The Qualitative Grounds of Quantitative Confidence

Edward Elliott
Slides: tinyurl.com/ndelliott

ND Internal Colloquium, 21st November 2025

Introduction

Introduction

	E_1	E_2	E_3
f	α	β	γ
g	γ	β	α

Expected Utility Theory

Symbols. $f \succsim g \;\Leftrightarrow\; \sum p(E)u\big(f(E)\big) \geq \sum p(E)u\big(g(E)\big)$

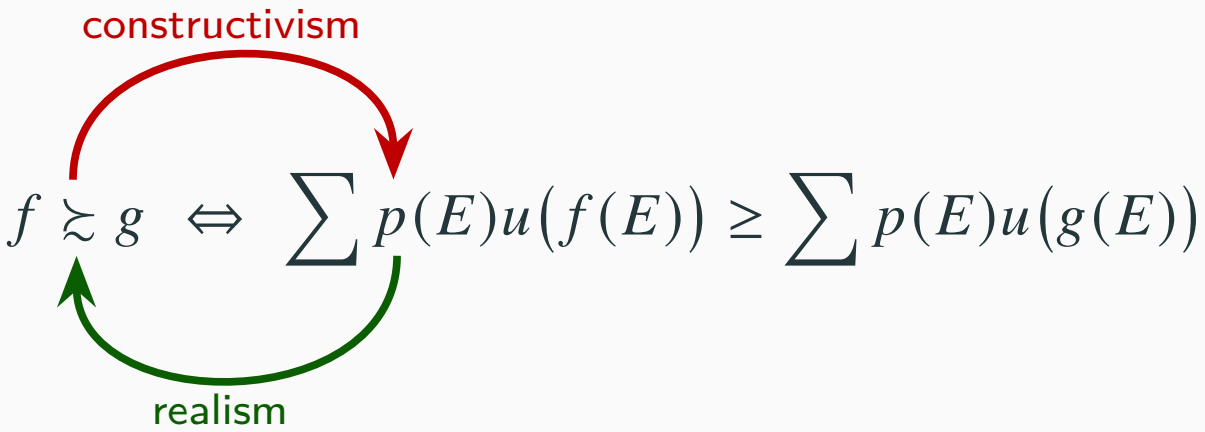
Words. Act f is **preferred** to act g iff the **probability**-weighted average **utility** of f's consequences is greater than that of g's consequences

Introduction

Q1. What is the nature of the relationship between preferences and subjective probabilities?

✗ Constructivism
Preferences are metaphysically prior; probabilities are 'mere constructions' for representing patterns in a coherent system of preferences

✓ Realism
Probability functions represent a 'real' psychological quantity (confidence), which is part of what causally determines preferences



Introduction

Q2. Given realism, how much does the *particular* choice of numbers matter when representing confidence?

✗ Naïve Realism
A numerical representation is correct iff it's *uniquely* correct — e.g., it reflects how the agent explicitly represents their own degrees of confidence when reasoning

✓ Sophisticated Realism
Confidence is a psychological quantity that can be equally-well represented in many ways — e.g., different scales, additive vs multiplicative representations

Introduction

Different scales, same additive representation...

$$f \succsim g \;\Leftrightarrow\; \sum p(E)u\big(f(E)\big) \geq \sum p(E)u\big(g(E)\big)$$

$$\Downarrow \qquad \text{if } c(E) = 10 \cdot p(E)$$

$$f \succsim g \;\Leftrightarrow\; \sum c(E)u\big(f(E)\big) \geq \sum c(E)u\big(g(E)\big)$$

- $p(\top) = 1$
- $p(E) \geq 0$
- $E \cap F = \varnothing$ implies $p(E \cup F) = p(E) + p(F)$

$\Longrightarrow$

- $c(\top) = 10$
- $c(E) \geq 0$
- $E \cap F = \varnothing$ implies $c(E \cup F) = c(E) + c(F)$

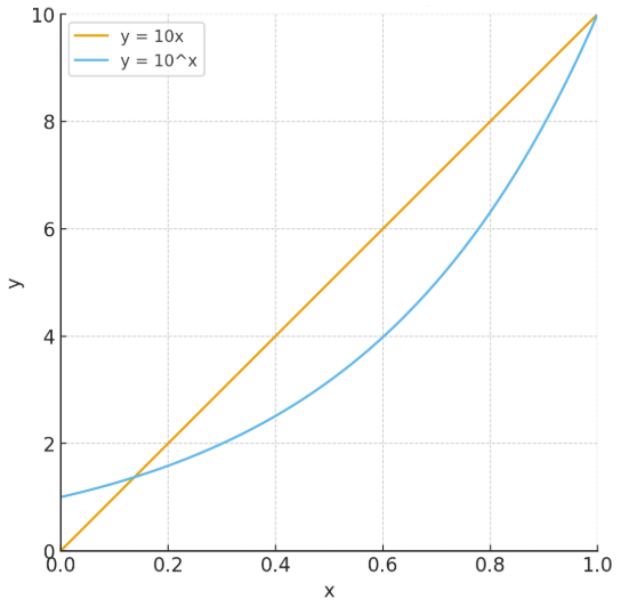
Introduction

Additive vs multiplicative representations...

$$f \succsim g \;\Leftrightarrow\; \sum p(E)u\big(f(E)\big) \geq \sum p(E)u\big(g(E)\big)$$

$$\Downarrow \qquad \text{if } c(E) = 10^{p(E)}$$

$$f \succsim g \;\Leftrightarrow\; \prod c(E)^{u(f(E))} \geq \prod c(E)^{u(g(E))}$$



Q3. Given sophisticated realism, what information in the *standard additive representation* is *psychologically meaningful*?

