* y to x ($y \succ x$)
3. $x \succeq y$ and $y \succeq x$: The agent is *indifferent* between x and y ($x \approx y$)

Preliminaries: Orderings

Let X be the set of outcomes (or options) and $\succeq$ an ordering ($\succeq \subseteq X \times X$).

Given two outcomes $x, y \in X$, there are four possibilities:

1. $x \succeq y$ and $y \not\succeq x$: The agent *strictly prefers* x to y ($x \succ y$)
2. $y \succeq x$ and $x \not\succeq y$: The agent *strictly prefers* y to x ($y \succ x$)
3. $x \succeq y$ and $y \succeq x$: The agent is *indifferent* between x and y ($x \approx y$)
4. $x \not\succeq y$ and $y \not\succeq x$: The agent *cannot compare* x and y ($x \perp y$)

Preliminaries: Utility Function

A **utility function** on a set X is a function $u : X \to \mathfrak{R}$

Preliminaries: Utility Function

A **utility function** on a set X is a function $u : X \rightarrow \mathfrak{R}$

The agent prefers x to y according to u provided $u(x) \geq u(y)$

Preliminaries: Utility Function

A **utility function** on a set X is a function $u : X \to \Re$

The agent prefers x to y according to u provided $u(x) \geq u(y)$

What properties does this preference ordering have?

Ordinal Utility Theory: Axioms

Ordinal Utility Theory: Axioms

1. The ordering is complete: the agent call always rank options (for any two options x and y, either (1) the agent strictly prefers x to y, (2) strictly prefers y to x or (3) is indifferent between x and y).

Ordinal Utility Theory: Axioms

1. The ordering is complete: the agent call always rank options (for any two options x and y, either (1) the agent strictly prefers x to y, (2) strictly prefers y to x or (3) is indifferent between x and y).

2. Strict preference is asymmetric: it is not the case that the agent strictly prefers x to y *and* strictly prefers y to x

Ordinal Utility Theory: Axioms

1. The ordering is complete: the agent call always rank options (for any two options x and y, either (1) the agent strictly prefers x to y, (2) strictly prefers y to x or (3) is indifferent between x and y).

2. Strict preference is asymmetric: it is not the case that the agent strictly prefers x to y *and* strictly prefers y to x

3. Weak preference is reflexive: the agent always thinks x is at least as good as x.

Ordinal Utility Theory: Axioms

1. The ordering is complete: the agent call always rank options (for any two options x and y, either (1) the agent strictly prefers x to y, (2) strictly prefers y to x or (3) is indifferent between x and y).

2. Strict preference is asymmetric: it is not the case that the agent strictly prefers x to y *and* strictly prefers y to x

3. Weak preference is reflexive: the agent always thinks x is at least as good as x.

4. Weak preference (and hence strict and indifference) is transitive

Ordinal Utility Theory: Axioms

1. The ordering is complete: the agent call always rank options (for any two options x and y, either (1) the agent strictly prefers x to y, (2) strictly prefers y to x or (3) is indifferent between x and y).
2. Strict preference is asymmetric: it is not the case that the agent strictly prefers x to y *and* strictly prefers y to x
3. Weak preference is reflexive: the agent always thinks x is at least as good as x.
4. Weak preference (and hence strict and indifference) is transitive

Why should we accept these axioms?

Why should we accept these axioms?

"Rather than trying to provide instrumental or pragmatic justifications for the axioms of ordinal utility, it is better...to see them as constitutive of our conception of a fully rational agent....those disposed to blatantly ignore transitivity are unintelligible to use: we can't understand their pattern of actions as sensible" (Gaus [OPPE], pg. 39)

Ordinal Utility Theory

Fact. Suppose that X is finite and $\succeq$ is a complete and transitive ordering over X, then there is a utility function $u : X \rightarrow \mathfrak{R}$ that represents $\succeq$ ($x \succeq y$ iff $u(x) \geq u(y)$)

Ordinal Utility Theory

Fact. Suppose that X is finite and $\succeq$ is a complete and transitive ordering over X, then there is a utility function $u : X \rightarrow \mathfrak{R}$ that represents $\succeq$ ($x \succeq y$ iff $u(x) \geq u(y)$)

Utility is *defined* in terms of preference (so it is an error to say that the agent prefers x to y *because* she assigns a higher utility to x than to y).

Ordinal Utility Theory

Fact. Suppose that X is finite and $\succeq$ is a complete and transitive ordering over X, then there is a utility function $u : X \rightarrow \mathfrak{R}$ that represents $\succeq$ ($x \succeq y$ iff $u(x) \geq u(y)$)

Utility is *defined* in terms of preference (so it is an error to say that the agent prefers x to y *because* she assigns a higher utility to x than to y).

Important point: consider $x \succ y \succ z$

Ordinal Utility Theory

Fact. Suppose that X is finite and $\succeq$ is a complete and transitive ordering over X, then there is a utility function $u : X \to \Re$ that represents $\succeq$ ($x \succeq y$ iff $u(x) \geq u(y)$)

Utility is *defined* in terms of preference (so it is an error to say that the agent prefers x to y *because* she assigns a higher utility to x than to y).

Important point: consider $x \succ y \succ z$, all three utility functions represent this ordering:

Preference	u_1	u_2	u_3
x	3	10	1000
y	2	5	99
z	1	0	1

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 y 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 y 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 y 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 y 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 y 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 y 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 y 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!*

Next week: more about utility theory