Sequential Extensions of Causal and Evidential Decision Theory

Tom Everitt, Jan Leike, and Marcus Hutter

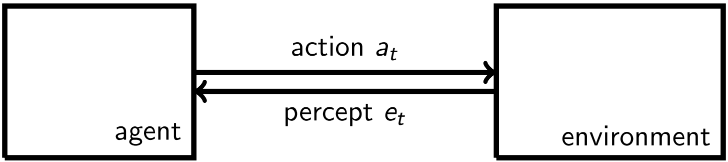
http://jan.leike.name/

Australian National University

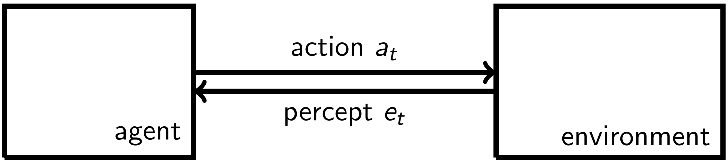
ADT'15 — 29 September 2015

Outline

Dualistic Agent Model

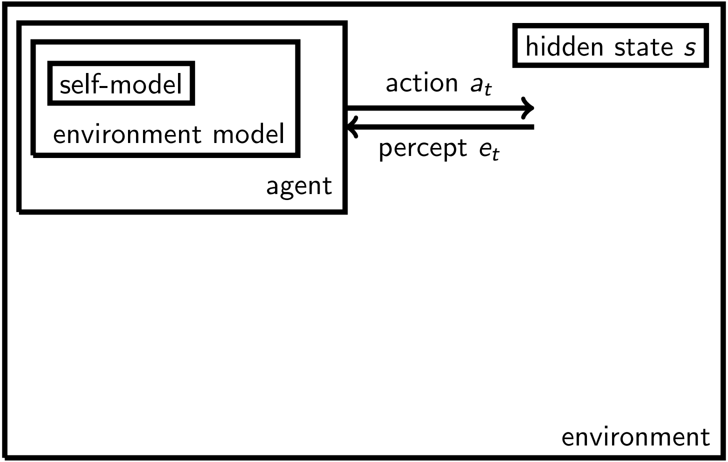


Dualistic Agent Model

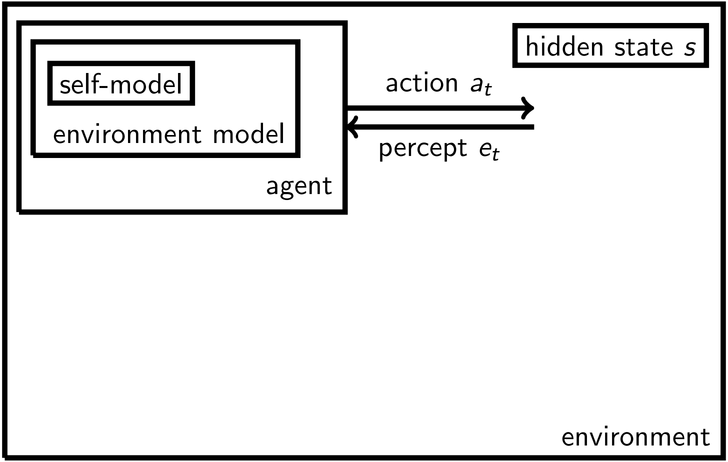


Goal: maximize expected utility $\mathbb{E}[\sum_{t=1}^{m} u(e_t)]$

Physicalistic Agent Model



Physicalistic Agent Model



Goal: maximize expected utility $\mathbb{E}[\sum_{t=1}^{m} u(e_t)]$

Outline

Newcomb's Problem

Presented by [Nozick, 1969]

Actions: (1) take the opaque box **or** (2) take both boxes

Reasoning Causally

Causal decision theory (CDT):

take the action that *causes* the best outcome

Reasoning Causally

Causal decision theory (CDT):

take the action that *causes* the best outcome

$$\arg\max_{a \in \mathcal{A}} \sum_{e \in \mathcal{E}} \mu(e \mid \mathtt{do}(a))\, u(e) \tag{CDT}$$

[Gibbard and Harper, 1978, Lewis, 1981, Skyrms, 1982, Joyce, 1999, Weirich, 2012]

Reasoning Causally

Causal decision theory (CDT):

take the action that *causes* the best outcome

$$\arg\max_{a \in \mathcal{A}} \sum_{e \in \mathcal{E}} \mu(e \mid \mathtt{do}(a))\, u(e) \tag{CDT}$$

