higher-order desire and preference change

Katie Steele

LSE

Outline

1. The starting question: preference change
2. Normal-form-Extensive-form equivalence
3. Ulysses and higher-order desire
4. Jeffrey on higher-order desire
5. Planned non-Bayesian changes of desire?

1. The starting question

- Can we model preference change that is not prompted by new information (as in, not *Bayesian conditionalisation*)?

- Such preference change could result from a change in one's prior beliefs or a change in one's 'basic' desires. Most people tend to concentrate on the latter. (Why?)

- What exactly do we want to achieve with the model?

  - Do we want to simply include preference change in, e.g., decision models?
  - Do we want to make the actual change explicable?
  - Do we want to make sense of reasoned or planned change?

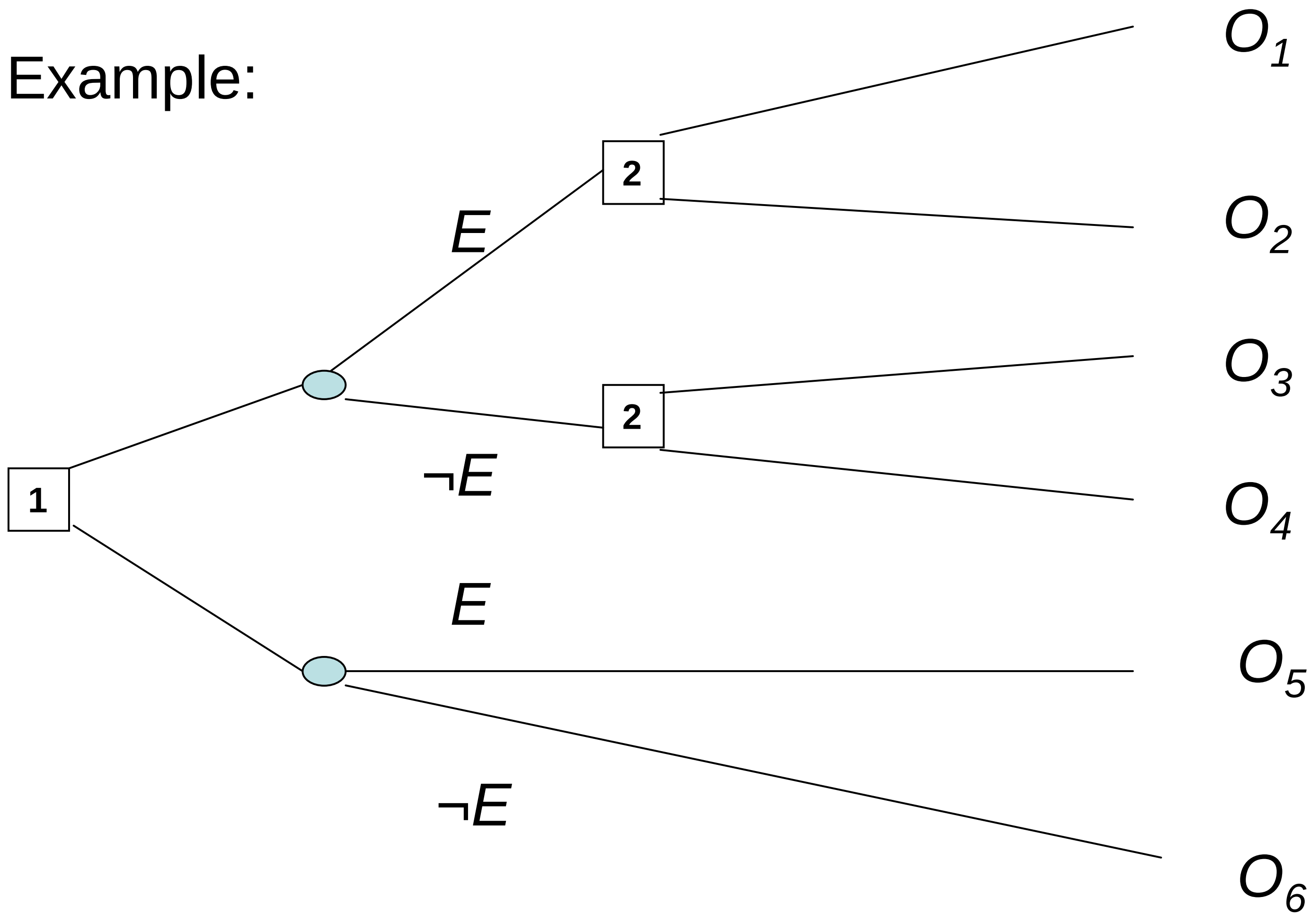
- My starting intuition was that *planned* changes of desire happen all the time, and are not irrational.