✗ **Ordinal**	✗ **Interval**	✓ **Ratio**	✗ **Absolute**
Moh hardness, hotel stars, Fujita scale	temperature, calendar year, IQ scores	length, mass, duration, area, speed, density	counting, percentages, relative humidity

- **Ordinal information:** "I have *more* confidence in E than in F"
- **Interval information:** "I have *much more* confidence in E than in F"
- **Ratio information:** "I have *twice as much* confidence in E than in F"

Q4. *Why* is it meaningful?

✗ Intrinsic-only Explanation
Appeals only to the intrinsic properties of the doxastic system — e.g., structure of the agent's relative confidence ordering

✓ Interactive Explanation
Appeals *at least in part* to how confidence and preference relate in decision making under uncertainty

Introduction

I will argue for two main theses...

Meaningful Ratios. In the context of expected utility theory (as standardly represented), confidence ratios are *at least sometimes* psychologically meaningful

Interactive Explanation. The explanation for this fact derives *at least in part* from the interactions between confidence and preference in decision making

Introduction

Structure of the talk...

Part 1. Representation theorem

Part 2. Two challenges

Part 3. Realism and meaningful confidence ratios

Part 4. Interactive versus intrinsic explanations

Part 1. Representation Theorem

Representation Theorem

Typical representation theorem

- If an agent's preferences ($\succsim$) satisfy certain axioms, then those preferences are **as if** the agent is maximizing expected utility given a certain probability function
- Also, there's exactly one probability function for which this is true

My representation theorem

- If an agent's preferences ($\succsim$) **and** relative confidences ($\succsim_c$) **jointly** satisfy certain axioms, then those preferences are **as if** the agent is maximizing expected utility given a certain probability function
- Also, there's exactly one probability function for which this is true, **and** it's a measure of relative confidence

Representation Theorem

Typical representation theorem

- If an agent's preferences ($\succsim$) satisfy certain axioms, then those preferences are **as if** the agent is maximizing expected utility given a certain probability function
- Also, there's exactly one probability function for which this is true

My representation theorem

- If an agent's preferences ($\succsim$) **and** relative confidences ($\succsim_c$) **jointly** satisfy certain axioms, then those preferences are **as if** the agent is maximizing expected utility given a certain probability function
- Also, there's exactly one probability function for which this is true, **and** it's a measure of relative confidence

A1. Correlation

Symbols. $\alpha \succsim \beta$ implies $\alpha E \beta \succsim \alpha F \beta \Leftrightarrow E \succsim_c F$

Words. Suppose you prefer apples (α) to bananas (β), and you have to choose between the following:

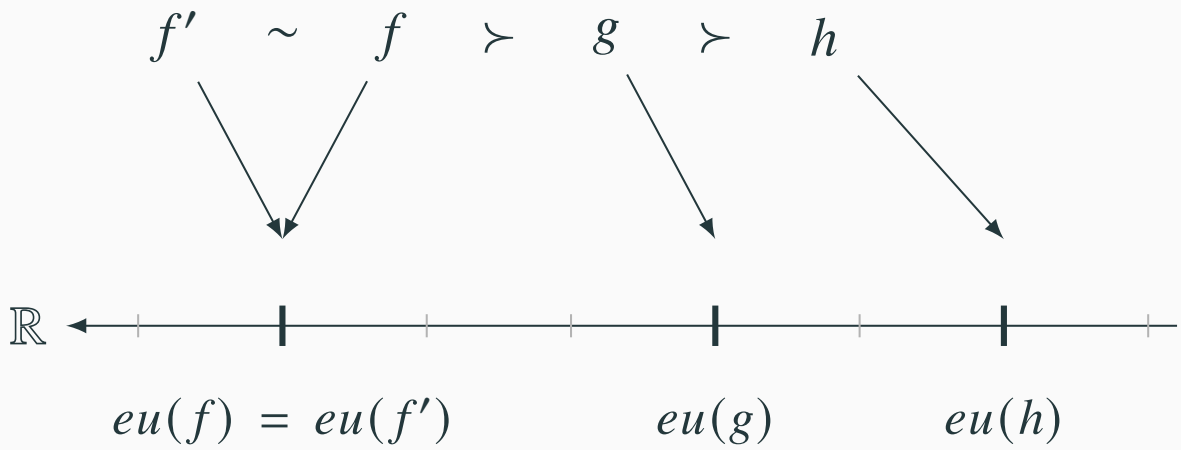
- $\alpha E \beta =$ get an apple if E is true, a banana otherwise
- $\alpha F \beta =$ get an apple if F is true, a banana otherwise

You prefer $\alpha E \beta$ to $\alpha F \beta$ just in case you have more confidence in E than in F

A2. Weak Ordering

Symbols. (i) $\succsim$ is transitive — $f \succsim g$ and $g \succsim h$ implies $f \succsim h$
(ii) $\succsim$ is connected — $f \succsim g$ or $g \succsim f$ for all f, g

