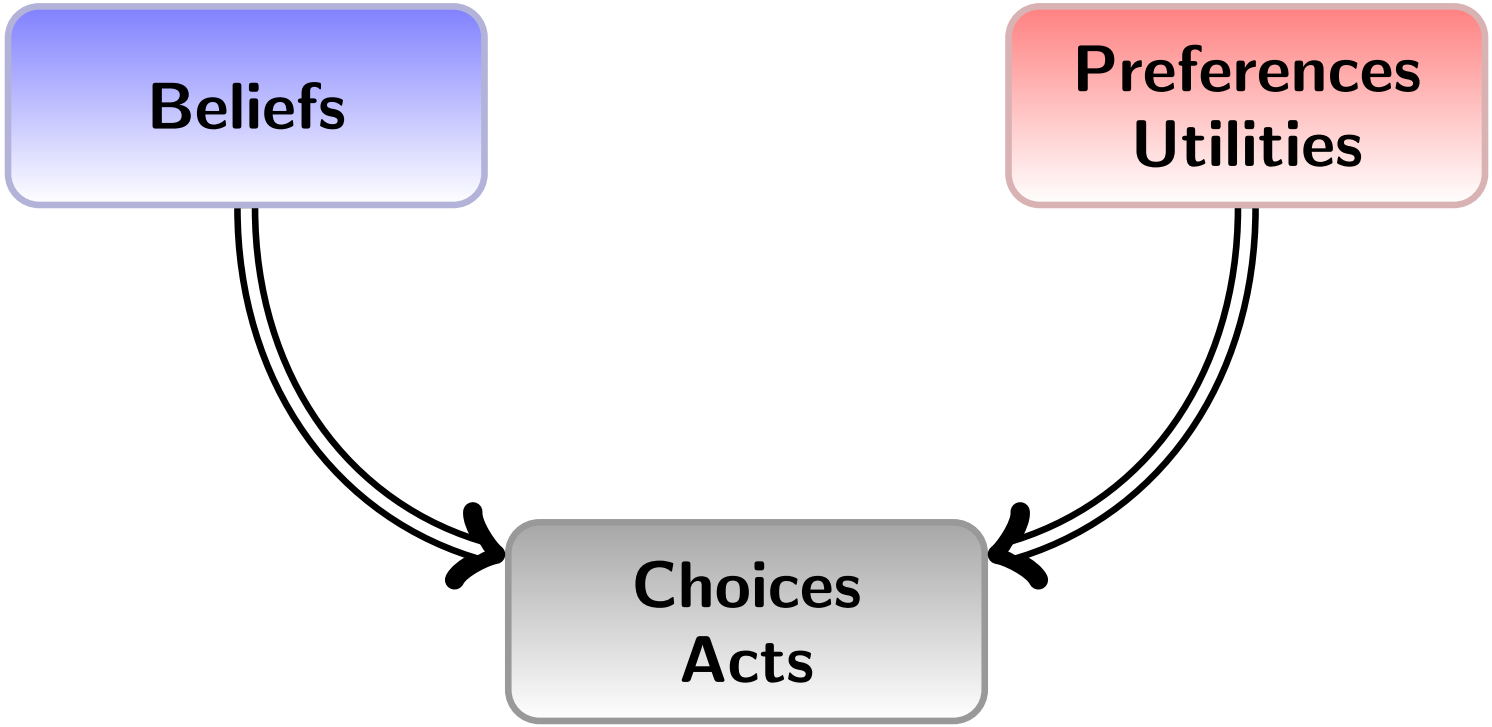
Rationality

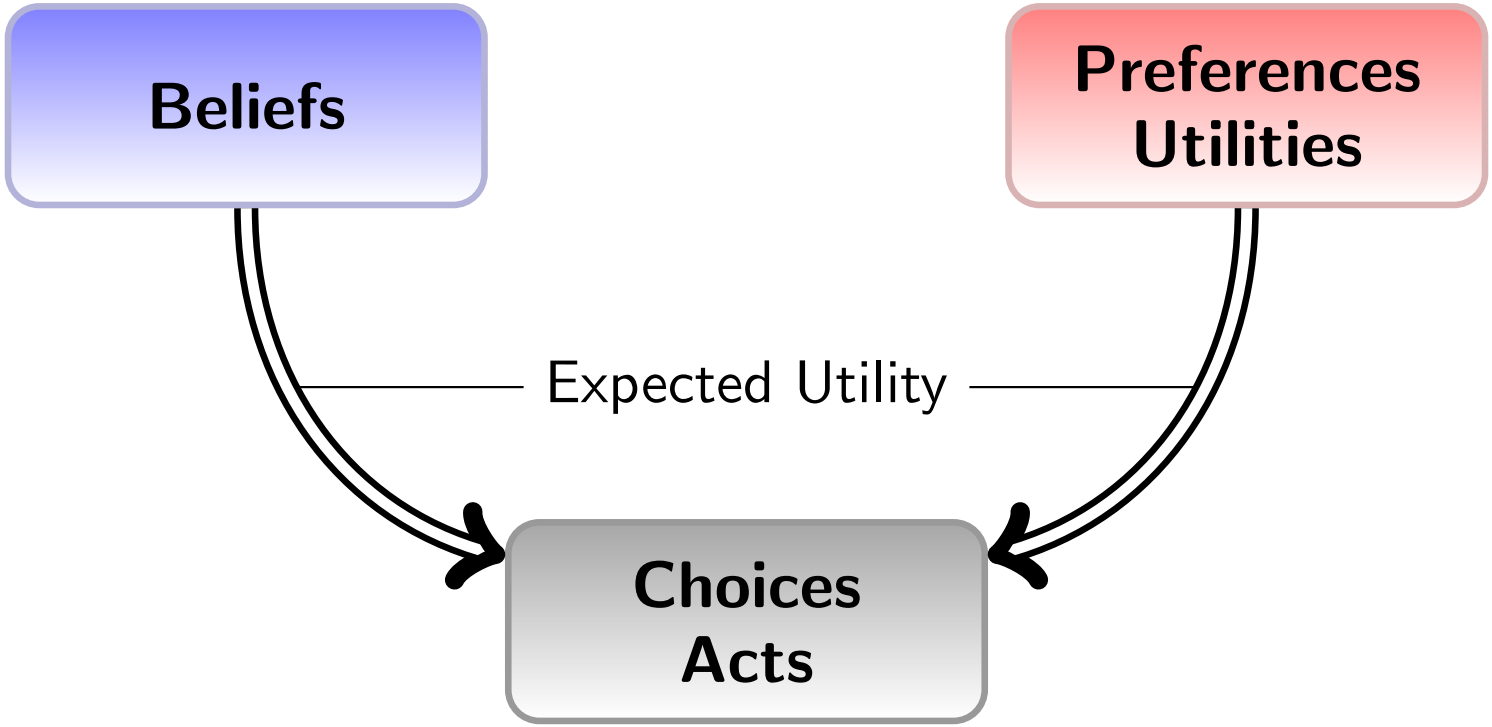
Lecture 7

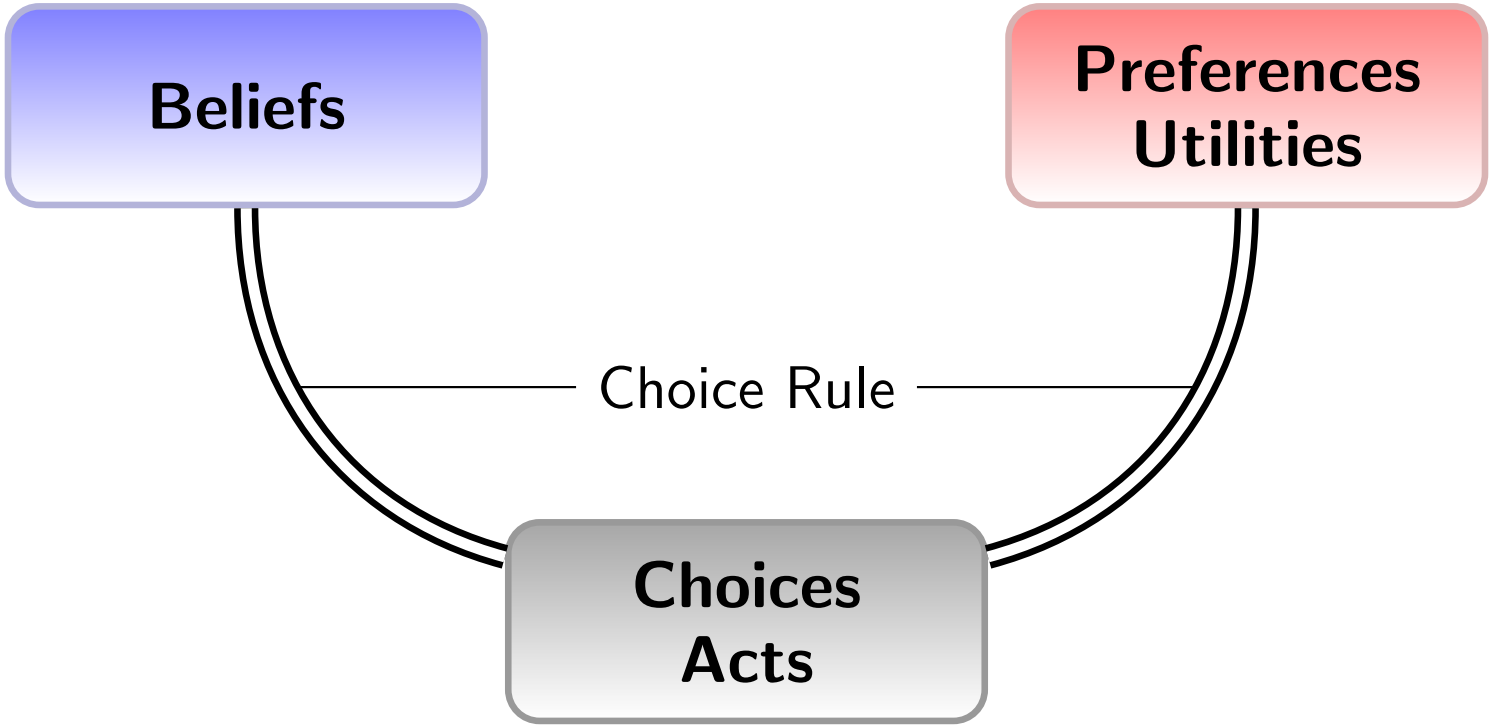