  - E.g. Wanting to desire opera more than one does now.

- But if we are going to talk about *plans* then we should employ the Bayesian decision model, and the sequential or extensive-form version is going to be most illuminating

2. Normal-form-Extensive-form equivalence

Example:



Normal-form-Extensive-form equivalence

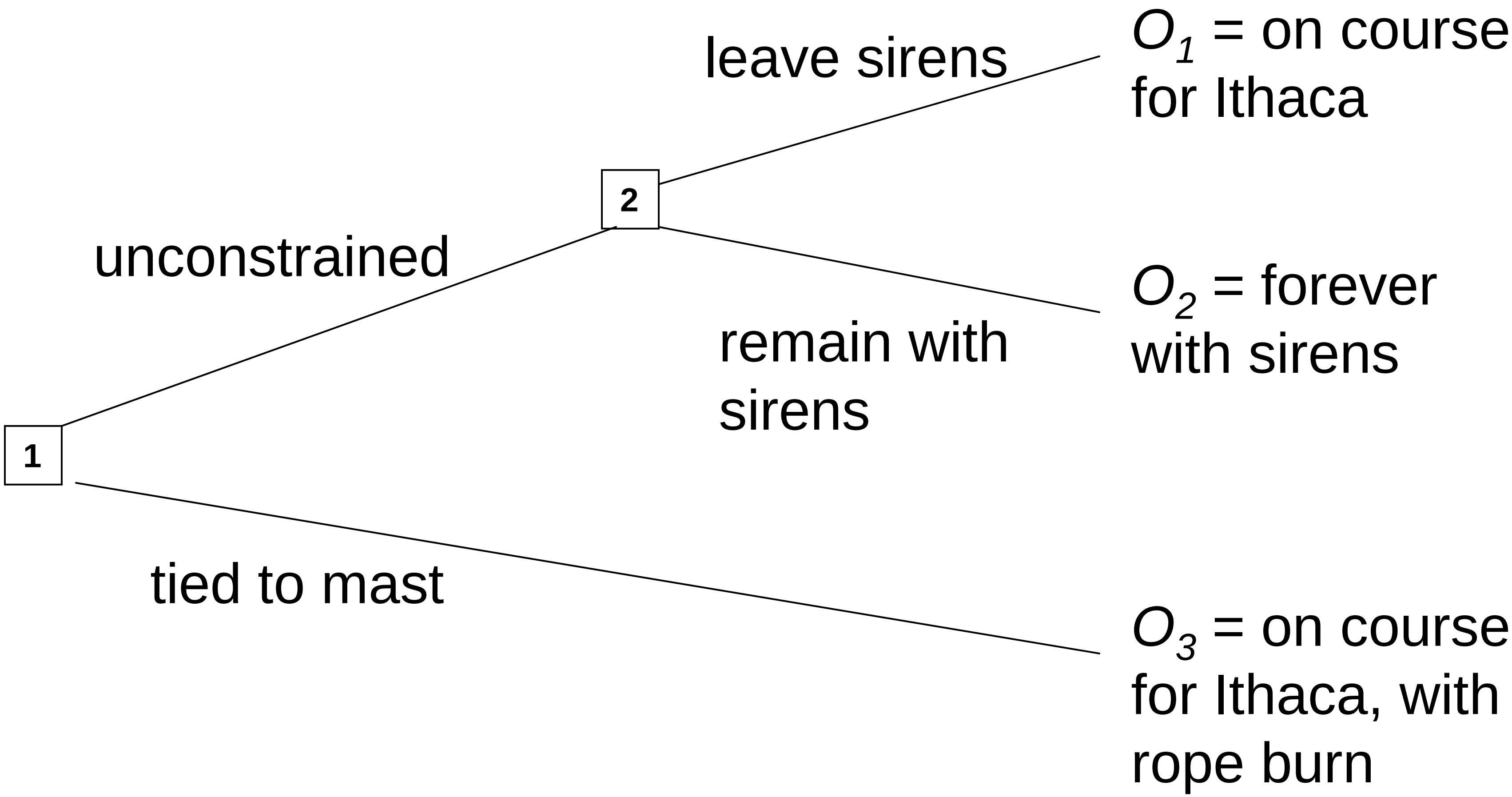
Example:

	E	***¬E***
D	O_5	O_6
(U,(U,U))	O_1	O_3
(U,(U,D))	O_1	O_4
(U,(D,U))	O_2	O_3
(U,(D,D))	O_2	O_4

In general, we have normal-form-extensive-form equivalence if:

- Agent is Bayesian (maximises EU, any preference change via conditionalisation).