[Gibbard and Harper, 1978, Lewis, 1981, Skyrms, 1982, Joyce, 1999, Weirich, 2012]

In Newcomb's problem: taking both boxes *causes* you to have $1000 more

Reasoning Evidentially

Evidential decision theory (EDT):

take the action that gives the best news about the outcome

Reasoning Evidentially

Evidential decision theory (EDT):

take the action that gives the best news about the outcome

$$\arg\max_{a \in \mathcal{A}} \sum_{e \in \mathcal{E}} \mu(e \mid a)\, u(e) \tag{EDT}$$

[Jeffrey, 1983, Briggs, 2014, Ahmed, 2014]

Reasoning Evidentially

Evidential decision theory (EDT):

take the action that gives the best news about the outcome

$$\arg\max_{a \in \mathcal{A}} \sum_{e \in \mathcal{E}} \mu(e \mid a)\, u(e) \qquad \text{(EDT)}$$

[Jeffrey, 1983, Briggs, 2014, Ahmed, 2014]

In Newcomb's problem: taking just the opaque box is good news because that means it likely contains $1,000,000

Newcomblike Problems

= problems where your actions are *not* independent of the (unobservable) environment state

Newcomblike Problems

= problems where your actions are *not* independent of the (unobservable) environment state

Newcomblike problems are actually quite common!

Newcomblike Problems

= problems where your actions are *not* independent of the (unobservable) environment state

Newcomblike problems are actually quite common!

- People predict each other all the time

Newcomblike Problems

= problems where your actions are *not* independent of the (unobservable) environment state

Newcomblike problems are actually quite common!

- People predict each other all the time
- Prediction does not need to be perfect

Newcomblike Problems

= problems where your actions are *not* independent of the (unobservable) environment state

Newcomblike problems are actually quite common!

- People predict each other all the time
- Prediction does not need to be perfect
- Example: Environment that knows your source code

Newcomblike Problems

= problems where your actions are *not* independent of the (unobservable) environment state

Newcomblike problems are actually quite common!

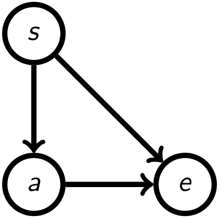
- People predict each other all the time
- Prediction does not need to be perfect
- Example: Environment that knows your source code
- Example: Multi-Agent setting with multiple copies of one agent

Outline

Sequential Decision Making

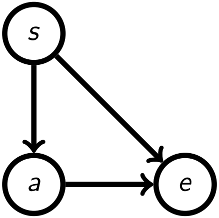
The Causal Graph

One-shot:

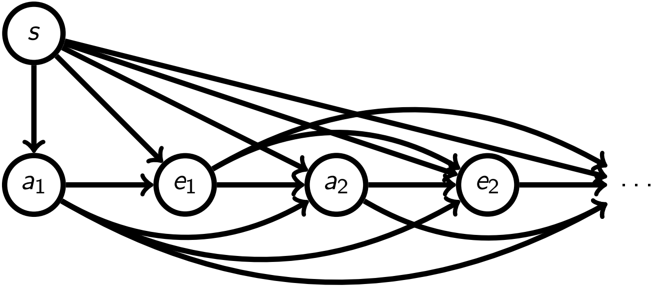


The Causal Graph

One-shot:



Sequential:



Notation

- $æ_{<t} = a_1 e_1 \dots a_{t-1} e_{t-1}$ denotes the history
- $\mu : (\mathcal{A} \times \mathcal{E})^* \times \mathcal{A} \rightarrow \Delta(\mathcal{E})$ denotes the environment model
- $\pi : (\mathcal{A} \times \mathcal{E})^* \rightarrow \mathcal{A}$ is my policy
- $m \in \mathbb{N}$ is the horizon

Sequential Evidential Decision Theory

- $æ_{<t} = a_1 e_1 \dots a_{t-1} e_{t-1}$ denotes the history
- $\mu : (\mathcal{A} \times \mathcal{E})^* \times \mathcal{A} \to \Delta(\mathcal{E})$ denotes the environment model
- $\pi : (\mathcal{A} \times \mathcal{E})^* \to \mathcal{A}$ is my policy
- $m \in \mathbb{N}$ is the horizon