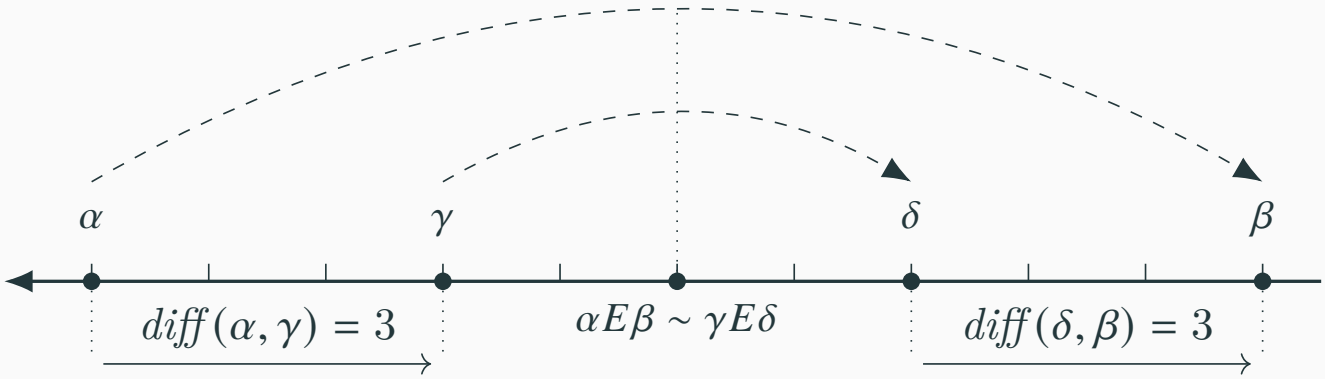
Words. The preference ordering can be coherently represented using points on the real number line



A3. Equal Differences

Symbols. If $E \sim_c \overline{E}$, then $\alpha E \beta \sim \gamma E \delta$ implies $diff(\alpha, \gamma) = diff(\delta, \beta)$

Words. If midway between the utilities of α and β is equal to midway between the utilities of γ and δ, then there's the same number of 'steps' between α and γ as there are between β and δ



A4. Strong Sure-Thing

Symbols. $diff(\alpha, \alpha') = diff(\beta, \beta')$ implies $\alpha E f \succsim \beta E g \Leftrightarrow \alpha' E f \succsim \beta' E g$

Words. If one pair of acts f, g has the same *pattern of utility differences* as another pair of acts f', g', then the preferences between f, g should match those between f', g'

	+2	−1	−3
	s_1	s_2	s_3
f	0	3	4
g	2	2	1

	s_1	s_2	s_3
f'	2	2	3
g'	4	1	0

$$\Rightarrow \qquad f \succsim g \Leftrightarrow f' \succsim g'$$

A5. Dominance

Symbols. $\alpha \succsim \beta \Leftrightarrow \alpha E f \succsim \beta E f$ unless E is null

Words. If act f has better consequences than act g at some event (*with positive probability*), and otherwise they have identical consequences, then f is preferred to g

	E_1	E_2	E_3
f	1	0	0
g	0	0	0

$\Rightarrow \quad f \succ g$ unless $p(E_1) = 0$

A6. Exchange coherence

Symbols. If there are $\alpha, \beta, \gamma, \delta$ such that $\alpha, \beta, \gamma \succ \delta$, $\gamma E_i \delta \sim \alpha F \delta$, and $\gamma E_j \delta \sim \beta F \delta$, then $\beta E_i \delta \sim \alpha E_j \delta$

Words. "Exchange rates" should be jointly coherent

$f \sim g$	E_1	E_2	E_3
f	1	0	0
g	0	0	3

+

$f' \sim g'$	E_1	E_2	E_3
f'	0	1	0
g'	0	0	2

$\Downarrow$

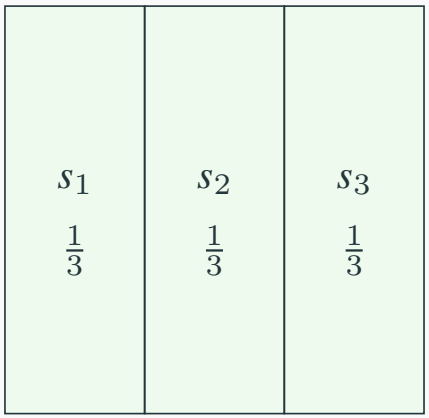
$f'' \sim g''$	E_1	E_2	E_3
f''	2	0	0
g''	0	3	0

A7. Factorization

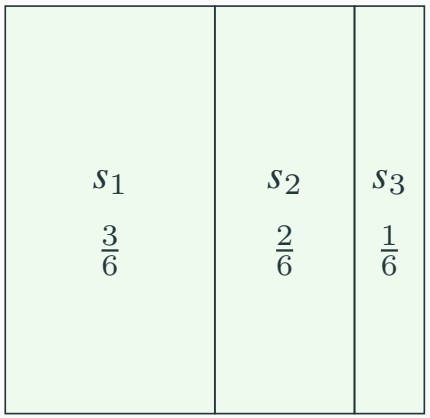
Symbols. There is a nonnull state s such that for all nonnull states s', there are α, β, γ where $diff(\beta, \gamma) \geq diff(\alpha, \gamma) = 1$ and $\alpha\{s'\}\gamma \sim \beta\{s\}\gamma$

Words. There is a state (read: *atomic event*) s with minimal positive probability, such that the probability of every other event is an integer multiple of the probability of s