Eric Pacuit

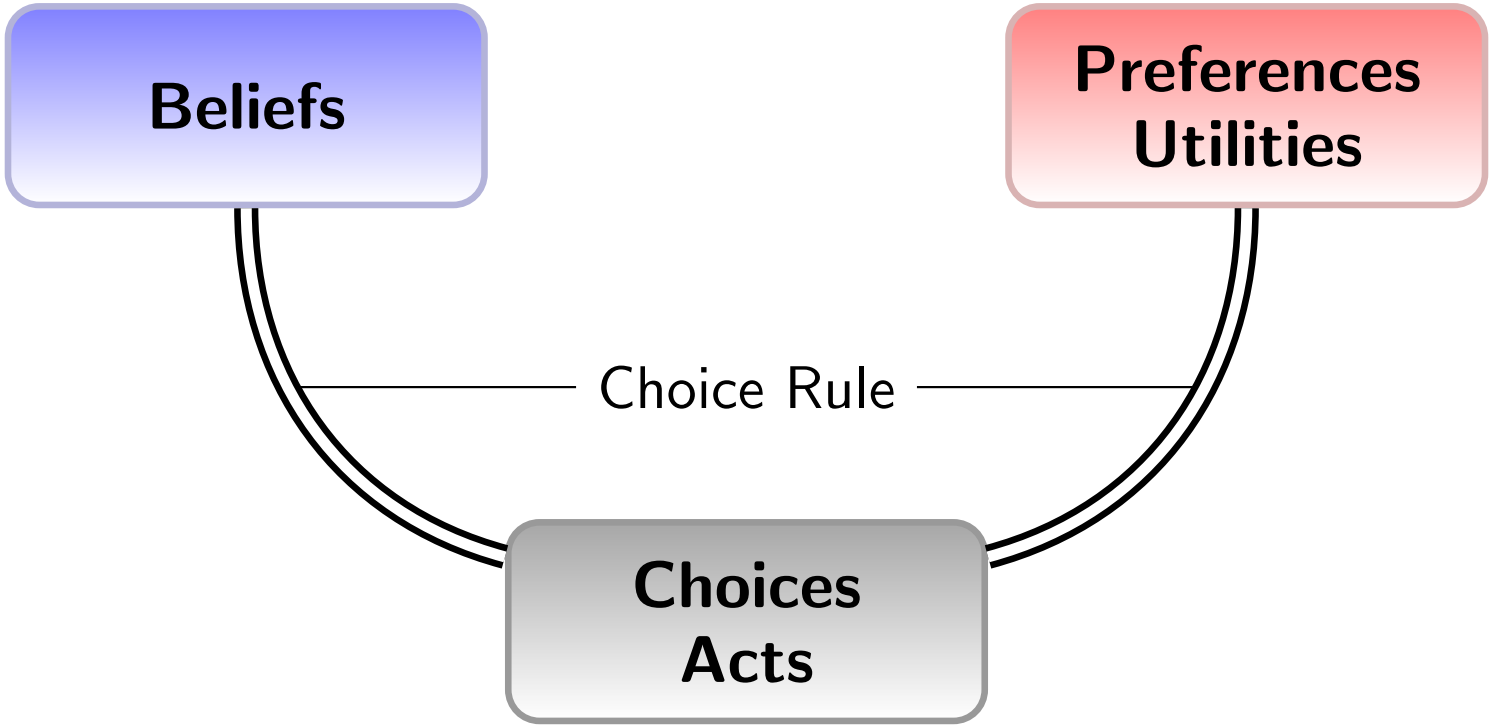
Center for Logic and Philosophy of Science
Tilburg University
ai.stanford.edu/∼epacuit
e.j.pacuit@uvt.nl