Sequential Evidential Decision Theory

- $æ_{<t} = a_1 e_1 \dots a_{t-1} e_{t-1}$ denotes the history
- $\mu : (\mathcal{A} \times \mathcal{E})^* \times \mathcal{A} \to \Delta(\mathcal{E})$ denotes the environment model
- $\pi : (\mathcal{A} \times \mathcal{E})^* \to \mathcal{A}$ is my policy
- $m \in \mathbb{N}$ is the horizon

Sequential action-evidential decision theory (SAEDT):

$$V^{\text{aev}}(æ_{<t} a_t) := \sum_{e_t} \underbrace{\mu(e_t \mid æ_{<t} a_t)}_{\mu(e_t \mid \text{past}, a_t)} \underbrace{\Big(u(e_t) + V^{\text{aev}}(æ_{<t} a_t e_t) \Big)}_{\text{future utility}}$$

Sequential Evidential Decision Theory

- $æ_{<t} = a_1 e_1 \dots a_{t-1} e_{t-1}$ denotes the history
- $\mu : (\mathcal{A} \times \mathcal{E})^* \times \mathcal{A} \to \Delta(\mathcal{E})$ denotes the environment model
- $\pi : (\mathcal{A} \times \mathcal{E})^* \to \mathcal{A}$ is my policy
- $m \in \mathbb{N}$ is the horizon

Sequential action-evidential decision theory (SAEDT):

$$V^{\mathrm{aev}}(æ_{<t} a_t) := \sum_{e_t} \underbrace{\mu(e_t \mid æ_{<t} a_t)}_{\mu(e_t \mid \text{past}, a_t)} \underbrace{\Big(u(e_t) + V^{\mathrm{aev}}(æ_{<t} a_t e_t) \Big)}_{\text{future utility}}$$

Sequential policy-evidential decision theory (SPEDT):

$$V^{\mathrm{pev}}(æ_{<t} a_t) := \sum_{e_t} \underbrace{\mu(e_t \mid æ_{<t} a_t, \pi_{t+1:m})}_{\mu(e_t \mid \text{past}, \pi)} \underbrace{\Big(u(e_t) + V^{\mathrm{pev}}(æ_{<t} a_t e_t) \Big)}_{\text{future utility}}$$

Sequential Causal Decision Theory

- $æ_{<t} = a_1 e_1 \dots a_{t-1} e_{t-1}$ denotes the history
- $\mu : (\mathcal{A} \times \mathcal{E})^* \times \mathcal{A} \to \Delta(\mathcal{E})$ denotes the environment model
- $\pi : (\mathcal{A} \times \mathcal{E})^* \to \mathcal{A}$ is my policy
- $m \in \mathbb{N}$ is the horizon

Sequential Causal Decision Theory

- $æ_{<t} = a_1 e_1 \dots a_{t-1} e_{t-1}$ denotes the history
- $\mu : (\mathcal{A} \times \mathcal{E})^* \times \mathcal{A} \to \Delta(\mathcal{E})$ denotes the environment model
- $\pi : (\mathcal{A} \times \mathcal{E})^* \to \mathcal{A}$ is my policy
- $m \in \mathbb{N}$ is the horizon

Sequential causal decision theory (SCDT):

$$V^{\text{cau}}(æ_{<t} a_t) := \sum_{e_t \in \mathcal{E}} \underbrace{\mu(e_t \mid æ_{<t}, \text{do}(a_t))}_{\mu(e_t \mid \text{past}, \text{do}(a_t))} \underbrace{\Big(u(e_t) + V^{\text{cau}}(æ_{<t} a_t e_t) \Big)}_{\text{future utility}}$$

Sequential Causal Decision Theory

- $æ_{<t} = a_1 e_1 \dots a_{t-1} e_{t-1}$ denotes the history
- $\mu : (\mathcal{A} \times \mathcal{E})^* \times \mathcal{A} \to \Delta(\mathcal{E})$ denotes the environment model
- $\pi : (\mathcal{A} \times \mathcal{E})^* \to \mathcal{A}$ is my policy
- $m \in \mathbb{N}$ is the horizon

Sequential causal decision theory (SCDT):

$$V^{\text{cau}}(æ_{<t} a_t) := \sum_{e_t \in \mathcal{E}} \underbrace{\mu(e_t \mid æ_{<t}, \mathtt{do}(a_t))}_{\mu(e_t|\text{past},\mathtt{do}(a_t))} \underbrace{\Big(u(e_t) + V^{\text{cau}}(æ_{<t} a_t e_t)\Big)}_{\text{future utility}}$$

Proposition (Policy-Causal = Action-Causal). For all histories $æ_{<t}$ and percepts e_t: $\mu(e_t \mid æ_{<t}, \mathtt{do}(a_t)) = \mu(e_t \mid æ_{<t}, \mathtt{do}(\pi_{t:m}))$.