Ulysses Again



Preferences at 1 ***(P1)****:* $O_1 > O_3 > O_2$

Preferences at 2 ***(P2)****:* $O_2 > O_1 > O_3$

Ulysses Again

	¬*E*
(unconstrained, leave sirens)	O_1
(unconstrained, remain with sirens)	O_2
(tied to mast)	O_3

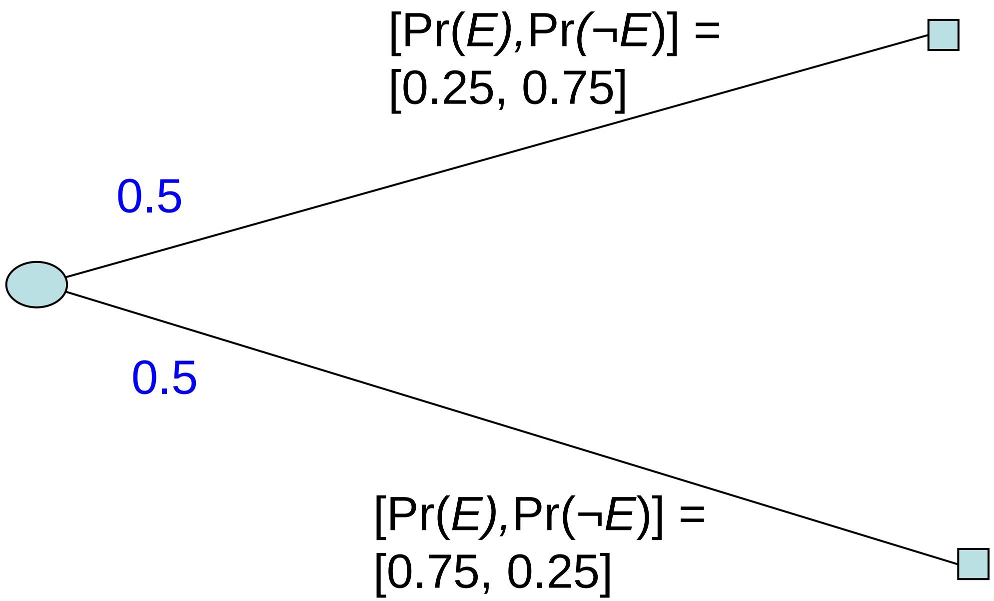
We can model preference change in the sequential-choice model, sure enough…

…but such changes are unfortunate because they do not necessarily preserve normal-form-extensive-form equivalence.

So from sequential-choice perspective, non-Bayesian changes look bad (irrational) insofar as they are *planned*.

A small aside:

