Rationality
Lecture 6

Eric Pacuit

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

February 28, 2011

Subjective Probabilities

Should a rational agent's graded beliefs satisfy the laws of probability?

J. Joyce. *Bayesianism*. in [HR].

Ann: "the probability it will rain tomorrow is 0.9" means "Ann's degree of belief is fairly high (0.9) that it will rain tomorrow. Of course whether it will actually rain, depends on objective events taking place in the external worlds."

Ramsey, de Finetti and Savage (1)

How do we *measure* a (rational) agent's subjective probabilities?

Ramsey, de Finetti and Savage (1)

How do we *measure* a (rational) agent's subjective probabilities?

Suppose we are wondering about Ann's degree of belief about whether a coin will land heads (H) or tails (T).

Ramsey, de Finetti and Savage (1)

How do we *measure* a (rational) agent's subjective probabilities?

Suppose we are wondering about Ann's degree of belief about whether a coin will land heads (H) or tails (T).

Why don't we just *ask* her?

Ramsey, de Finetti and Savage (1)

How do we *measure* a (rational) agent's subjective probabilities?

Suppose we are wondering about Ann's degree of belief about whether a coin will land heads (H) or tails (T).

Why don't we just *ask* her? reported vs. "actual" degrees of belief.

Ramsey, de Finetti and Savage (1)

How do we *measure* a (rational) agent's subjective probabilities?

Suppose we are wondering about Ann's degree of belief about whether a coin will land heads (H) or tails (T).

Why don't we just *ask* her? reported vs. "actual" degrees of belief.

What we need: systematic procedures for linking the probability calculus (graded beliefs) to claims about objectively observable behavior, such as preferences revealed by choice behavior.

Ramsey, de Finetti and Savage (2)

Suppose we are wondering about Ann's degree of belief about whether a coin will land heads (H) or tails (T).

Offer Ann two bets:

- L_1 If the coin lands heads, you win a sports car; otherwise you win nothing
- L_2 If the coin does not land heads, you win a sports car; otherwise you win nothing.

Ramsey, de Finetti and Savage (2)

Suppose we are wondering about Ann's degree of belief about whether a coin will land heads (H) or tails (T).

Offer Ann two bets:

L_1 If the coin lands heads, you win a sports car; otherwise you win nothing

L_2 If the coin does not land heads, you win a sports car; otherwise you win nothing.

If Ann chooses L_1, she believes H is more probable than T
If Ann chooses L_2, she believes T is more probable than H
If Ann is indifferent, she believes H and T are equally probable (i.e., $p_A(H) = p_A(T) = 1/2$)

The Dutch Book Argument

But, why *should* a rational agent's graded beliefs satisfy the Kolmogorov axioms?

The Dutch Book Argument

But, why *should* a rational agent's graded beliefs satisfy the Kolmogorov axioms?

Anyone whose beliefs violate the laws of probability is *practically irrational*.

F. P. Ramsey. *Truth and Probability*. 1931.

B. de Finetti. *La prévision: Ses lois logiques, ses sources subjectives*. 1937.

Alan Hájek. *Dutch Book Arguments*. Oxford Handbook of Rational and Social Choice, 2008.

The Dutch Book Argument

1. **Coherence** A practically rational agent will never freely perform any action when another act is certain to leave her better off in *all* possible circumstances

The Dutch Book Argument

1. **Coherence** A practically rational agent will never freely perform any action when another act is certain to leave her better off in *all* possible circumstances
2. **Belief/Desire Psychology** A practically rational agent will always act in ways that she estimates will best satisfy her desires

The Dutch Book Argument

1. **Coherence** A practically rational agent will never freely perform any action when another act is certain to leave her better off in *all* possible circumstances
2. **Belief/Desire Psychology** A practically rational agent will always act in ways that she estimates will best satisfy her desires
3. **The EU-Thesis** A practically rational agent will estimate that an act best satisfies her desires iff that act maximizes her subjective expected utility

The Dutch Book Argument

1. **Coherence** A practically rational agent will never freely perform any action when another act is certain to leave her better off in *all* possible circumstances
2. **Belief/Desire Psychology** A practically rational agent will always act in ways that she estimates will best satisfy her desires
3. **The EU-Thesis** A practically rational agent will estimate that an act best satisfies her desires iff that act maximizes her subjective expected utility
4. **Dutch Book Theorem**. An agent who tries to maximize her subjective expected utility using beliefs that violate the laws of probability will freely preform an act that is sure to leave her worse off than some alternative act would in all circumstances.

Belief/Desire Psychology

Belief/Desire Psychology

But, there is ample evidence that actions are affected by emotions, habits, decision-making heuristics, and judgmental bias....

Belief/Desire Psychology

But, there is ample evidence that actions are affected by emotions, habits, decision-making heuristics, and judgmental bias....

Bayesianism is a **normative** theory of **rational** behavior not an empirical theory of actual behavior.

Belief/Desire Psychology

But, there is ample evidence that actions are affected by emotions, habits, decision-making heuristics, and judgmental bias....

Bayesianism is a **normative** theory of **rational** behavior not an empirical theory of actual behavior.

Emotions, habits, etc. do *cause actions*, BUT the DBA does not rely on the belief-desire psychology as a causal theory of action.

Belief/Desire Psychology

But, there is ample evidence that actions are affected by emotions, habits, decision-making heuristics, and judgmental bias....

Bayesianism is a **normative** theory of **rational** behavior not an empirical theory of actual behavior.

Emotions, habits, etc. do *cause actions*, BUT the DBA does not rely on the belief-desire psychology as a causal theory of action.

Assumption 2 makes no claim abut what causes actions....it says that what *makes* an act rational is that it bears the right relationship to the actor's beliefs and desires.