Outline

Examples

	action-evidential	policy-evidential	causal
Newcomb	✓	✓	×
Newcomb w/ precommit	✓	✓	×
Newcomb w/ looking	×	×	×
Toxoplasmosis	×	×	✓
Seq. Toxoplasmosis	×	×	✓

Formal description in [Everitt et al., 2015] and
source code at `http://jan.leike.name`

Conclusion

- How should physicalistic agents make decisions?

Conclusion

- How should physicalistic agents make decisions?
- Answer from philosophy: EDT, CDT

Conclusion

- How should physicalistic agents make decisions?
- Answer from philosophy: EDT, CDT
- Extended to sequential decision making

Conclusion

- How should physicalistic agents make decisions?
- Answer from philosophy: EDT, CDT
- Extended to sequential decision making

Which decision theory is better?

Conclusion

- How should physicalistic agents make decisions?
- Answer from philosophy: EDT, CDT
- Extended to sequential decision making

Which decision theory is better?

- In the end it matters whether you *win* (get the most utility)

Conclusion

- How should physicalistic agents make decisions?
- Answer from philosophy: EDT, CDT
- Extended to sequential decision making

Which decision theory is better?

- In the end it matters whether you *win* (get the most utility)
- Neither EDT nor CDT win on every example

Conclusion

- How should physicalistic agents make decisions?
- Answer from philosophy: EDT, CDT
- Extended to sequential decision making

Which decision theory is better?

- In the end it matters whether you *win* (get the most utility)
- Neither EDT nor CDT win on every example
- Neither EDT nor CDT model the environment containing themselves

Conclusion

- How should physicalistic agents make decisions?
- Answer from philosophy: EDT, CDT
- Extended to sequential decision making

Which decision theory is better?

- In the end it matters whether you *win* (get the most utility)
- Neither EDT nor CDT win on every example
- Neither EDT nor CDT model the environment containing themselves
- How physicalistic agents make decisions optimally is unsolved

Conclusion

- How should physicalistic agents make decisions?
- Answer from philosophy: EDT, CDT
- Extended to sequential decision making

Which decision theory is better?

- In the end it matters whether you *win* (get the most utility)
- Neither EDT nor CDT win on every example
- Neither EDT nor CDT model the environment containing themselves
- How physicalistic agents make decisions optimally is unsolved
- We need a better decision theory! E.g. timeless decision theory [Yudkowsky, 2010] or updateless decision theoy [Soares and Fallenstein, 2014]

Outline

References I

Ahmed, A. (2014).
Evidence, Decision and Causality.
Cambridge University Press.

Briggs, R. (2014).
Normative theories of rational choice: Expected utility.
In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Fall 2014 edition.

Everitt, T., Leike, J., and Hutter, M. (2015).
Sequential extensions of causal and evidential decision theory.
Technical report, Australian National University.
http://arxiv.org/abs/1506.

Gibbard, A. and Harper, W. L. (1978).
Counterfactuals and two kinds of expected utility.
In *Foundations and Applications of Decision Theory*, pages 125–162. Springer.

References II

Jeffrey, R. C. (1983).
The Logic of Decision.
University of Chicago Press, 2nd edition.

Joyce, J. M. (1999).
The Foundations of Causal Decision Theory.
Cambridge University Press.

Lewis, D. (1981).
Causal decision theory.
Australasian Journal of Philosophy, 59(1):5–30.

Nozick, R. (1969).
Newcomb's problem and two principles of choice.
In *Essays in honor of Carl G. Hempel*, pages 114–146. Springer.

Pearl, J. (2009).
Causality.
Cambridge University Press, 2nd edition.

References III

Skyrms, B. (1982).
Causal decision theory.
The Journal of Philosophy, pages 695–711.

Soares, N. and Fallenstein, B. (2014).
Toward idealized decision theory.
Technical report, Machine Intelligence Research Institute.
`http://intelligence.org/files/TowardIdealizedDecisionTheory.pdf`.

Weirich, P. (2012).
Causal decision theory.
In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Winter 2012 edition.

Yudkowsky, E. (2010).
Timeless decision theory.
Technical report, Machine Intelligence Research Institute.
`http://intelligence.org/files/TDT.pdf`.