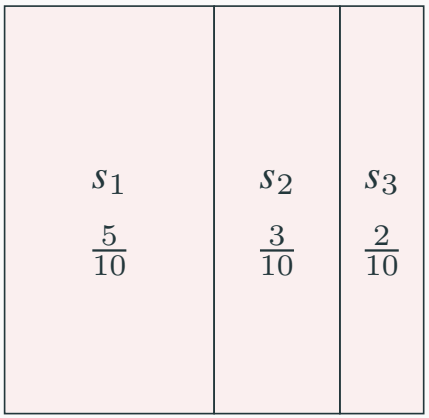
✓



✓



✗



Representation Theorem

Main Theorem

If **A1–A7** are satisfied, then there exists a confidence function c and a utility function u such that

$$f \succsim g \;\Leftrightarrow\; \sum c(E)u\big(f(E)\big) \geq \sum c(E)u\big(g(E)\big)$$

Furthermore,

1. u is unique up to a interval-preserving transformation
2. c is unique up to a ratio-preserving transformation
3. c is additive, non-negative, and $E \succsim_c F \Leftrightarrow c(E) \geq c(F)$

Corollary

If **A1–A7** are satisfied, then $\succsim$ maximizes expected utility relative to *exactly one* probability function

Representation Theorem

Main Theorem

If **A1–A7** are satisfied, then there exists a confidence function c and a utility function u such that

$$f \succsim g \;\Leftrightarrow\; \sum c(E)u\big(f(E)\big) \geq \sum c(E)u\big(g(E)\big)$$

Furthermore,

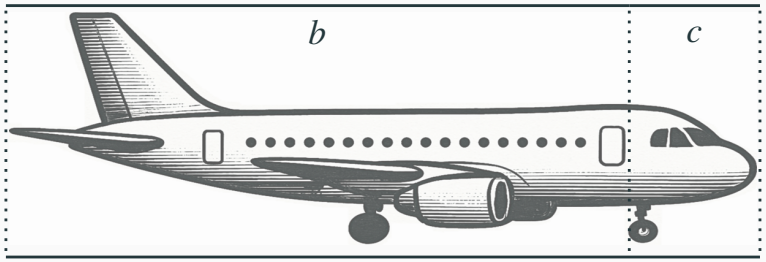
1. u is unique up to a interval-preserving transformation
2. c is unique up to a ratio-preserving transformation
3. c is additive, non-negative, and $E \succsim_c F \Leftrightarrow c(E) \geq c(F)$

Corollary

If **A1–A7** are satisfied, then $\succsim$ maximizes expected utility relative to *exactly one* probability function

Part 2. Two Challenges

Two Challenges



$a = b + c$

Two-part structure to the measurement of length...

Part A. Assign numbers to lengths to reflect the ordering ($a \succ b \succ c$) — this is *what it is* for assignment to be a measure of length

Part B. Restrict to measures that also reflect additive part-whole structure ($a = b + c$) — this is *what makes ratios meaningful*

Two Challenges

Generalizing…

- Every quantity Q is associated with an ordering — e.g., relative length, relative temperature, relative utility, etc.
- A *measure of* Q is any assignment of numbers that reflects the ordering — that's *what it is* to be a measure of Q [Part A]
- Every measure carries ordinal information about Q; it carries *further information* only if it also represents some *further structure* [Part B]

Two Challenges

Generalizing…

- Every quantity Q is associated with an ordering — e.g., relative length, relative temperature, relative utility, etc.
- A *measure of* Q is any assignment of numbers that reflects the ordering — that's *what it is* to be a measure of Q [Part A]
- Every measure carries ordinal information about Q; it carries *further information* only if it also represents some *further structure* [Part B]

Two Challenges

Generalizing...

- Every quantity Q is associated with an ordering — e.g., relative length, relative temperature, relative utility, etc.
- A *measure of* Q is any assignment of numbers that reflects the ordering — that's *what it is* to be a measure of Q [Part A]
- Every measure carries ordinal information about Q; it carries *further information* only if it also represents some *further structure* [Part B]

Two Challenges

Generalizing...

- Every quantity Q is associated with an ordering — e.g., relative length, relative temperature, relative utility, etc.
- A *measure of* Q is any assignment of numbers that reflects the ordering — that's *what it is* to be a measure of Q [Part A]
- Every measure carries ordinal information about Q; it carries *further information* only if it also represents some *further structure* [Part B]

Two Challenges

Meaningful Ratios. In the context of expected utility theory (as standardly represented), confidence ratios are *at least sometimes* psychologically meaningful

→ Need to show the probability function which 'fits' the preference ordering is *also* a measure of confidence on a ratio scale

Interactive Explanation. The explanation for this fact derives *at least in part* from the interactions between confidence and preference in decision making

→ Need to show there are meaningful confidence ratios that *cannot* be given an intrinsic-only explanation

Two Challenges

Meaningful Ratios. In the context of expected utility theory (as standardly represented), confidence ratios are *at least sometimes* psychologically meaningful

→ Need to show the probability function which 'fits' the preference ordering is *also* a measure of confidence on a ratio scale

Interactive Explanation. The explanation for this fact derives *at least in part* from the interactions between confidence and preference in decision making

→ Need to show there are meaningful confidence ratios that *cannot* be given an intrinsic-only explanation

Part 3. Meaningful Ratios

Meaningful Ratios

Typical Representation Theorem

If an agent's preferences satisfy certain axioms, then those preferences are **as if** the agent is maximizing expected utility given a certain probability function

Problem: Unless we assume constructivism, nothing implies that the probability function actually reflects the agent's confidence ordering $\succsim_c$!

Meaningful Ratios

The solution...

1. Posit *relative confidence* ($\succsim_c$) as an independent primitive relation alongside *preference* ($\succsim$)
2. Specify conditions under which relative confidence *correlates* with certain preference patterns

Meaningful Ratios

Implementing the solution...