March 14, 2011


Beliefs
Preferences
Utilities
Choices
Acts


Beliefs
Preferences
Utilities
Expected Utility
Choices
Acts


Beliefs
Preferences
Utilities
Choice Rule
Choices
Acts



- Is this picture "correct"?

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.

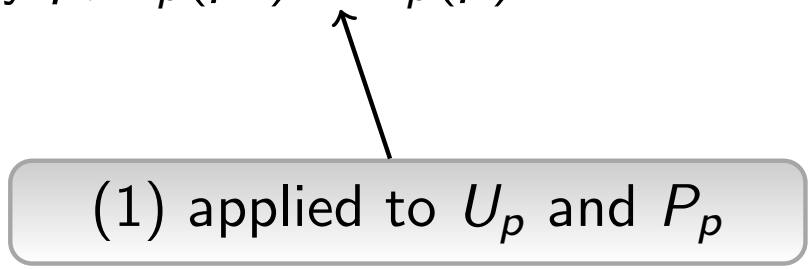
Desire-As-Belief

(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, $P_p(p^\circ) = U_p(p)$



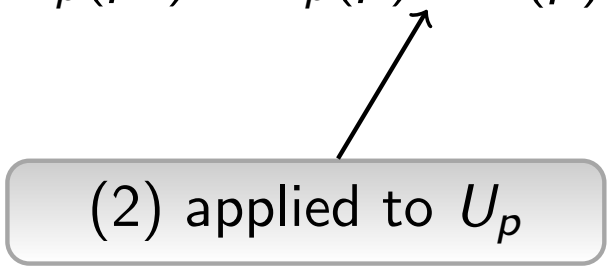
Desire-As-Belief

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



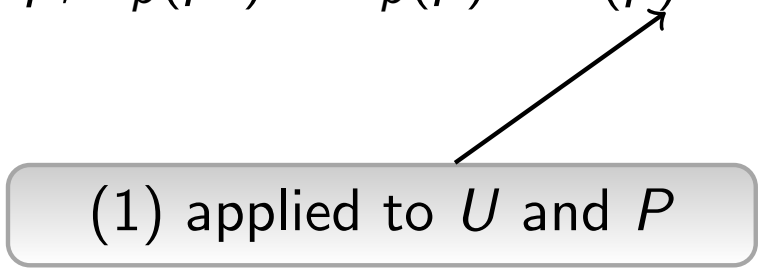
Desire-As-Belief

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



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 $P_p(p^\circ) = P(p^\circ)$)
$0 = P_q(p^\circ \mid p) \neq P_q(p^\circ) > 0$.