The EU-Thesis

Expected Money/Value/Utility: Given an agent's beliefs and desires, the **expected utility** of an **action** leading to a set of outcomes *Out* is:

$$\sum_{o \in Out} [\text{how likely the act will lead to } o] \times [\text{how much the agent desires } o]$$

The EU-Thesis

Expected Money/Value/Utility: Given an agent's beliefs and desires, the **expected utility** of an **action** leading to a set of outcomes *Out* is:

$$\sum_{o \in Out} \text{[how likely the act will lead to } o\text{]} \times \text{[how much the agent desires } o\text{]}$$

1. principle of maximizing expected monetary value
2. principle of maximizing expected value
3. principle of maximizing expected utility

Comments on Expected Utility

Options	1/2	1/2
L_1	$1M$	$1M$
L_2	$3M$	$0M$

Comments on Expected Utility

Options	1/2	1/2
L_1	$1M$	$1M$
L_2	$3M$	$0M$

$EVM(L_1) = 1/2 \cdot 1 + 1/2 \cdot 1 = 1$
$EVM(L_1) = 1/2 \cdot 3 + 1/2 \cdot 0 = 1.5$

Comments on Expected Utility

Options	1/2	1/2
L_1	$1M$	$1M$
L_2	$3M$	$0M$

$EVM(L_1) = 1/2 \cdot 1 + 1/2 \cdot 1 = 1$
$EVM(L_1) = 1/2 \cdot 3 + 1/2 \cdot 0 = 1.5$

What numbers should we use in place of monetary value?

Comments on Expected Utility

Options	1/2	1/2
L_1	$1M$	$1M$
L_2	$3M$	$0M$

$EVM(L_1) = 1/2 \cdot 1 + 1/2 \cdot 1 = 1$
$EVM(L_1) = 1/2 \cdot 3 + 1/2 \cdot 0 = 1.5$

What numbers should we use in place of monetary value? (moral) value? personal utility?

Why maximize expected utility?

Law of Large Numbers: everyone who maximizes expected utility will *almost certainly* be better off in the long run. By performing a random experiment sufficiently many times, the probability that the average outcome differs from the expected outcome can be rendered *arbitrarily* small.

Why maximize expected utility?

Law of Large Numbers: everyone who maximizes expected utility will *almost certainly* be better off in the long run. By performing a random experiment sufficiently many times, the probability that the average outcome differs from the expected outcome can be rendered *arbitrarily* small.

Gambler's Ruin: Suppose Ann and Bob start with 1000EUR each and flip a fair coin.

Why maximize expected utility?

Law of Large Numbers: everyone who maximizes expected utility will *almost certainly* be better off in the long run. By performing a random experiment sufficiently many times, the probability that the average outcome differs from the expected outcome can be rendered *arbitrarily* small.

Gambler's Ruin: Suppose Ann and Bob start with 1000EUR each and flip a fair coin. Ann gives Bob 1EUR if H and Bob gives Ann 1EUR if T.

Why maximize expected utility?

Law of Large Numbers: everyone who maximizes expected utility will *almost certainly* be better off in the long run. By performing a random experiment sufficiently many times, the probability that the average outcome differs from the expected outcome can be rendered *arbitrarily* small.

Gambler's Ruin: Suppose Ann and Bob start with 1000EUR each and flip a fair coin. Ann gives Bob 1EUR if H and Bob gives Ann 1EUR if T. If they flip the coin a *sufficiently* large number of times, each player is *guaranteed* to face a sequence of flips that bankrupts them.

Why maximize expected utility?

Law of Large Numbers: everyone who maximizes expected utility will *almost certainly* be better off in the long run. By performing a random experiment sufficiently many times, the probability that the average outcome differs from the expected outcome can be rendered *arbitrarily* small.

Gambler's Ruin: Suppose Ann and Bob start with 1000EUR each and flip a fair coin. Ann gives Bob 1EUR if H and Bob gives Ann 1EUR if T. If they flip the coin a *sufficiently* large number of times, each player is *guaranteed* to face a sequence of flips that bankrupts them. The player that faces such a sequence first, will never have an opportunity to feel the effects of the Law of Large Numbers.

Why maximize expected utility?

Law of Large Numbers: everyone who maximizes expected utility will *almost certainly* be better off in the long run. By performing a random experiment sufficiently many times, the probability that the average outcome differs from the expected outcome can be rendered *arbitrarily* small.

Gambler's Ruin: Suppose Ann and Bob start with 1000EUR each and flip a fair coin. Ann gives Bob 1EUR if H and Bob gives Ann 1EUR if T. If they flip the coin a *sufficiently* large number of times, each player is *guaranteed* to face a sequence of flips that bankrupts them. The player that faces such a sequence first, will never have an opportunity to feel the effects of the Law of Large Numbers.

Direct axiomatizations of expected utility vs. *indirect* axiomatization of expected utility

Returning to the Dutch Book Argument

Assumptions:

1. the agent desires *only* money
2. her desire for money does not vary with changes in her fortune
3. she is not averse to risk or uncertainty

Betting Behavior

the *EU*-thesis entails that a person satisfying 1-3 will reveal the strengths of her beliefs in her betting behavior.

Betting Behavior

the *EU*-thesis entails that a person satisfying 1-3 will reveal the strengths of her beliefs in her betting behavior.

A **wager**: $W_X = [\ a \text{ if } X,\ b \text{ otherwise}]$: "you get a EUR if X is true and b EUR otherwise.
(X's truth does not depend causally on W)

Betting Behavior

the *EU*-thesis entails that a person satisfying 1-3 will reveal the strengths of her beliefs in her betting behavior.

A **wager**: $W_X = [\ a \text{ if } X, b \text{ otherwise}]$: "you get a EUR if X is true and b EUR otherwise.
(X's truth does not depend causally on W)

The EU-thesis entails that the agent's level of confidence in X will be revealed by the monetary value she puts on W_X.