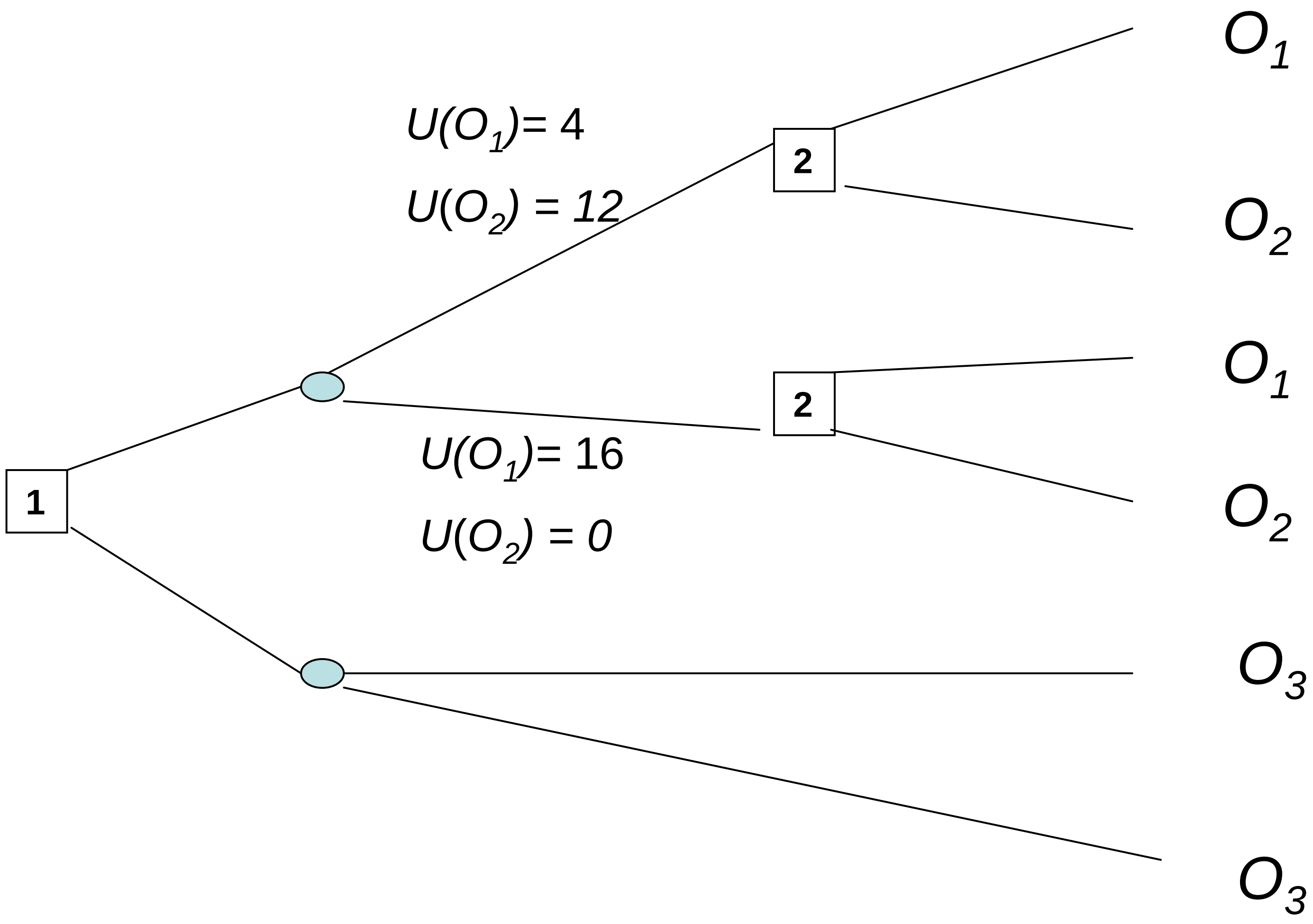
Can we plan to update according to Jeffrey-conditionalisation?



By reflection: Pr(*E*) = 0.5.

If update Jeffrey-style do we ensure normal-form-ext-form equivalence in decision problems like the above?

A modified Ulysses:



Agent satisfies reflection with respect to their utilities, but normal-form-extensive-form equivalence is not respected.

- So we have violations of normal-form/extensive form equivalence that do not seem irrational (esp. since reflection is satisfied)