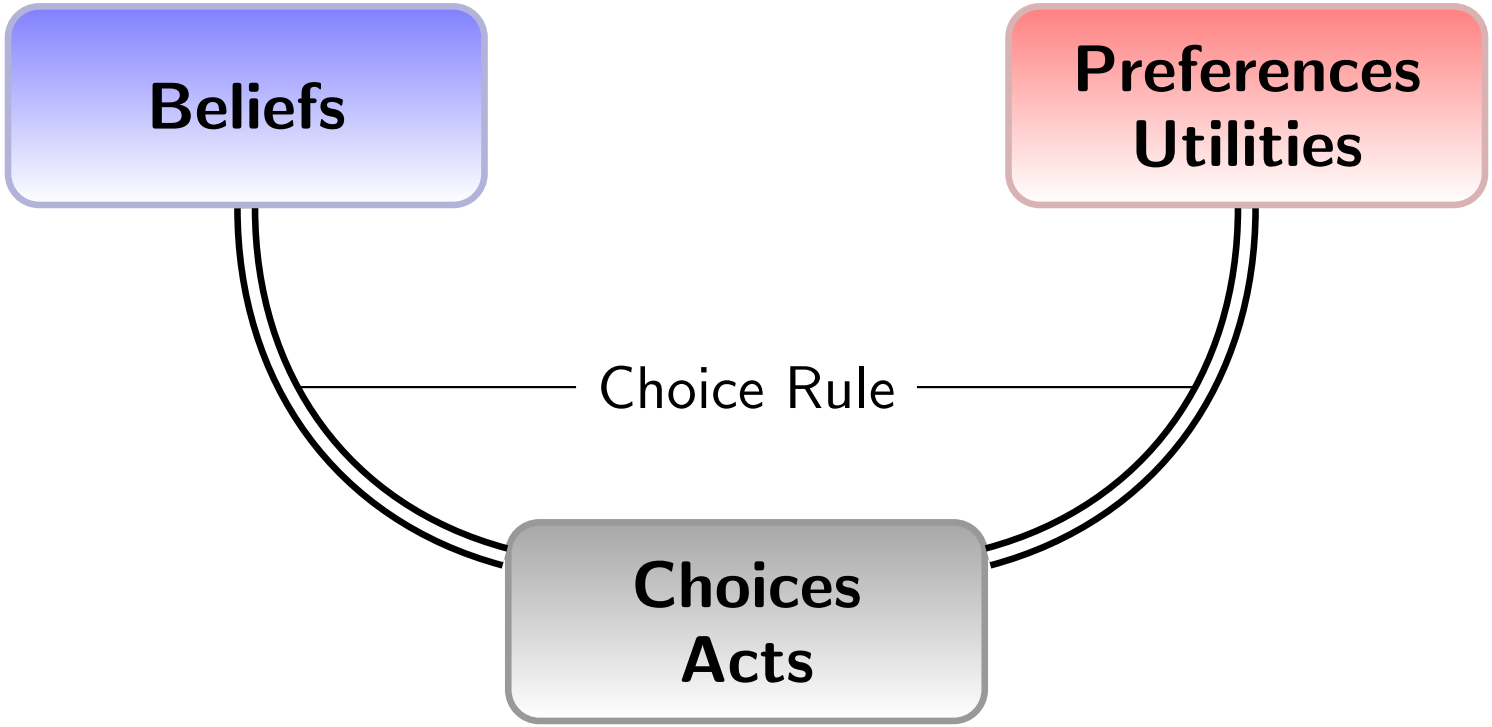
Analyzing the Argument

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

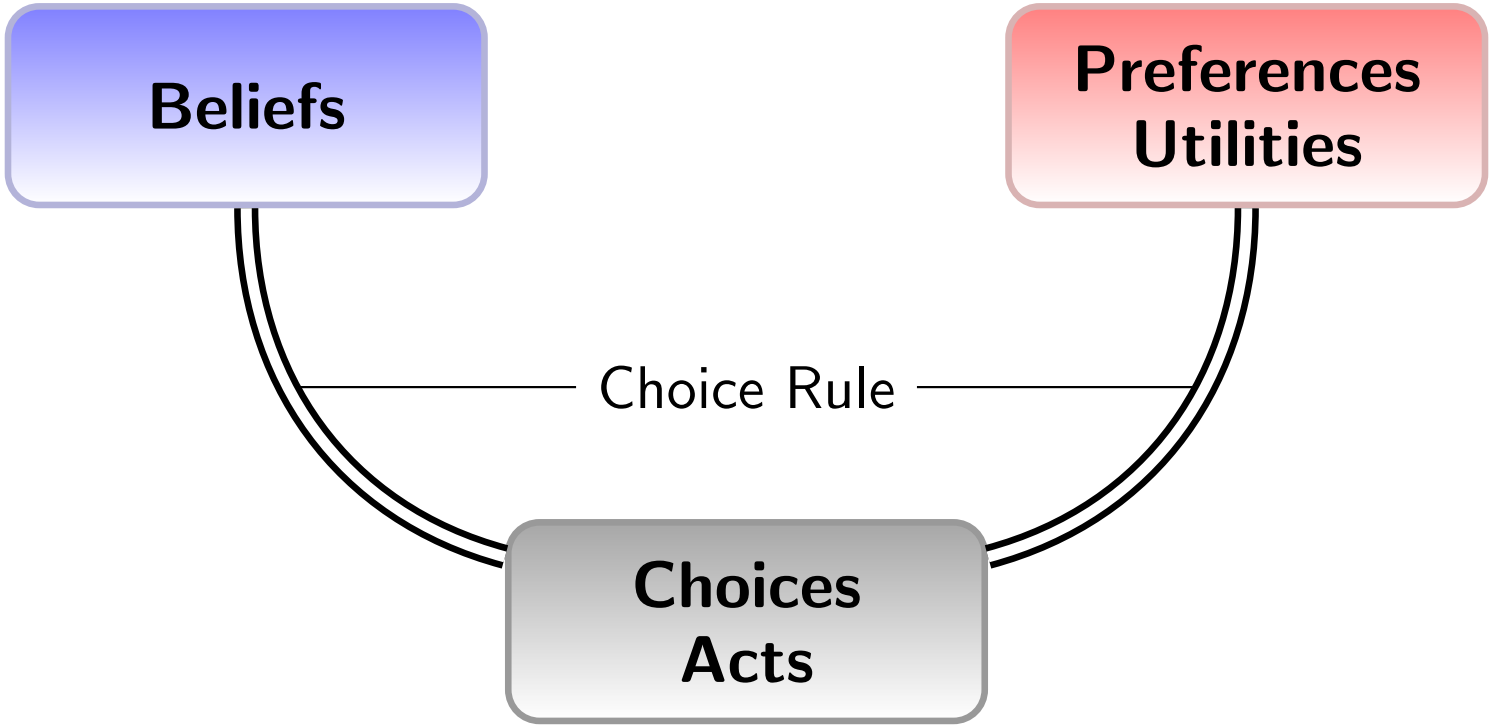
A. Hájek and P. Pettit. *Desire Beyond Belief*. Australasian Journal of Philosophy, 82(1), pgs. 77 - 92, 2004.

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

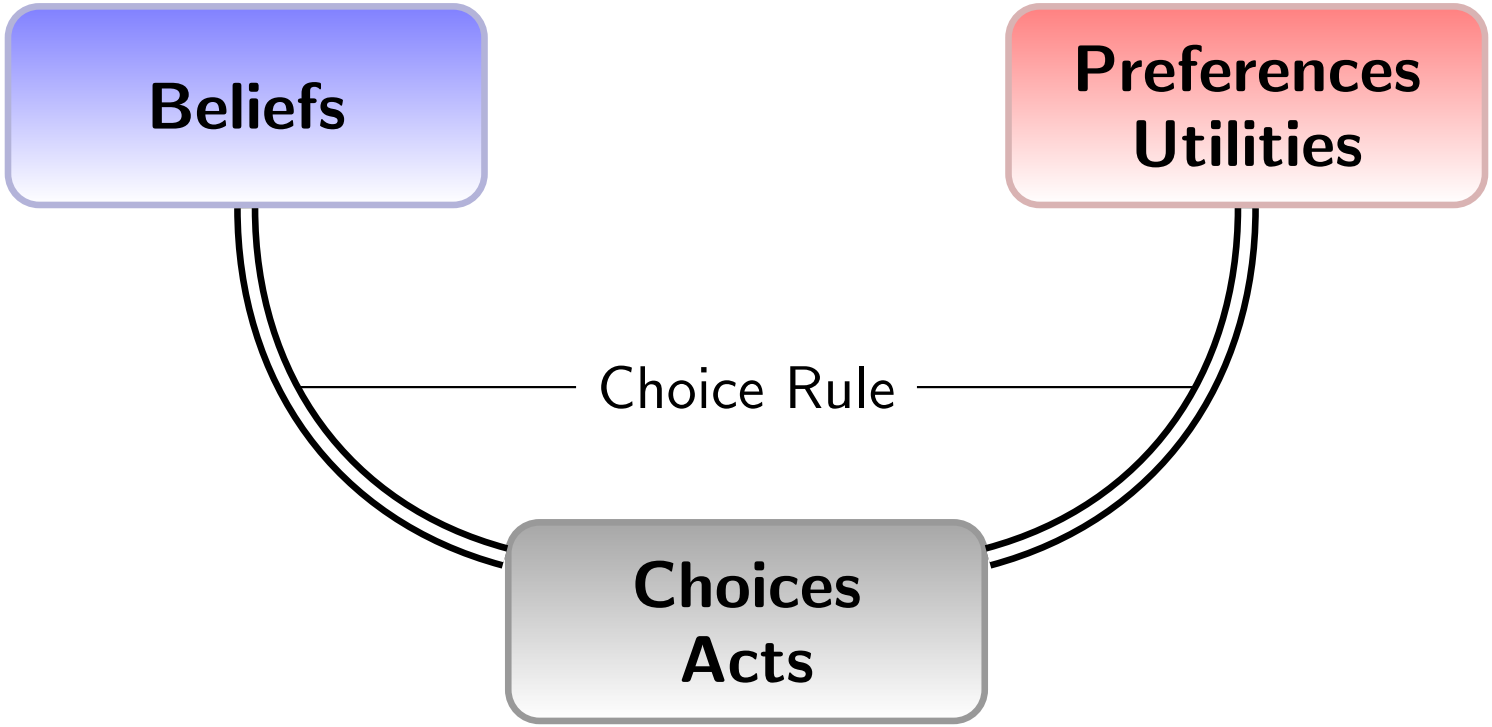
J. Collins. *Desire, Belief and Expectation*. Mind, 100, pgs. 333 - 342, 1997.