Betting Behavior

fair price f **for** W_X: the sum of money at which she is indifferent between receiving a payment of f EUR or having W_X go into effect.

Betting Behavior

fair price f for W_X: the sum of money at which she is indifferent between receiving a payment of f EUR or having W_X go into effect.

$f = ExpVal(W_X) = C(X) \cdot a + (1 - C(X)) \cdot b$ implies $C(X) = \frac{f-b}{a-b}$

Betting Behavior

fair price f for W_X: the sum of money at which she is indifferent between receiving a payment of f EUR or having W_X go into effect.

$f = ExpVal(W_X) = C(X) \cdot a + (1 - C(X)) \cdot b$ implies $C(X) = \frac{f-b}{a-b}$

If she is indifferent between 63,81 EUR and
[100 EUR if it rains, 0 EUR otherwise], then she believes to degree 0.6381 that it will rain.

Dutch Book

An agent will swap an (set of) wagers with the (sum of) their fair prices.

Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Dutch Book

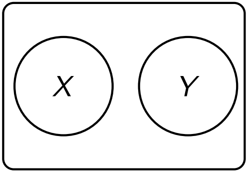
Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

$$W_X = [1 \text{ if } X, 0 \text{ else}]$$
$$W_Y = [1 \text{ if } Y, 0 \text{ else}]$$
$$W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$$

Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

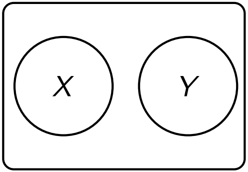
$$W_X = [1 \text{ if } X, 0 \text{ else}]$$
$$W_Y = [1 \text{ if } Y, 0 \text{ else}]$$
$$W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$$



Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$
Fair price: $f = 0.25$ for $W_Y = [1 \text{ if } Y, 0 \text{ else}]$
Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

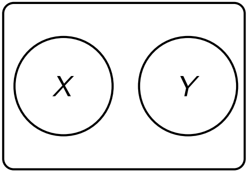


Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$
Fair price: $f = 0.25$ for $W_Y = [1 \text{ if } Y, 0 \text{ else}]$
Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

Consider $\mathcal{W}_1 = \{0.6, W_X, W_Y\}$ and $\mathcal{W}_2 = \{0.5, W_{X \vee Y}\}$



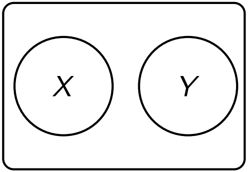
Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$
Fair price: $f = 0.25$ for $W_Y = [1 \text{ if } Y, 0 \text{ else}]$
Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

Consider $\mathcal{W}_1 = \{0.6, W_X, W_Y\}$ and $\mathcal{W}_2 = \{0.5, W_{X \vee Y}\}$

- indifferent between $\mathcal{W}_1$ and $\mathcal{W}_2$



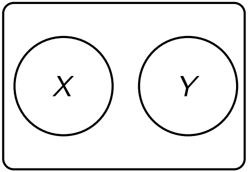
Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$
Fair price: $f = 0.25$ for $W_Y = [1 \text{ if } Y, 0 \text{ else}]$
Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

Consider $\mathcal{W}_1 = \{0.6, W_X, W_Y\}$ and $\mathcal{W}_2 = \{0.5, W_{X \vee Y}\}$

- indifferent between $\mathcal{W}_1$ and $\mathcal{W}_2$
- swap $\mathcal{W}_1$ for $\mathcal{W}_2$

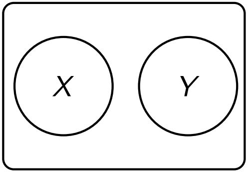


Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$
Fair price: $f = 0.25$ for $W_Y = [1 \text{ if } Y, 0 \text{ else}]$
Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

Consider $\mathcal{W}_1 = \{0.6, W_X, W_Y\}$ and $\mathcal{W}_2 = \{0.5, W_{X \vee Y}\}$



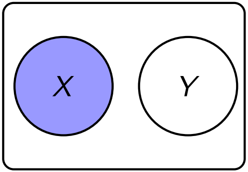
- indifferent between $\mathcal{W}_1$ and $\mathcal{W}_2$
- swap $\mathcal{W}_1$ for $\mathcal{W}_2$
- But $\mathcal{W}_1$ is always better:

Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$
Fair price: $f = 0.25$ for $W_Y = [1 \text{ if } Y, 0 \text{ else}]$
Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

Consider $\mathcal{W}_1 = \{0.6, W_X, W_Y\}$ and $\mathcal{W}_2 = \{0.5, W_{X \vee Y}\}$



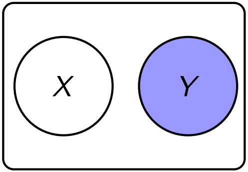
- indifferent between $\mathcal{W}_1$ and $\mathcal{W}_2$
- swap $\mathcal{W}_1$ for $\mathcal{W}_2$
- But $\mathcal{W}_1$ is always better:
  - If X is true
    payoff($\mathcal{W}_1$) $= 1.6 >$ payoff($\mathcal{W}_2$)$=1.5$

Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$
Fair price: $f = 0.25$ for $W_Y = [1 \text{ if } Y, 0 \text{ else}]$
Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

Consider $\mathcal{W}_1 = \{0.6, W_X, W_Y\}$ and $\mathcal{W}_2 = \{0.5, W_{X \vee Y}\}$



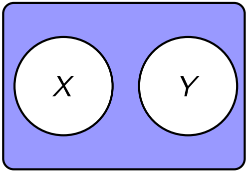
- indifferent between $\mathcal{W}_1$ and $\mathcal{W}_2$
- swap $\mathcal{W}_1$ for $\mathcal{W}_2$
- But $\mathcal{W}_1$ is always better:
  - If X is true
    payoff($\mathcal{W}_1$) $= 1.6 >$ payoff($\mathcal{W}_2$)$=1.5$
  - If Y is true
    payoff($\mathcal{W}_1$) $= 1.6 >$ payoff($\mathcal{W}_2$)$= 1.5$