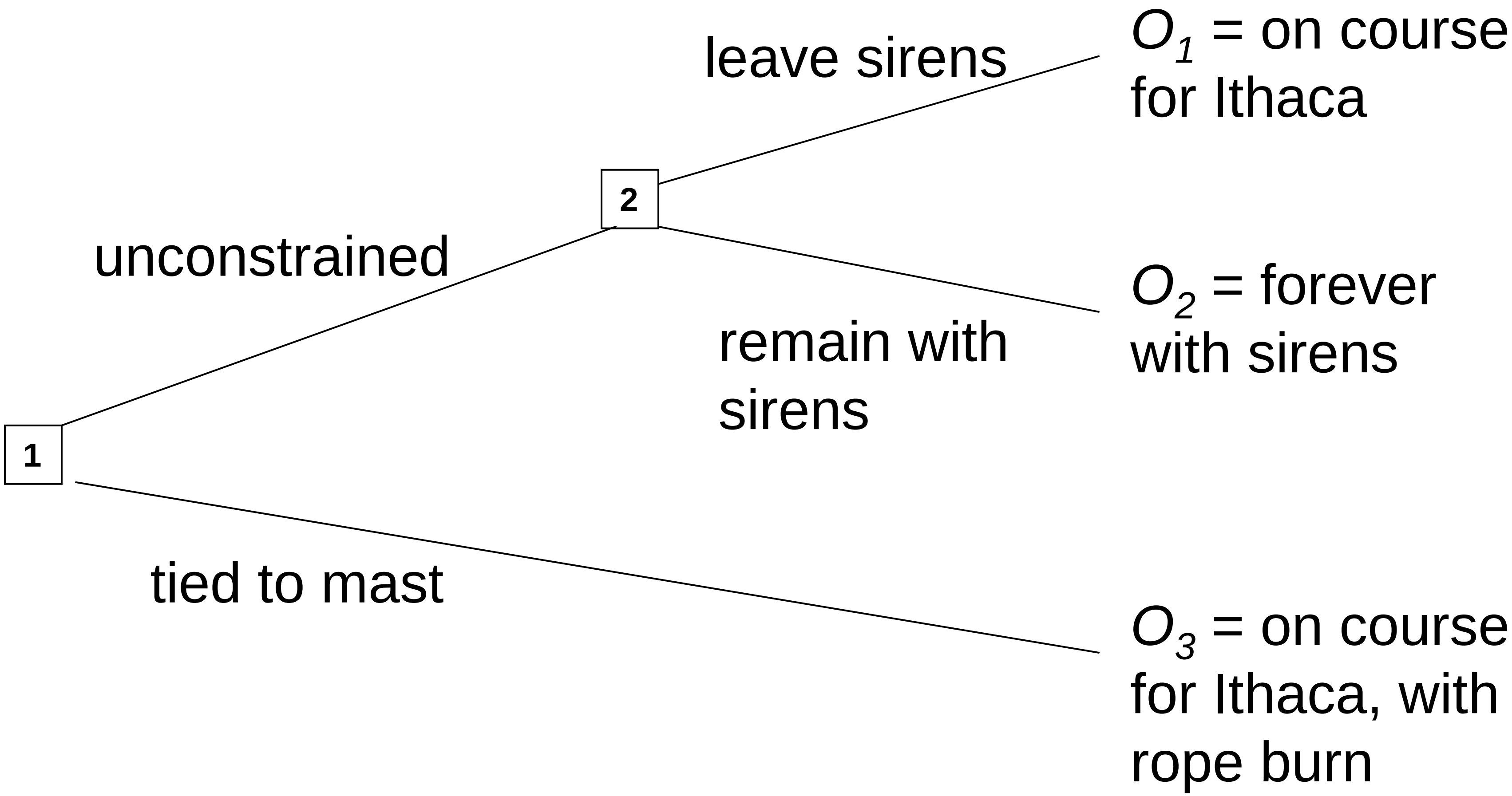
- Also, perhaps agent need not satisfy reflection in cases where they predict a Jeffrey-style probability/utility change
- Might they then hope/plan for such a change?

3. Higher-order desire and Ulysses

- Back to the traditional Ulysses decision tree…
- Could it be the case that Ulysses actually prefers his later preferences?
- That is, he might have a higher order desire contrary to the way the story is generally told…

As the story is told:	P1 > P2
But perhaps:	P2 > P1

Higher-order desire and Ulysses



Preferences at 1 ***(P1)****:* $O_1 > O_3 > O_2$

Preferences at 2 ***(P2)****:* $O_2 > O_1 > O_3$

Higher-order desire and Ulysses

- A complication to solving the sequential-choice problem…

  …OR reason to think that HO preferences must simply favour one’s current preferences?

- Jeffrey, at least, thinks P2 > P1 would be legitimate, but only in certain settings.

4. Jeffrey's 'Preferences Among Preferences'

- Argues for a distinction between the following 2 preference orderings (decreasing preference down the page):

1.	$\neg S$ pref S	2.	S
	S		$\neg S$ pref S
	$\neg S$		$\neg S$
	$\neg S$ pref S		$\neg S$ pref S

- Jeffrey: 2 is rational but 1 is not.

Jeffrey's 'Preferences Among Preferences'

Jeffrey on Ulysses case:

1. $P2$ ($O2 > O1 > O3$)
   $O3$
   $O2$
   $P1$ ($O1 > O3 > O2$)

2. $O3$
   $P2$ ($O2 > O1 > O3$)
   $O2$
   $P1$ ($O1 > O3 > O2$)