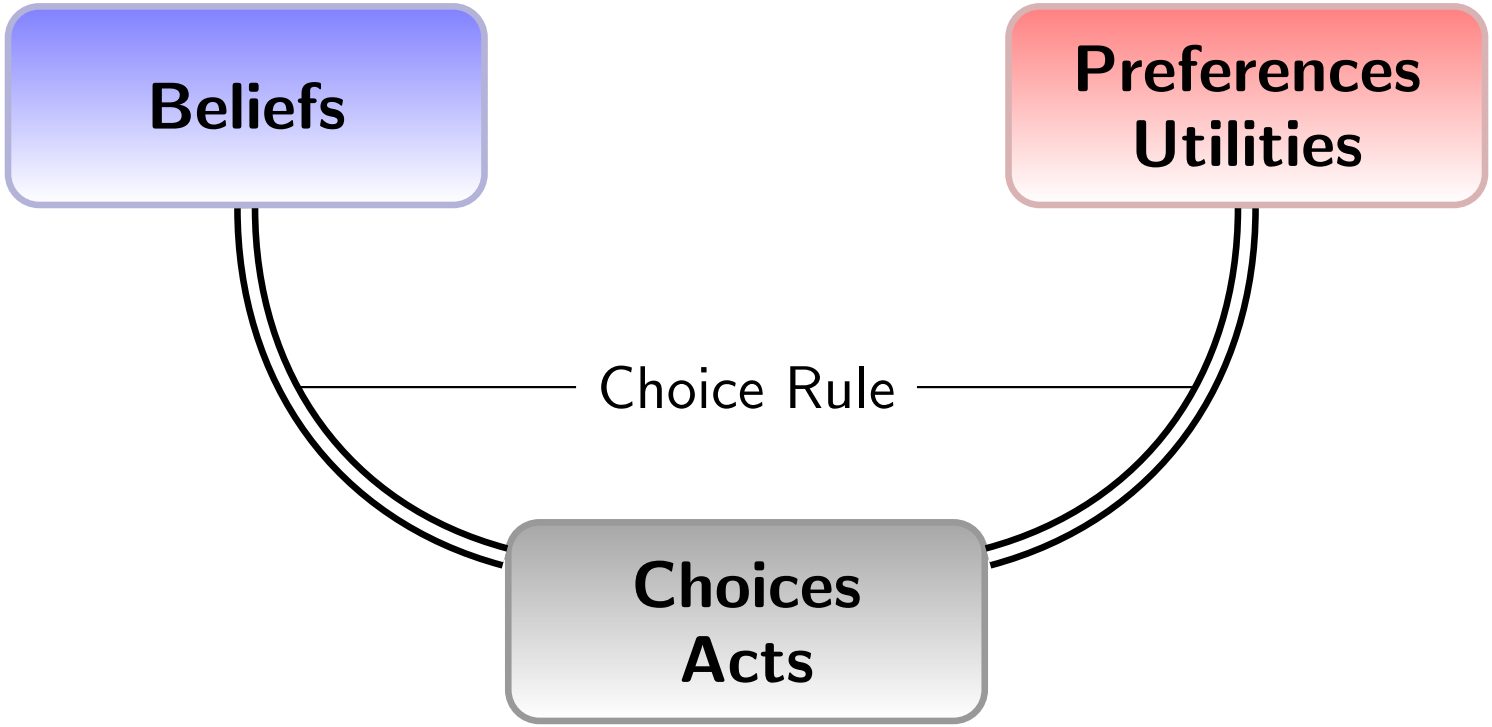
- Is this picture "correct"?
- Basic building blocks: states, outcomes and acts.



- Is this picture "correct"?
- Basic building blocks: states, outcomes and acts.
- How do we *measure* beliefs and utilities?



- Is this picture "correct"?
- Basic building blocks: states, outcomes and acts.
- How do we *measure* beliefs and utilities?
- What are the (derived) normative constraints?



- Is this picture "correct"?
- Basic building blocks: states, outcomes and acts.
- How do we *measure* beliefs and utilities?
- What are the (derived) normative constraints?
- How should we understand normative constraints?

Normative Concepts

Normative repertoire:

Normative Concepts

Normative repertoire: ought, should, must, duty, obligation, right, wrong;

Normative Concepts

Normative repertoire: ought, should, must, duty, obligation, right, wrong; claims about what is justified, warranted, merited, reasonable, permissible;

Normative Concepts

Normative repertoire: ought, should, must, duty, obligation, right, wrong; claims about what is justified, warranted, merited, reasonable, permissible; evaluative concepts: good, bad, better, and worse.

We will focus on two exemplarily normative concepts: *ought* and *a reason*.

Normative Concepts

Normative repertoire: ought, should, must, duty, obligation, right, wrong; claims about what is justified, warranted, merited, reasonable, permissible; evaluative concepts: good, bad, better, and worse.