Dutch Book

Suppose that X and Y are logically incompatible ($X \cap Y = \emptyset$)

Fair price: $f = 0.25$ for $W_X = [1 \text{ if } X, 0 \text{ else}]$
Fair price: $f = 0.25$ for $W_Y = [\ 1 \text{ if } Y, 0 \text{ else}]$
Fair price: $f = 0.6$ for $W_{X \vee Y} = [1 \text{ if } X \vee Y, 0 \text{ else}]$

Consider $\mathcal{W}_1 = \{0.6, W_X, W_Y\}$ and $\mathcal{W}_2 = \{0.5, W_{X \vee Y}\}$



- indifferent between $\mathcal{W}_1$ and $\mathcal{W}_2$
- swap $\mathcal{W}_1$ for $\mathcal{W}_2$
- But $\mathcal{W}_1$ is always better:
  - If X is true
    payoff($\mathcal{W}_1$) $= 1.6 >$ payoff($\mathcal{W}_2$)$=1.5$
  - If Y is true
    payoff($\mathcal{W}_1$) $= 1.6 >$ payoff($\mathcal{W}_2$)$= 1.5$
  - If neither X nor Y is true
    payoff($\mathcal{W}_1$) $= 0.6 >$ payoff($\mathcal{W}_2$)$=0.5$

Dutch Book Theorem

Theorem. Imagine and EU-maximizer who satisfies 1-3 and has a precise degree of belief for every proposition she considers. If these beliefs violate the laws of probability, then she will make Dutch Book against herself.

Dutch Book Theorem

Theorem. Imagine and EU-maximizer who satisfies 1-3 and has a precise degree of belief for every proposition she considers. If these beliefs violate the laws of probability, then she will make Dutch Book against herself.

This assumes there is an agent who

1. Meets conditions 1-3

Dutch Book Theorem

Theorem. Imagine and EU-maximizer who satisfies 1-3 and has a precise degree of belief for every proposition she considers. If these beliefs violate the laws of probability, then she will make Dutch Book against herself.

This assumes there is an agent who

1. Meets conditions 1-3
2. sets a fair price for every wager she considers

Dutch Book Theorem

Theorem. Imagine and EU-maximizer who satisfies 1-3 and has a precise degree of belief for every proposition she considers. If these beliefs violate the laws of probability, then she will make Dutch Book against herself.

This assumes there is an agent who

1. Meets conditions 1-3
2. sets a fair price for every wager she considers
3. maximizes expected utility

Dutch Book Theorem

Theorem. Imagine and EU-maximizer who satisfies 1-3 and has a precise degree of belief for every proposition she considers. If these beliefs violate the laws of probability, then she will make Dutch Book against herself.

This assumes there is an agent who

1. Meets conditions 1-3
2. sets a fair price for every wager she considers
3. maximizes expected utility

allow agents to have incomplete or imprecise preferences

Dutch Book Theorem

Theorem. Imagine and EU-maximizer who satisfies 1-3 and has a precise degree of belief for every proposition she considers. If these beliefs violate the laws of probability, then she will make Dutch Book against herself.

This assumes there is an agent who

1. Meets conditions 1-3
2. sets a fair price for every wager she considers
3. maximizes expected utility

justify probabilistic coherence and EU simultaneously: Savage's Representation Theorem (discussed later in the semester)

Converse Dutch Book Theorem

If you violate Φ, then you are susceptible to Dutch Book.

Converse Dutch Book Theorem

If you violate Φ, then you are susceptible to Dutch Book.
Therefore, you should *obey* Φ.

Converse Dutch Book Theorem

If you violate Φ, then you are susceptible to Dutch Book.
Therefore, you should *obey* Φ.

We need: "If you *obey* Φ, then you are *not* susceptible to a Dutch Book (or possibly you are not susecptible to a Dutch Book)."

Converse Dutch Book Theorem

If you violate Φ, then you are susceptible to Dutch Book.
Therefore, you should *obey* Φ.

We need: "If you *obey* Φ, then you are *not* susceptible to a Dutch Book (or possibly you are not susecptible to a Dutch Book)."

Converse Dutch Book Theorem. If your degrees of beliefs (fair betting prices) satisfy the laws of probability, then there does not exist a Dutch Book consisting of bets at those prices.

Issues

- Epistemic vs. Practical Rationality
- What sort of requirement is "maximize expected utility"?
- Observed failures of maximizing subjective expected utility: Allais Paradox (more on this later in the semester)
- Rational principles of belief *change*
- Is the desire/belief "ontology" *correct*? Lewis on the desire as belief thesis.

Epistemic vs. Practical Rationality

"...The fact (if it is a fact) that one will be bound to lose money unless one's degrees of belief [obey the laws of probability] just isn't relevant to the truth of those beliefs" (Kennedy and Chihara, 1979)

Epistemic vs. Practical Rationality

"...The fact (if it is a fact) that one will be bound to lose money unless one's degrees of belief [obey the laws of probability] just isn't relevant to the truth of those beliefs" (Kennedy and Chihara, 1979)

- *Behavorism*: Savage *defines* (graded) beliefs beliefs in terms of characteristic patterns of preferences for actions.

Epistemic vs. Practical Rationality

"...The fact (if it is a fact) that one will be bound to lose money unless one's degrees of belief [obey the laws of probability] just isn't relevant to the truth of those beliefs" (Kennedy and Chihara, 1979)

- *Behavorism*: Savage *defines* (graded) beliefs beliefs in terms of characteristic patterns of preferences for actions.
- *Pragmatism*: De Finetti: it is "trivial and obvious...that the degree of probability attributed by an individual to a given event is revealed by the conditions under which he would be disposed to bet on that event."