Suppose you prefer apples (α) to bananas (β), and you have to choose between the following:

- $\alpha E \beta =$ get an apple if E is true, a banana otherwise
- $\alpha F \beta =$ get an apple if F is true, a banana otherwise

If $\succsim$ satisfies **A2–A7**, then there's a probability p such that...

$$p(E) \geq p(F) \Leftrightarrow \alpha E \beta \succsim \alpha F \beta$$

If $\succsim$ and $\succsim_c$ jointly satisfy **A1**, then...

$$\alpha E \beta \succsim \alpha F \beta \Leftrightarrow E \succsim_c F$$

Representation Theorem

Main Theorem

If **A1–A7** are satisfied, then there exists a confidence function c and a utility function u such that

$$f \succsim g \;\Leftrightarrow\; \sum c(E)u\big(f(E)\big) \geq \sum c(E)u\big(g(E)\big)$$

Furthermore,

1. u is unique up to a interval-preserving transformation
2. c is unique up to a ratio-preserving transformation (**A2–A7**)
3. c is additive, non-negative, and $E \succsim_c F \Leftrightarrow c(E) \geq c(F)$ (**A1**)

Upshot

c is a measure of confidence (**A1**), and c-ratios are meaningful (**A2–A7**)

Comparison to literature...

- A few 'dual-primitive' representation theorems exist, with axioms on both preference and relative confidence
- These include axioms that *directly* impose additive structure on relative confidence — such as the *qualitative additivity* axiom:

$$E \succsim_c E' \Leftrightarrow E \cup F \succsim_c E' \cup F \text{ whenever } E \cap F = E' \cap F = \varnothing$$

- For my theorem, *almost* all structure in $\succsim_c$ derives (via **A1**) from axioms on preference — so no ratio information is smuggled in through direct constraints on $\succsim_c$

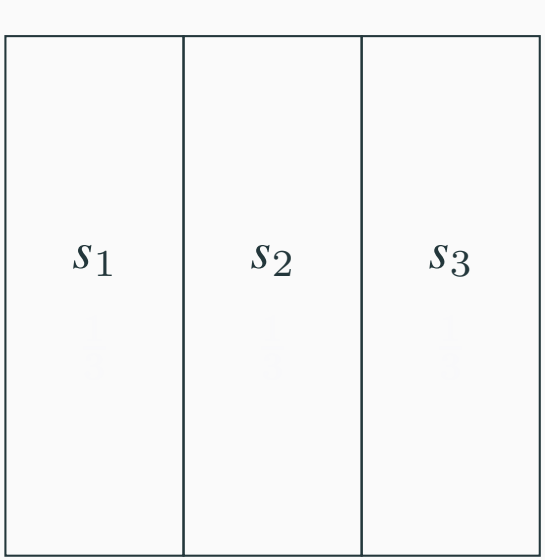
Part 4: Intrinsic vs Interactive Explanations

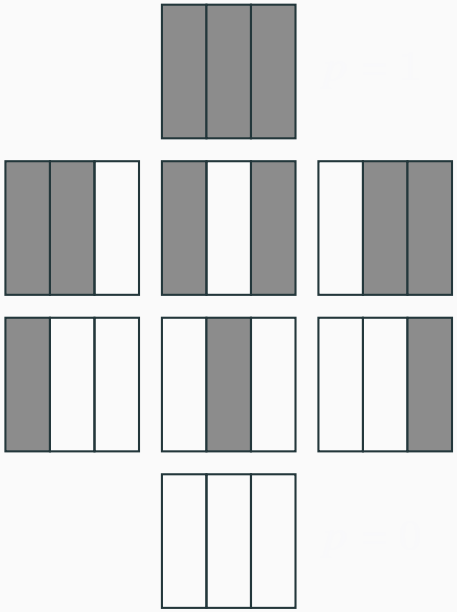
Intrinsic vs Interactive Explanations

Interactive Explanation. The explanation for the meaningfulness of confidence ratios derives *at least in part* from the interactions between confidence and preference in decision making

→ Need to show there are meaningful confidence ratios that *cannot* be given an intrinsic-only explanation — e.g., by appealing only to the structure of the relative confidence relation