We will focus on two exemplarily normative concepts: *ought* and *a reason*.

Normative claim: "You ought to go help her."
Descriptive claim: "You are helping her."

Ought

Oughts specify *requirements on agents*.

Ought

Oughts specify *requirements on agents*.

Non-normative uses of 'ought':

- It ought to rain today
- He ought to be on time

Reasons and Oughts

Reasons can be *pro tanto*: "as far as it goes" or *conclusive*

Oughts entail reason but not the other way around.

That you ought to perform a particular action entails that you have reason to do so; nonetheless there may also be reasons not to perform the action, as well as reasons to perform other other actions you ought not to perform.

A reason can be only one reason amongst may others

Rationality Constraints

We use the term 'rational' to signal something about the connections between an agent's various attitudes (both normative and non-normative): eg. *belief states*

Rationality Constraints

We use the term 'rational' to signal something about the connections between an agent's various attitudes (both normative and non-normative): eg. *belief states*

We asses as rational and irrational an agent's *occurrent* mental states

"it is irrational to believe both that p and that $\neg p$."

Rationality also involves the capacity that takes an agent from one mental state to another (either explicitly or implicitly through reasoning):

"Move from believing $\{p, q\}$ to believing $\{p, q, p \wedge q\}$"

Philosophy of Normativity

Philosophy of Normativity

How should we understand what it means to be **rationally required** to be in a relevant attitudinal state of mind?

Philosophy of Normativity

How should we understand what it means to be **rationally required** to be in a relevant attitudinal state of mind?

- How do we make sense of the fact that deliberative reflection can directly give rise to action?
- Which norms for the assessment of action are binding on us as agents? What about *moral norms*?
- Which normative attitude is "primary"? (ought, reason)
- Should you, or even do you have any reason to, comply with *rational requirements* — requirements to have attitudes which fit together in a coherent way?

Philosophy of Normativity

How should we understand what it means to be **rationally required** to be in a relevant attitudinal state of mind?

- How do we make sense of the fact that deliberative reflection can directly give rise to action?
- Which norms for the assessment of action are binding on us as agents? What about *moral norms*?
- Which normative attitude is "primary"? (ought, reason)
- Should you, or even do you have any reason to, comply with *rational requirements* — requirements to have attitudes which fit together in a coherent way?

The Problem of Bootstrapping

If rationality requires to A, then you have conclusive reasons to A.

The Problem of Bootstrapping

If rationality requires to A, then you have conclusive reasons to A.

Suppose that you believe p

The Problem of Bootstrapping

If rationality requires to A, then you have conclusive reasons to A.

Suppose that you believe p

'If you believe that p, then you are rationally required to believe what p entails'

The Problem of Bootstrapping

If rationality requires to A, then you have conclusive reasons to A.

Suppose that you believe p

'If you believe that p, then you are rationally required to believe what p entails'

p entails p

The Problem of Bootstrapping

If rationality requires to A, then you have conclusive reasons to A.

Suppose that you believe p

'If you believe that p, then you are rationally required to believe what p entails'

p entails p

You have conclusive reason to believe that p.

The Detachement Problem (1)

Normativity If rationality requires you to A, you should A

The Detachement Problem (1)

Normativity If rationality requires you to A, you should A

Inconsistency in belief: You believe that p and believe that not-p

Belief Consistency: If you believe that p, rationality requires you not to believe that not-p

The Detachement Problem (1)

Normativity If rationality requires you to A, you should A

Inconsistency in belief: You believe that p and believe that not-p

Belief Consistency: If you believe that p, rationality requires you not to believe that not-p

Means-end incoherence You intend to E, believe that M is necessary for E, but do not intend to M

Means-End If you intend to E and believe that M is necessary for E, rationality requires you to intend to M

The Detachement Problem (2)

Normativity If rationality requires you to A, you should A

Inconsistency in belief: You believe that p and believe that not-p

Belief Consistency: If you believe that p, rationality requires you not to believe that not-p

The Detachement Problem (2)

Normativity If rationality requires you to A, you should A

Inconsistency in belief: You believe that p and believe that not-p

Belief Consistency: If you believe that p, rationality requires you not to believe that not-p

Suppose the evidence for $\neg p$ is conclusive.

The Detachement Problem (2)

Normativity If rationality requires you to A, you should A

Inconsistency in belief: You believe that p and believe that not-p

Belief Consistency: If you believe that p, rationality requires you not to believe that not-p

Suppose the evidence for $\neg p$ is conclusive.

But you believe p anyway.

The Detachement Problem (2)

Normativity If rationality requires you to A, you should A

Inconsistency in belief: You believe that p and believe that not-p

Belief Consistency: If you believe that p, rationality requires you not to believe that not-p

Suppose the evidence for $\neg p$ is conclusive.

But you believe p anyway.

By belief consistency, you are rationally required to not believe that not-p

The Detachement Problem (2)

Normativity If rationality requires you to A, you should A

Inconsistency in belief: You believe that p and believe that not-p

Belief Consistency: If you believe that p, rationality requires you not to believe that not-p

Suppose the evidence for $\neg p$ is conclusive.

But you believe p anyway.

By belief consistency, you are rationally required to not believe that not-p

By normativity, you should not believe that not-p

The Detachement Problem (2)

Normativity If rationality requires you to A, you should A

Inconsistency in belief: You believe that p and believe that not-p

Belief Consistency: If you believe that p, rationality requires you not to believe that not-p

Suppose the evidence for $\neg p$ is conclusive.

But you believe p anyway.

By belief consistency, you are rationally required to not believe that not-p

By normativity, you should not believe that not-p

But, the evidence for not-p is conclusive, that is what you *should* believe.

The Detachement Problem (3)

Normativity If rationality requires you to A, you should A

Means-end incoherence You intend to E, believe that M is necessary for E, but do not intend to M

Means-End If you intend to E and believe that M is necessary for E, rationality requires you to intend to M

The Detachement Problem (3)

Normativity If rationality requires you to A, you should A

Means-end incoherence You intend to E, believe that M is necessary for E, but do not intend to M

Means-End If you intend to E and believe that M is necessary for E, rationality requires you to intend to M

Suppose that you intend to get the job

The Detachement Problem (3)

Normativity If rationality requires you to A, you should A

Means-end incoherence You intend to E, believe that M is necessary for E, but do not intend to M

Means-End If you intend to E and believe that M is necessary for E, rationality requires you to intend to M

Suppose that you intend to get the job

and believe that assassinating your rival is necessary for getting the job.