Epistemic vs. Practical Rationality

"...The fact (if it is a fact) that one will be bound to lose money unless one's degrees of belief [obey the laws of probability] just isn't relevant to the truth of those beliefs" (Kennedy and Chihara, 1979)

- *Behavorism*: Savage *defines* (graded) beliefs beliefs in terms of characteristic patterns of preferences for actions.
- *Pragmatism*: De Finetti: it is "trivial and obvious...that the degree of probability attributed by an individual to a given event is revealed by the conditions under which he would be disposed to bet on that event."

There is nothing more to the rationality of beliefs than their propensity to produce practically rational actions.

Epistemic vs. Practical Rationality

- Skyrms: "you evaluate a betting arrangement independent of how it is described."

Epistemic vs. Practical Rationality

- Skyrms: "you evaluate a betting arrangement independent of how it is described."

  The agent believes X more likely than Y and $X \vee Z$ more likely than $Y \vee Z$ and suppose that O_1 strictly preferred to O_2

Epistemic vs. Practical Rationality

- Skyrms: "you evaluate a betting arrangement independent of how it is described."

  The agent believes X more likely than Y and $X \vee Z$ more likely than $Y \vee Z$ and suppose that O_1 strictly preferred to O_2

  P1=$\{[O_1 \text{ if } X, \text{ else } O_2], [O_1 \text{ if } Z, \text{ else } O_2]\}$

  P2=$\{[O_1 \text{ if } Y, \text{ else } O_2], [O_1 \text{ if } Z, \text{ else } O_2]\}$

  W1=$[O_1 \text{ if } Y \vee Z, \text{ else } O_2]$

  W2=$[O_1 \text{ if } X \vee Z, \text{ else } O_2]$

Epistemic vs. Practical Rationality

- Skyrms: "you evaluate a betting arrangement independent of how it is described."

  The agent believes X more likely than Y and $X \vee Z$ more likely than $Y \vee Z$ and suppose that O_1 strictly preferred to O_2

  P1=$\{[O_1 \text{ if } X, \text{ else } O_2], [O_1 \text{ if } Z, \text{ else } O_2]\}$

  P2=$\{[O_1 \text{ if } Y, \text{ else } O_2], [O_1 \text{ if } Z, \text{ else } O_2]\}$

  W1=$[O_1 \text{ if } Y \vee Z, \text{ else } O_2]$

  W2=$[O_1 \text{ if } X \vee Z, \text{ else } O_2]$

  Suppose $P1 \succeq P2$ but $W_1 \succeq W_2$.

Epistemic vs. Practical Rationality

- Skyrms: "you evaluate a betting arrangement independent of how it is described."

  The agent believes X more likely than Y and $X \vee Z$ more likely than $Y \vee Z$ and suppose that O_1 strictly preferred to O_2

  P1=$\{[O_1 \text{ if } X, \text{ else } O_2], [O_1 \text{ if } Z, \text{ else } O_2]\}$

  P2=$\{[O_1 \text{ if } Y, \text{ else } O_2], [O_1 \text{ if } Z, \text{ else } O_2]\}$

  W1=$[O_1 \text{ if } Y \vee Z, \text{ else } O_2]$

  W2=$[O_1 \text{ if } X \vee Z, \text{ else } O_2]$

  Suppose $P1 \succeq P2$ but $W_1 \succeq W_2$. This is inconsistent as they are simply rephrasing of the same event!

Epistemic vs. Practical Rationality

- Skyrms: "you evaluate a betting arrangement independent of how it is described."

  The agent believes X more likely than Y and $X \vee Z$ more likely than $Y \vee Z$ and suppose that O_1 strictly preferred to O_2

  P1=$\{[O_1 \text{ if } X, \text{ else } O_2], [O_1 \text{ if } Z, \text{ else } O_2]\}$

  P2=$\{[O_1 \text{ if } Y, \text{ else } O_2], [O_1 \text{ if } Z, \text{ else } O_2]\}$

  W1=$[O_1 \text{ if } Y \vee Z, \text{ else } O_2]$

  W2=$[O_1 \text{ if } X \vee Z, \text{ else } O_2]$

  Suppose $P1 \succeq P2$ but $W_1 \succeq W_2$. This is inconsistent as they are simply rephrasing of the same event!

- Howson and Urbach: Define beliefs as the betting odds the agent *believes* are fair.

Epistemic vs. Practical Rationality

- Skyrms: "you evaluate a betting arrangement independent of how it is described."

  The agent believes X more likely than Y and $X \vee Z$ more likely than $Y \vee Z$ and suppose that O_1 strictly preferred to O_2

  P1=$\{[O_1 \text{ if } X, \text{ else } O_2], [O_1 \text{ if } Z, \text{ else } O_2]\}$

  P2=$\{[O_1 \text{ if } Y, \text{ else } O_2], [O_1 \text{ if } Z, \text{ else } O_2]\}$

  W1=$[O_1 \text{ if } Y \vee Z, \text{ else } O_2]$

  W2=$[O_1 \text{ if } X \vee Z, \text{ else } O_2]$

  Suppose $P1 \succeq P2$ but $W_1 \succeq W_2$. This is inconsistent as they are simply rephrasing of the same event!

- Howson and Urbach: Define beliefs as the betting odds the agent *believes* are fair.
- Joyce: relate probability consistency to *accuracy* of graded beliefs.

Maximizing Expected Utility

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

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.

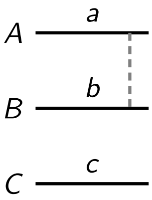
Maximizing

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.