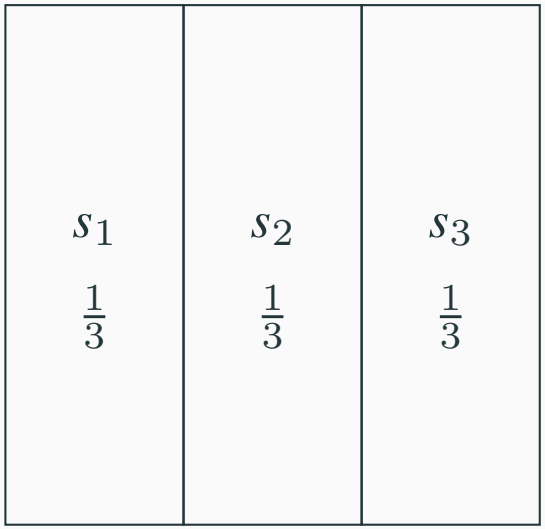
Intrinsic vs Interactive Explanations

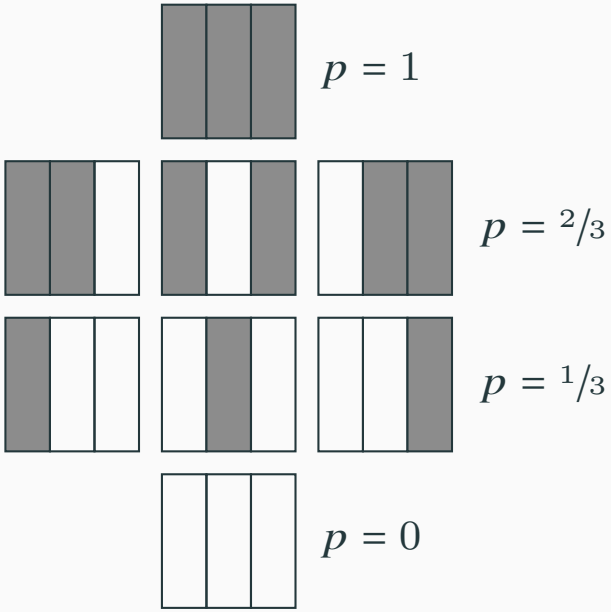




$$\begin{array}{ccccccc} & & \{s_1, s_2\} & & s_1 & & \\ \top & \succ_c & \{s_1, s_3\} & \succ_c & s_2 & \succ_c & \emptyset \\ & & \{s_2, s_3\} & & s_3 & & \end{array}$$