The Detachement Problem (3)

Normativity If rationality requires you to A, you should A

Means-end incoherence You intend to E, believe that M is necessary for E, but do not intend to M

Means-End If you intend to E and believe that M is necessary for E, rationality requires you to intend to M

Suppose that you intend to get the job

and believe that assassinating your rival is necessary for getting the job.

Means-End and Normativity imply that you *should* intend to assassinate your rival.

Incoherence Patterns

Inconsistency in intentions: If you intend to A, intend to B, and believe that you cannot do both A and B.

Modus ponens incoherence: You believe that p, believe p implies q and do not believe q, although it matters to you whether q.

Akrasia You believe that you should A, but do not intend to A.

Incoherence Patterns

Inconsistency in intentions: If you intend to A, intend to B, and believe that you cannot do both A and B.

Modus ponens incoherence: You believe that p, believe p implies q and do not believe q, although it matters to you whether q.

Akrasia You believe that you should A, but do not intend to A.

Rationality Requirements

Intention consistency: If you intend to A and believe that you cannot do both A and B, rationality requires that you not to intend B.

Modus ponens: You believe that p and believe p implies q, and it matters to you whether q, rationality requires that you believe q.

Enkrasia If you believe that you should A, then rationality requires you to intend to A.

Weak Normativity

If rationality requires you to A, there is a reason for you to A.

Weak Normativity

If rationality requires you to A, there is a reason for you to A.

- You can believe that p, but have no reason at all to believe not-p.
- There is nothing necessarily wrong with failing to do what there is *some* reason to do.

Wide-Scoping

Belief Consistency Rationality requires of you that [if you believe that p, you do not believe that not-p]
Means-end Rationality requires of you that [if you intend to E and believe that M is necessary for E, you intend to M]

Wide-Scoping

Belief Consistency Rationality requires of you that [if you believe that p, you do not believe that not-p]
Means-end Rationality requires of you that [if you intend to E and believe that M is necessary for E, you intend to M]

Thus, you must *avoid* the incoherent patterns (eg., inconsistency in beliefs and means-end incoherence).

Wide-Scoping

Belief Consistency Rationality requires of you that [if you believe that p, you do not believe that not-p]
Means-end Rationality requires of you that [if you intend to E and believe that M is necessary for E, you intend to M]

Thus, you must *avoid* the incoherent patterns (eg., inconsistency in beliefs and means-end incoherence).

But now, there is more than one way to avoid such patterns.

Rational Belief Change

If an agent believes p and $p \rightarrow q$, 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 $p \rightarrow q$, 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.

The Revision Problem

$$State * Input = State$$

The Revision Problem

$$State * Input = State$$

1. How should we *describe* the "mental state"?
2.
3.
4.

The Revision Problem

$$State * Input = State$$

1. How should we *describe* the "mental state"?
2. What is a *coherent* description?
3.
4.

The Revision Problem

$$State * Input = State$$

1. How should we *describe* the "mental state"?
2. What is a *coherent* description?
3. What are the inputs?
4.

The Revision Problem

$$State * Input = State$$

1. How should we *describe* the "mental state"?
2. What is a *coherent* description?
3. What are the inputs?
4. How should we characterize the revision operator?

The Revision Problem

$$\mathcal{B} * \varphi = \mathcal{B}'$$

1. How should we *describe* the "mental state"?
   Propositional formulas
2. What is a *coherent* description?
   A belief state is a consistent and deductively closed set of propositional formulas
3. What are the inputs?
   Consistent propositional formula
4. How should we characterize the revision operator?
   AGM postulates

The Revision Problem

$$State * Input = State$$

1. How should we *describe* the "mental state"?
   Propositional formulas
2. What is a *coherent* description?
   A belief state is a consistent and deductively closed set of propositional formulas
3. What are the inputs?
   Consistent propositional formula
4. How should we characterize the revision operator?
   AGM postulates

AGM Postulates

C. E. Alchourròn, P. Gärdenfors, and D. Makinson,. *On the logic of theory change: Partial meet contraction and revision functions*. Journal of Symbolic Logic, vol. 50, pp. 510 - 530, 1985..

AGM Postulates

AGM 1: $\mathcal{B} * \varphi$ is deductively closed

AGM 2: $\varphi \in \mathcal{B} * \varphi$

AGM 3: $\mathcal{B} * \varphi \subseteq Cn(\mathcal{B} \cup \{\varphi\})$

AGM 4: If $\neg\varphi \notin \mathcal{B}$ then $\mathcal{B} * \varphi = Cn(\mathcal{B} \cup \{\varphi\})$

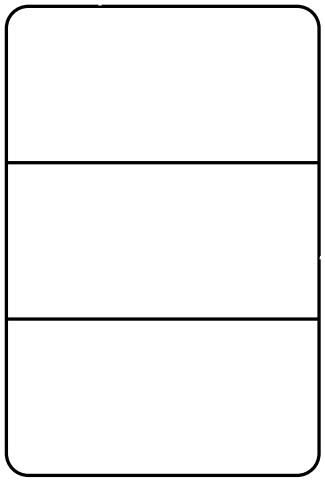
AGM 5: $\mathcal{B} * \varphi$ is inconsistent only if φ is inconsistent

AGM 6: If φ and ψ are logically equivalent then $\mathcal{B} * \varphi = \mathcal{B} * \psi$

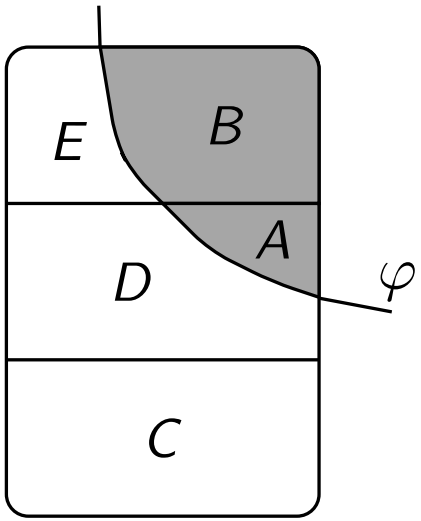
AGM 7: $\mathcal{B} * (\varphi \wedge \psi) \subseteq Cn(\mathcal{B} * \varphi \cup \{\psi\})$

AGM 8 if $\neg\psi \notin \mathcal{B} * \varphi$ then $Cn(\mathcal{B} * \varphi \cup \{\psi\}) \subseteq \mathcal{B} * (\varphi \wedge \psi)$