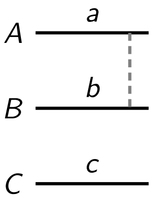
1. A is rationally obligatory iff A has a higher expectation value than any act incompatible with A
2. A is rationally permissible iff A is at least as great as that of any act incompatible with A

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

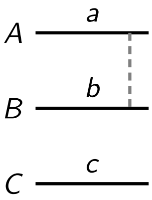


1. A is rationally obligatory iff A has a higher expectation value than any act incompatible with A
2. A is rationally permissible iff A is at least as great as that of any act incompatible with A



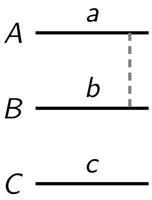
- Assume E_A, E_B and E_C are all positive

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



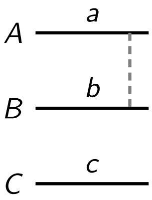
- Assume E_A, E_B and E_C are all positive
- Assume $E_A = E_B$, so by (2), both A and B are rationally permissible

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!

“It is worth emphasizing that all of these paradoxical cases are non-realistic in important ways. The wine example requires that the wine get better without limit, and do so rapidly enough to counteract the decrease in expectation value normally resulting from the extreme improbability of our surviving to very great ages (e.g., one million years). The examples illustrate logically possible difficulties for (1) and (2), and hence show that (1) and (2) are not logically necessary, but they are not difficulties that are apt to arise in practice. ” (Pollock, pg. 420)

Allais Paradox

M. Allais. *Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine*. Econometrica 21, 503-546, 1953.

Allais Paradox

Suppose there are three possible outcomes:

1. O_1 you receive \$0
2. O_2 you receive \$1M
3. O_3 you receive \$5M

A lottery is a triple (p_1, p_2, p_3) meaning the player wins O_i with probability p_i.

Allais Paradox

Suppose there are three possible outcomes:

1. O_1 you receive \$0
2. O_2 you receive \$1M
3. O_3 you receive \$5M

A lottery is a triple (p_1, p_2, p_3) meaning the player wins O_i with probability p_i.

Which lottery do you prefer?

1. $L_1 = (0.00, 1.00, 0.00)$ or $L_2 = (0.01, 0.89, 0.10)$?

Allais Paradox

Suppose there are three possible outcomes:

1. O_1 you receive \$0
2. O_2 you receive \$1M
3. O_3 you receive \$5M

A lottery is a triple (p_1, p_2, p_3) meaning the player wins O_i with probability p_i.

Which lottery do you prefer?

1. $L_1 = (0.00, 1.00, 0.00)$ or $L_2 = (0.01, 0.89, 0.10)$?
2. $L_3 = (0.90, 0.00, 0.10)$ or $(0.89, 0.11, 0.00)$?

Allais Paradox

1. O_1 you receive \$0
2. O_2 you receive \$1M
3. O_3 you receive \$5M

1. $L_1 = (0.00, 1.00, 0.00)$ or $L_2 = (0.01, 0.89, 0.10)$?
2. $L_3 = (0.90, 0.00, 0.10)$ or $(0.89, 0.11, 0.00)$?

Allais Paradox

1. O_1 you receive \$0
2. O_2 you receive \$1M
3. O_3 you receive \$5M

1. $L_1 = (0.00, 1.00, 0.00)$ or $L_2 = (0.01, 0.89, 0.10)$?
2. $L_3 = (0.90, 0.00, 0.10)$ or $(0.89, 0.11, 0.00)$?

Many subjects report $L_1 \succ L_2$ and $L_3 \succ L_4$.

Allais Paradox

1. O_1 you receive \$0
2. O_2 you receive \$1M
3. O_3 you receive \$5M

1. $L_1 = (0.00, 1.00, 0.00)$ or $L_2 = (0.01, 0.89, 0.10)$?
2. $L_3 = (0.90, 0.00, 0.10)$ or $(0.89, 0.11, 0.00)$?

Many subjects report $L_1 \succ L_2$ and $L_3 \succ L_4$.

Why does this contradict standard expected utility calculations?

Allais Paradox

1. O_1 you receive \$0
2. O_2 you receive \$1M
3. O_3 you receive \$5M

1. $L_1 = (0.00, 1.00, 0.00)$ or $L_2 = (0.01, 0.89, 0.10)$?
2. $L_3 = (0.90, 0.00, 0.10)$ or $(0.89, 0.11, 0.00)$?

Many subjects report $L_1 \succ L_2$ and $L_3 \succ L_4$.

Why does this contradict standard expected utility calculations?
(Explanation on the next slide)

Allais Paradox

If $L_1 \succ L_2$ and the decision makers is maximizing expected utility, then we have
$0.00 \cdot u_0 + 1.00 \cdot u_{1M} + 0.00 \cdot u_{5M} > 0.01 \cdot u_0 + 0.89 \cdot u_{1M} + 0.10 \cdot u_{5M}$.
So, (after some algebraic manipulations)

$$0.11 \cdot u_{1M} > 0.01 \cdot u_0 + 0.10 u_{5M}$$

If $L_3 \succ L_4$ and the decision makers is maximizing expected utility, then we have
$0.90 \cdot u_0 + 0.00 \cdot u_{1M} + 0.10 \cdot u_{5M} > 0.89 \cdot u_0 + 0.11 \cdot u_{1M} + 0.00 \cdot u_{5M}$.
So, (after some algebraic manipulations)

$$0.01 \cdot u_0 + 0.10 \cdot u_{5M} > 0.11 \cdot u_{1M}$$

Putting these inequalities together, we have

$$0.11 \cdot u_{1M} > 0.01 \cdot u_0 + 0.10 u_{5M} > 0.11 \cdot u_{1M}$$

which implies $0.11 \cdot u_{1M} > 0.11 \cdot u_{1M}$, which is a contradiction.