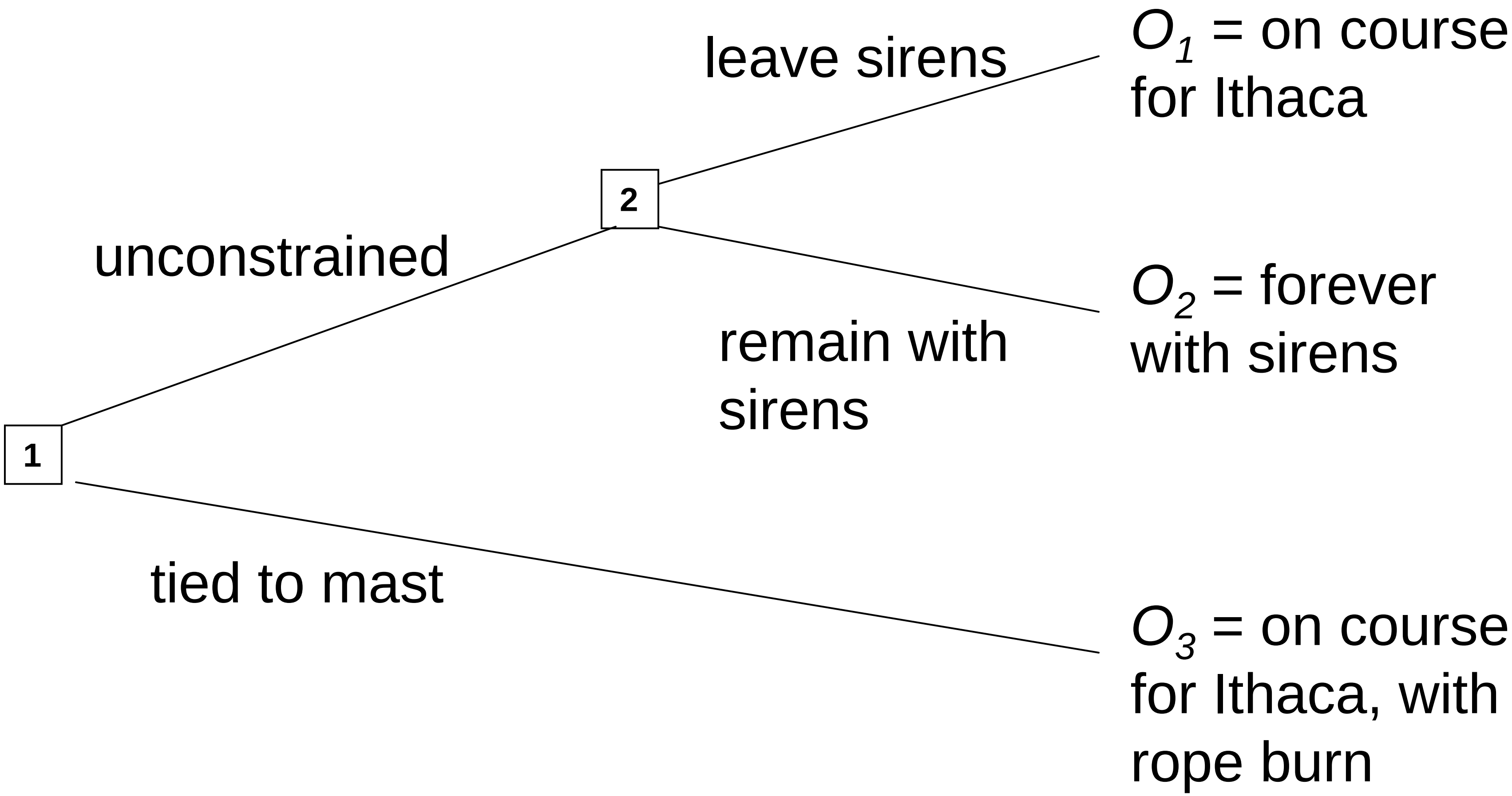
- Jeffrey: 2 is rational but 1 is not.

Jeffrey's 'Preferences Among Preferences'

- In favour of Jeffrey's analysis, it does not look to require any change to the way we solve sequential-choice problems.

  And yet it permits substantive higher-order preference—higher-order prefs that differ from current preferences.

Higher-order desire and Ulysses



Preferences at 1 ***(P1)****:* $O_1 > O_3 > O_2$

Preferences at 2 ***(P2)****:* $O_2 > O_1 > O_3$

Jeffrey's 'Preferences Among Preferences'

Against Jeffrey: Either both orderings 1 and 2 are reasonable, or neither is.

Informal perspective

- Surely HO prefs should always be in sync with 1st-order prefs.

Formal perspective:

- Does it even make sense to have these higher-order preference propositions in our preference ordering?
- Do such propositions have news value?

What is at stake again?

If Ulysses can sensibly prefer his future preferences to his current ones (according to Jeffrey's ordering 1), then

- There is a case that he should defer to his future preferences in the decision problem
- This would be a *planned* change in desire