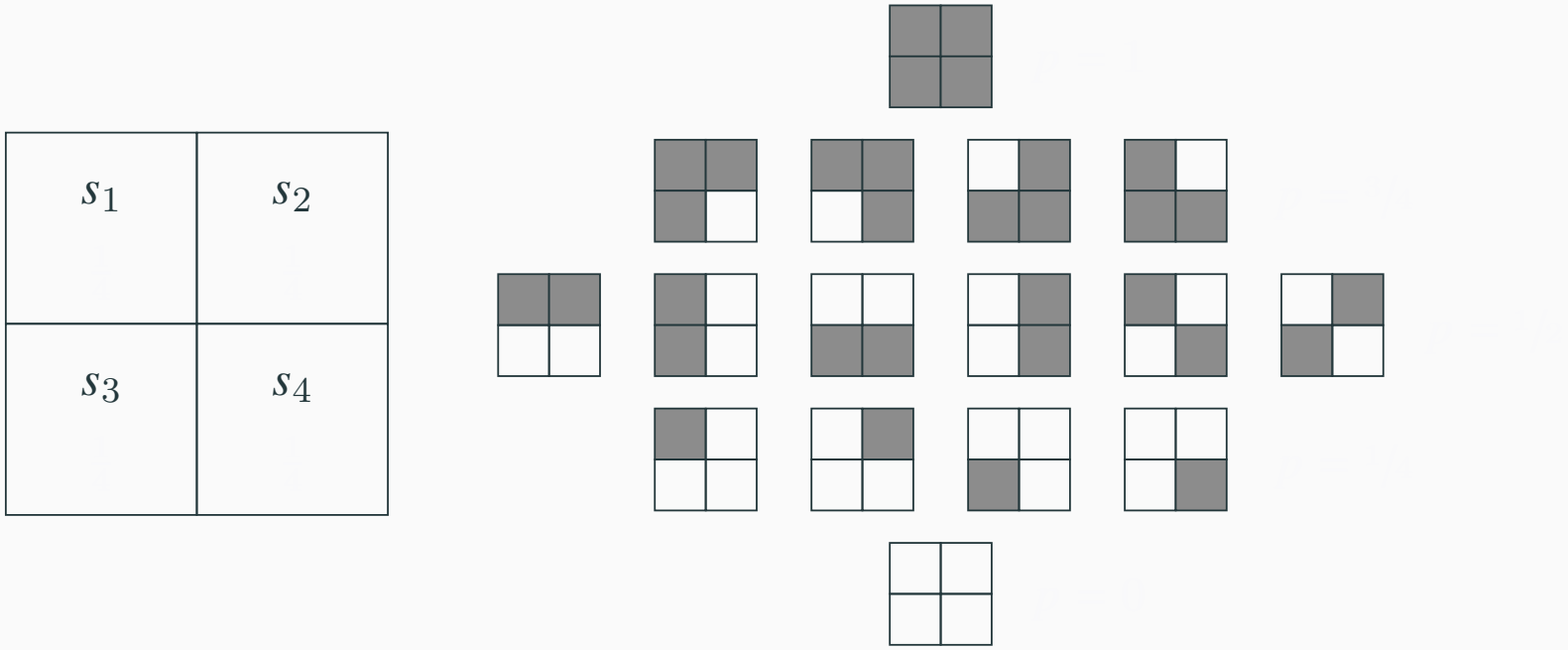
Intrinsic vs Interactive Explanations





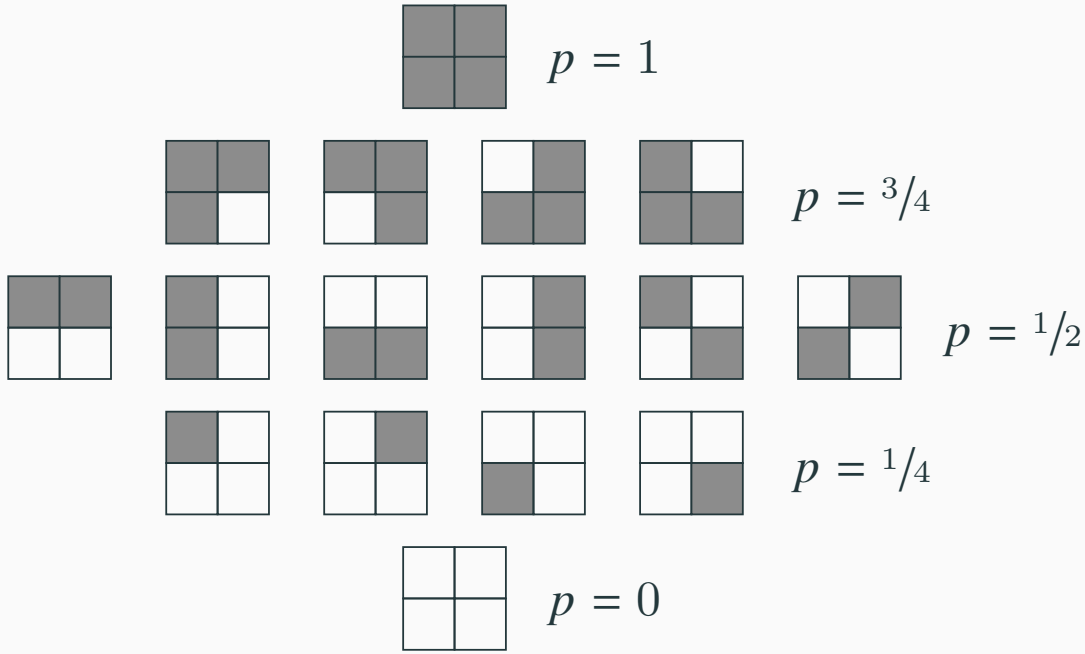
$$\top \succ_c \begin{matrix} \{s_1, s_2\} \\ \{s_1, s_3\} \\ \{s_2, s_3\} \end{matrix} \succ_c \begin{matrix} s_1 \\ s_2 \\ s_3 \end{matrix} \succ_c \varnothing$$