Rational Belief Change

If an agent believes p and $p \to q$, what should an ideally rational agent do if she comes to (reliably) believe that $\neg q$?

Rational Belief Change

If an agent believes p and $p \to q$, what should an ideally rational agent do if she comes to (reliably) believe that $\neg q$?

Given a subjective probability measure P, how should an ideally rational agent revise its probability measure upon learning an event E?

Revising Probabilities

Suppose that P is an initial probability measure and $P_E(\cdot) = P(\cdot \mid E)$ is the conditional probability measure.

Revising Probabilities

Suppose that P is an initial probability measure and $P_E(\cdot) = P(\cdot \mid E)$ is the conditional probability measure.

Simple conditioning: If a person (with $0 < P(E) < 1$) has a learning experience whose sole immediate effect is to raise her subjective probability for E to 1, then her post-learning "posterior" for any proposition H should be $P_{new}(H) = P_E(H)$.

Revising Probabilities

Richard Jeffrey has argued, the evidence we receive is often too vague or ambiguous to justify such “dogmatism.” On more realistic models, the direct effect of a learning experience will be to alter the subjective probability of some proposition without raising it to 1 or lowering it to 0.

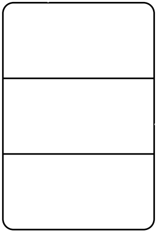
Revising Probabilities

Richard Jeffrey has argued, the evidence we receive is often too vague or ambiguous to justify such "dogmatism." On more realistic models, the direct effect of a learning experience will be to alter the subjective probability of some proposition without raising it to 1 or lowering it to 0.

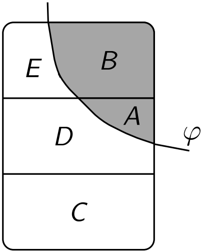
Jeffrey Conditioning: If a person (with $0 < P(E) < 1$) has a learning experience whose sole immediate effect is to change her subjective probability for E to q, then her post-learning posterior for any H should be $P_{new}(H) = qP_E(H) + (1-q)P_{-E}(H)$.

R. Jeffrey. *Alias Smith and Jones: The Testimony of the Senses*. Erkenntnis 26, 391-399, 1987.

Revising (All-out) Beliefs

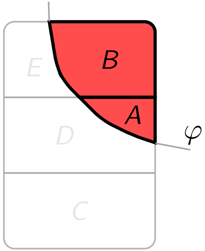


Revising (All-out) Beliefs



Incorporate the new information φ

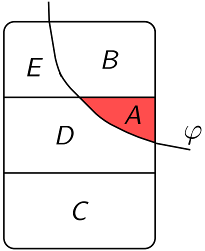
Revising (All-out) Beliefs



Public Announcement: Information from an infallible source

$(!\varphi)$: $A \prec_i B$

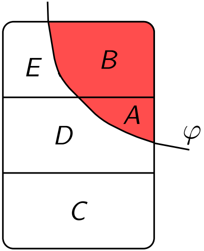
Revising (All-out) Beliefs



Public Announcement: Information from an infallible source ($!\varphi$): $A \prec_i B$

Conservative Upgrade: Information from a trusted source ($\uparrow\varphi$): $A \prec_i C \prec_i D \prec_i B \cup E$

Revising (All-out) Beliefs



Public Announcement: Information from an infallible source ($!\varphi$): $A \prec_i B$

Conservative Upgrade: Information from a trusted source ($\uparrow\varphi$): $A \prec_i C \prec_i D \prec_i B \cup E$

Radical Upgrade: Information from a strongly trusted source ($\Uparrow\varphi$): $A \prec_i B \prec_i C \prec_i D \prec_i E$

Desire as Belief Thesis

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.

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 Week: No Class (consult website for reading)

Savage's Representation Theorem

A set of states S, a set of consequences O, **acts** are functions from S to O.

Savage's Representation Theorem

A set of states S, a set of consequences O, **acts** are functions from S to O.

1. each act/state pair produces a unique consequence that settles every issue the agent cares about
2. she is convinced that her behavior will make no causal difference to which state obtains.

Savage's Representation Theorem

A set of states S, a set of consequences O, **acts** are functions from S to O.

1. each act/state pair produces a unique consequence that settles every issue the agent cares about
2. she is convinced that her behavior will make no causal difference to which state obtains.

The agent is assumed to have preference ordering $\succeq$ over the set of acts.

Savage's Representation Theorem

A set of states S, a set of consequences O, **acts** are functions from S to O.

1. each act/state pair produces a unique consequence that settles every issue the agent cares about
2. she is convinced that her behavior will make no causal difference to which state obtains.

The agent is assumed to have preference ordering $\succeq$ over the set of acts.

Expected Utility:

$$Exp_{P,u}(\alpha) = \sum_{w \in W} P(w) \times u(\alpha, w)$$

Small Worlds

States: {the sixth egg is good, the sixth egg is rotten}
Consequences { 6-egg omelet, no omelet and five good eggs destroyed, 6-egg omelet and a saucer to wash....}
Acts: { break egg into bowl, break egg into saucer, throw egg away}

Small Worlds

States: {the sixth egg is good, the sixth egg is rotten}
Consequences { 6-egg omelet, no omelet and five good eggs destroyed, 6-egg omelet and a saucer to wash....}
Acts: { break egg into bowl, break egg into saucer, throw egg away}

	Good Egg	Rotten Egg
Break into bowl	6-egg omelet	No Omelet and five good eggs destroyed
Break into saucer	6-egg omelet and a saucer to wash	5-egg omelet and a saucer to wash
Throw away	5-egg omelet and one good egg destroyed	5-egg omelet

Representation

EU-coherence: There must be at least on probability P defined on states and one utility function for consequences that **represent** the agent's preferences in the sense that, for any acts α and β, she strictly (weakly) prefers α to β only if $Exp_{P,u}(\alpha)$ is greater (as great as) $Exp_{P,u}(\beta)$.