Revising (All-out) Beliefs

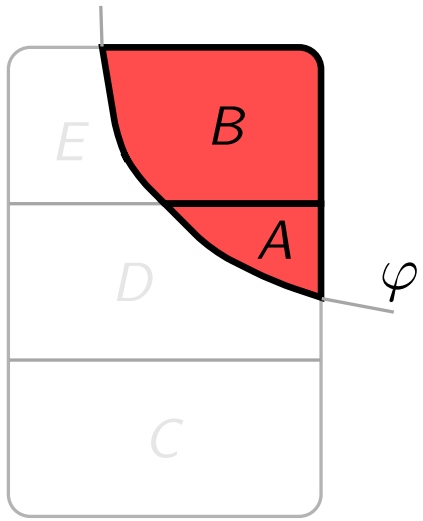


Revising (All-out) Beliefs



Incorporate the new information φ

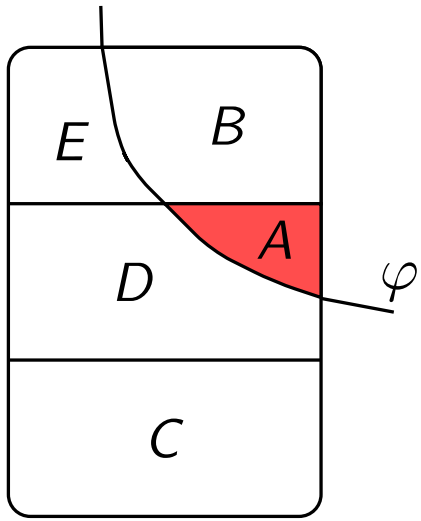
Revising (All-out) Beliefs



Public Announcement: Information from an infallible source

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

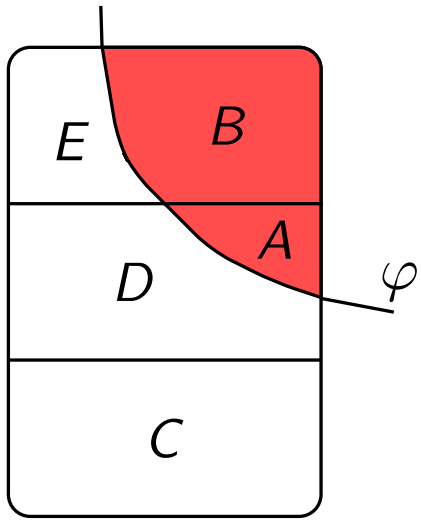
Revising (All-out) Beliefs



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

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

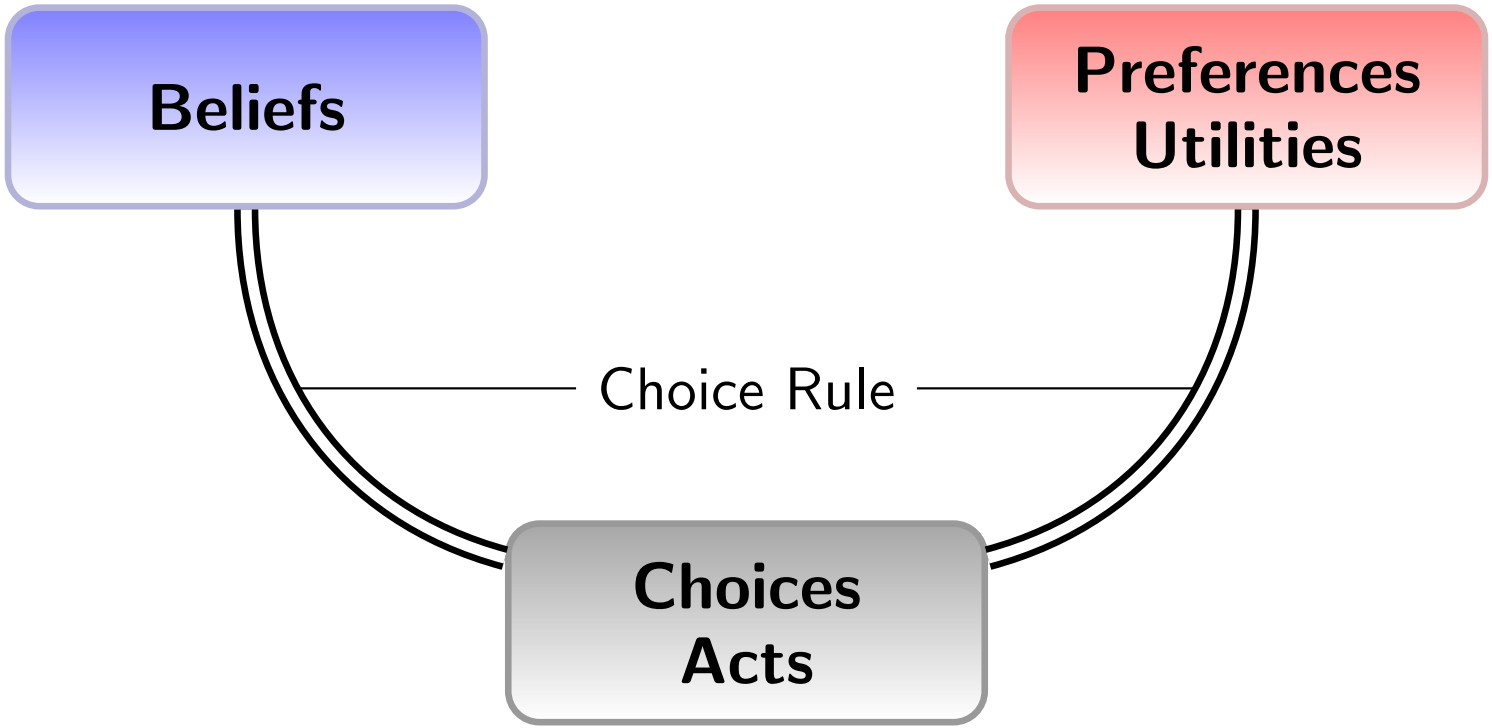
Revising (All-out) Beliefs



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 $(\Uparrow\varphi)$: $A \prec_i B \prec_i C \prec_i D \prec_i E$



- Is this picture "correct"?
- Basic building blocks: states, outcomes and acts.
- How do we *measure* beliefs and utilities?
- What are the (derived) normative constraints?
- How should we understand normative constraints?

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.