Intrinsic vs Interactive Explanations

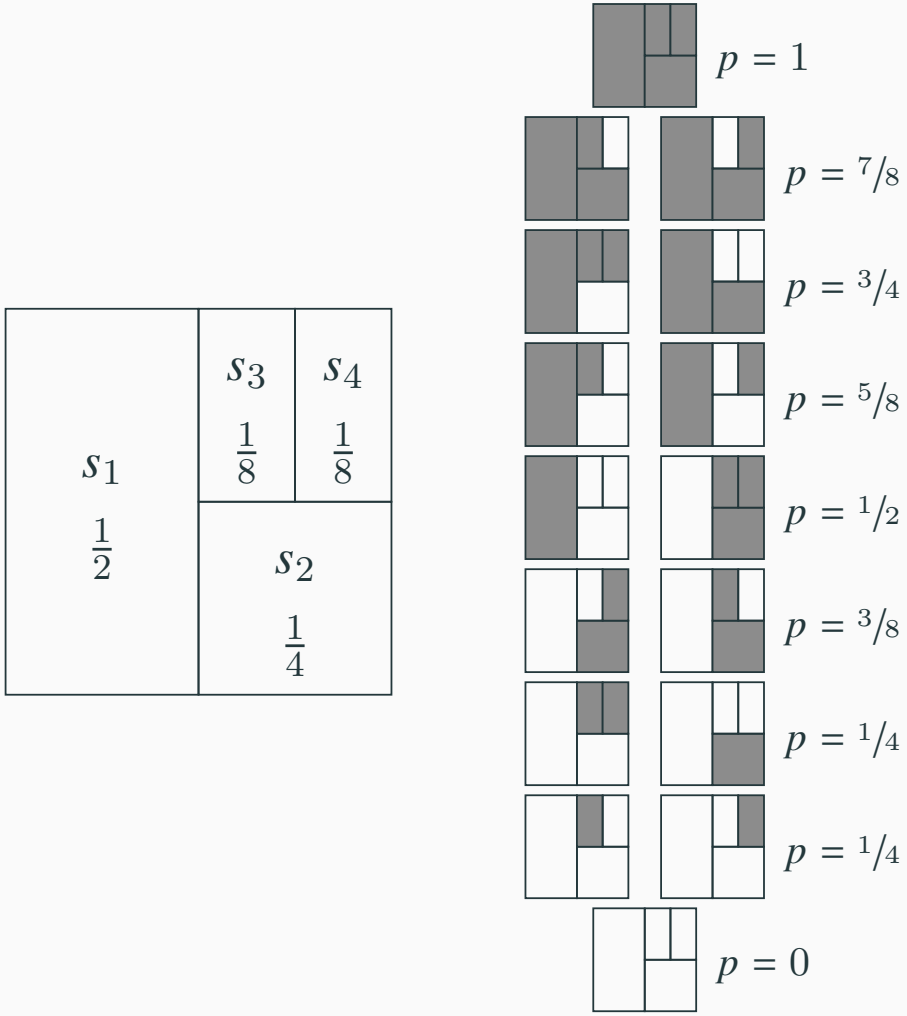


Intrinsic vs Interactive Explanations

s_1 $\frac{1}{4}$	s_2 $\frac{1}{4}$
s_3 $\frac{1}{4}$	s_4 $\frac{1}{4}$



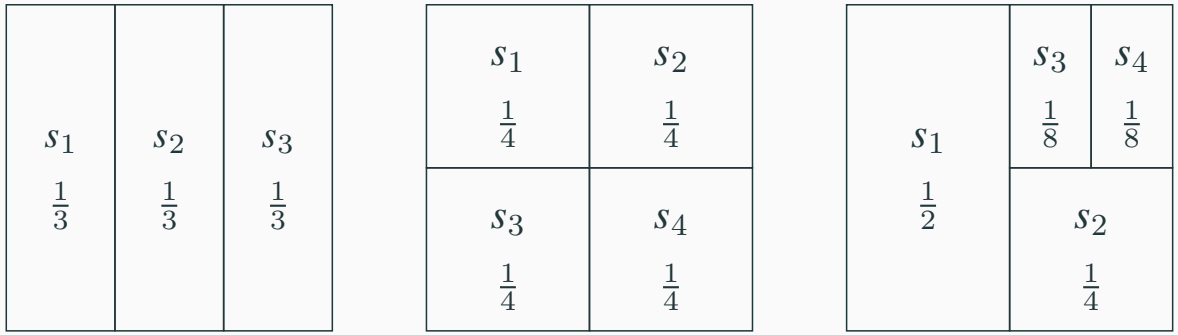
Intrinsic vs Interactive Explanations



Meaningful Ratios

Ratios and relative confidence...

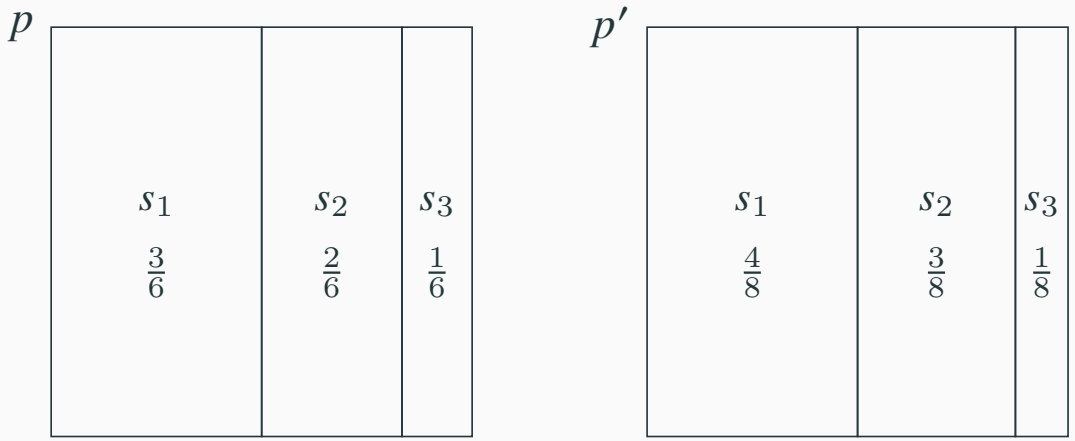
- Relative confidence *alone* determines meaningful ratios iff the probabilistic representation is **ordinally determined**
- p is **ordinally determined** iff no other probability function p' corresponds to the same confidence ordering



Meaningful Ratios

Ratios and relative confidence...

- Consider p and p' below, which are **ordinally equivalent** but represent different ratio information
- If the ratios in p and/or p' are meaningful, then this fact *cannot* be explained by relative confidence alone



Intrinsic vs Interactive Explanations

On the handout...

1. One comparative confidence ordering $\succsim_c$
2. Two distinct preference orderings, $\succsim$ and $\succsim'$
3. Both $\succsim$ and $\succsim'$ satisfy all axioms **A1–A7** with $\succsim_c$

From the theorem...

1. $\succsim$ maximizes expected utility relative to p (and not p')
2. $\succsim'$ maximizes expected utility relative to p' (and not p)
3. Ratios in both p and p' are meaningful, but they are ordinally equivalent

Upshot: The explanation for meaningful ratios in p and p' *cannot* appeal only to the structure of the confidence ordering

Intrinsic vs Interactive Explanations

But wait, there's more!

The same result applies to both of the other 'intrinsic-only' explanations on the market, which add further doxastic structure...

- **Enrich** $\succsim_c$ with quarternary *conditional confidence* comparisons — $E|F \succ_c E'|F'$ iff more confidence in E *given* F than E' *given* F'
- **Supplement** $\succsim_c$ with a primitive *independence* relation — $E \perp F$ iff E is evidentially independent of F

Intrinsic vs Interactive Explanations

Comparison to literature...

Savage's theorem

- Order of construction...
  1. Derives relative confidence from preference
  2. Shows confidence has unique probabilistic representation
  3. Derives interval-scale utilities from unique probabilities
- Savage's probabilities are always ordinally determined — cannot differentiate intrinsic versus interactive explanations

Ramsey's theorem

- Order of construction...
  1. Derives interval-scale utilities from preference
  2. Derives unique probabilities from interval-scale utilities
- Ramsey's probabilities needn't be ordinally determined — but the construction requires highly simplified decision-theoretic framework

Comparison to literature...

Savage's theorem

- Order of construction...
    1. Derives relative confidence from preference
    2. Shows confidence has unique probabilistic representation
    3. Derives interval-scale utilities from unique probabilities
- Savage's probabilities are always ordinally determined — cannot differentiate intrinsic versus interactive explanations

Ramsey's theorem

- Order of construction...
    1. Derives interval-scale utilities from preference
    2. Derives unique probabilities from interval-scale utilities
- Ramsey's probabilities needn't be ordinally determined — but the construction requires highly simplified decision-theoretic framework

Conclusion

Meaningful Ratios. In the context of expected utility theory (as standardly represented), confidence ratios are *at least sometimes* psychologically meaningful

Interactive Explanation. The explanation for this fact derives *at least in part* from the interactions between confidence and preference in decision making