Axioms of Preference

For all acts α, β, γ and events X, Y

Axioms of Preference

For all acts α, β, γ and events X, Y

1. **Trichotomy** One of the following is true: $\alpha \succ \beta$ or $\beta \succ \alpha$ or $\alpha \approx \beta$,

Axioms of Preference

For all acts α, β, γ and events X, Y

1. **Trichotomy** One of the following is true: $\alpha \succ \beta$ or $\beta \succ \alpha$ or $\alpha \approx \beta$,
2. **Transitivity** $\succeq$ is transitive

Axioms of Preference

For all acts α, β, γ and events X, Y

1. **Trichotomy** One of the following is true: $\alpha \succ \beta$ or $\beta \succ \alpha$ or $\alpha \approx \beta$,
2. **Transitivity** $\succeq$ is transitive
3. **Sure-thing Principle** If α and β produce the same consequences in every state consistent with X, then the agent's preference between the two acts depends only on their consequences when X obtains.

Axioms of Preference

For all acts α, β, γ and events X, Y

1. **Trichotomy** One of the following is true: $\alpha \succ \beta$ or $\beta \succ \alpha$ or $\alpha \approx \beta$,
2. **Transitivity** $\succeq$ is transitive
3. **Sure-thing Principle** If α and β produce the same consequences in every state consistent with X, then the agent's preference between the two acts depends only on their consequences when X obtains.
4. **Wagers** For consequences O_1 and O_2 and any event X, there is an act [O_1 if X, O_2 else] that produces O_1 in any state that entails X and O_2 in any state that entails $\neg X$

Axioms of Preference

For all acts α, β, γ and events X, Y

1. **Trichotomy** One of the following is true: $\alpha \succ \beta$ or $\beta \succ \alpha$ or $\alpha \approx \beta$,
2. **Transitivity** $\succeq$ is transitive
3. **Sure-thing Principle** If α and β produce the same consequences in every state consistent with X, then the agent's preference between the two acts depends only on their consequences when X obtains.
4. **Wagers** For consequences O_1 and O_2 and any event X, there is an act [O_1 if X, O_2 else] that produces O_1 in any state that entails X and O_2 in any state that entails $\neg X$
5. **Savage's P4** If the agent prefers [O_1 if X, O_2 else] to [O_1 if Y, O_2 else] when O_1 is more desirable than O_2, then she will also prefer [O_1^* if X, O_2^* else] to [O_1^* if Y, O_2^* else] for any other outcomes such that O_1^* is more desirable than O_2^*.

The Sure-Thing Principle

A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant.

The Sure-Thing Principle

A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant. So, to clarify the matter to himself, he asks whether he would buy if he knew *that the Democratic candidate were going to win, and decides that he would.*

The Sure-Thing Principle

> *A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant. So, to clarify the matter to himself, he asks whether he would buy if he* knew *that the Democratic candidate were going to win, and decides that he would. Similarly, he considers whether he would buy if he* knew *a Republican candidate were going to win, and again he finds that he would.*

The Sure-Thing Principle

> *A businessman contemplates buying a certain piece of property. He considers the outcome of the next presidential election relevant. So, to clarify the matter to himself, he asks whether he would buy if he* knew *that the Democratic candidate were going to win, and decides that he would. Similarly, he considers whether he would buy if he* knew *a Republican candidate were going to win, and again he finds that he would. Seeing that he would buy in either event, he decides that he should buy, even though he does not know which event obtains, or will obtain, as we would ordinarily say.* (Savage, 1954)

Representation Theorem

If an agent satisfies all of the above postulates (including some technical ones not discussed), then the agent acts *as if* she is maximizing an expected utility.

These axioms (along with a few others) guarantee the existence of a unique probability P and utility u, unique up to the arbitrary choice of a unit and zero-point, whose associated expectation represents the agent's preferences.

Defining Beliefs from Preferences

Savage's P4 If the agent prefers [O_1 if X, O_2 else] to [O_1 if Y, O_2 else] when O_1 is more desirable than O_2, then she will also prefer [O_1^* if X, O_2^* else] to [O_1^* if Y, O_2^* else] for any other outcomes such that O_1^* is more desirable than O_2^*.

Defining Beliefs from Preferences

Savage's P4 If the agent prefers [O_1 if X, O_2 else] to [O_1 if Y, O_2 else] when O_1 is more desirable than O_2, then she will also prefer [O_1^* if X, O_2^* else] to [O_1^* if Y, O_2^* else] for any other outcomes such that O_1^* is more desirable than O_2^*.

Definition A practically rational agent **believes** X **more strongly than she believes** Y if and only if she strictly prefers [O_1 if X, O_2 else] to [O_1 if Y, O_2 else] for some (hence any by P4) outcome with O_1 more desirable than O_2.

Defining Beliefs from Preferences

Savage's P4 If the agent prefers [O_1 if X, O_2 else] to [O_1 if Y, O_2 else] when O_1 is more desirable than O_2, then she will also prefer [O_1^* if X, O_2^* else] to [O_1^* if Y, O_2^* else] for any other outcomes such that O_1^* is more desirable than O_2^*.

Definition A practically rational agent **believes** X **more strongly than she believes** Y if and only if she strictly prefers [O_1 if X, O_2 else] to [O_1 if Y, O_2 else] for some (hence any by P4) outcome with O_1 more desirable than O_2.

If O_1 is preferred to O_2 then the agent *has a good reason* for preferring [O_1 if X, O_2 else] to [O_1 if Y, O_2 else] exactly if she is more confident in X than in Y.