Department of Philosophy, Tsinghua University

A Logic for Desire Based on Causal Inference

Kaibo Xie & Jialiang Yan
xiekaibozju@gmail.com
yan-ji19@mails.tsinghua.edu.cn

8th International Conference on Logic, Rationality and Interaction (LORI-VIII)
October 16, 2021

Outline

Motivation

Desire is the primitive component that drives the agent's behavior. Theories of desire have been established in many disciplines. We study the reasoning about desire in order to predict and characterize tendency of the agent's action.

Basic idea

In this paper, we propose an interpretation of "desiring ϕ" based on all of the most relevant alternatives that satisfies ϕ which are preferred by the agent. The relevance is evaluated by the causality. And then a logic based on causal inferences is given.

Having a desire means, as we explain, this desire can bring about what the agent most prefers at the moment.

There are two main feature capturing the desire in our proposal:

- What people desire is not for its own sake.
- Causality should be taken into consideration.

Example-1

Robin was in charge of an important project in the company recently, and the tremendous pressure made him sleepless every night. Robin is a healthy - conscious person and poor sleep will make him sluggish the next day which bothers Robin most. So having a good sleep is what he needs most at the moment. Therefore, on this sleepless night, Robin wants to take some sleep-helping pills to help him sleep well, even if these pills have some side effects.

Example-1

- P: “pills are taken”
- S: “Robin sleeps well”

Now assume that there are four possible worlds:

- w_1: $P \wedge \neg S$
- w_2: $\neg P \wedge S$
- w_3: $P \wedge S$
- w_4: $\neg P \wedge \neg S$

The preference structure can be characterized by the following preference order:

$$w_1 < w_4 < w_3 < w_2$$

Example-2

Robin's company offered him a chance to study in the Netherlands and he attached great importance to it. But his savings are small, so he doesn't want to pay for the tuition. Unfortunately, due to the epidemic, he cannot go abroad. Even after paying the tuition, he can only stay at home and take online lessons.

(The original version of this example comes from Maria Aloni.)

Example-2

- P': "paying the tuition"
- S': "studying in Netherlands"

Now assume that there are four possible worlds:

- w'_1: $P' \wedge \neg S'$
- w'_2: $\neg P' \wedge S'$
- w'_3: $P' \wedge S'$
- w'_4: $\neg P' \wedge \neg S'$

The preference structure can be characterized by the following preference order:

$$w'_1 < w'_4 < w'_3 < w'_2$$

The Basic Idea of Modelling

To capture these two features of instrumental desire, we introduce a **desire-causality model** in our paper.

- Following the traditional approach of formalizing preference, e.g. [von Wright, 1972] and [van Benthem & Liu, 2007], we use a pre order over all the possible worlds to represent the preference structure of an agent.
- We adopt a causal model which makes use of the interventionist approach to causality from [Halpern, 2000] and [Pearl, 2002].

Halpern (2000) and Pearl (2002)

Causal Variables

- $\mathcal{U} = \{U_1, \ldots, U_m\}$ is a set of *exogenous* variables,
- $\mathcal{V} = \{V_1, \ldots, V_n\}$ is a set of *endogenous* variables
- $\mathcal{R}(X)$ is the non-empty range of the variable $X \in \mathcal{U} \cup \mathcal{V}$.

Desire-Causality Model

Let $\mathcal{S} = \langle \mathcal{U}, \mathcal{V}, \mathcal{R} \rangle$. A desire-causality model for $\mathcal{S}$ is a tuple $\langle \mathcal{F}, \mathcal{A}, \leq \rangle$.

- **Structural Functions** $\mathcal{F} = \{\mathcal{F}_{V_j} \mid V_j \in \mathcal{V}\}$. For each endogenous variable V_j, $\mathcal{F}_{V_j}$ is a mapping from all assignments to $\mathcal{U} \cup \mathcal{V} \setminus \{V_j\}$ to $\mathcal{R}(V_j)$. $\mathcal{F}$ is assumed to be recursive.
- **Valuation Function** $\mathcal{A}$ assigns to every $X \in \mathcal{U} \cup \mathcal{V}$ a value $\mathcal{A}(X) \in \mathcal{R}(X)$. $\mathcal{A}$ has to *comply with* $\mathcal{F}_{V_j}$.
- **Preference Ordering** $\leq$ is a pre order over all possible assignments to $\mathcal{U} \cup \mathcal{V}$.

Language for instrumental desire

Language of Logic ID

Formulas ϕ of the language $\mathcal{L}$ based on $\mathcal{S}$

$$\phi ::= X{=}x \mid \neg\phi \mid \phi \wedge \phi \mid D(\vec{X}{=}\vec{x}) \mid (\vec{X}{=}\vec{x}) \boxdot\!\!\rightarrow \phi$$

Adopting interventionist approach

The semantics of "desire $X = x$": after an intervention forcing $X = x$ to be true, the world results from it will be more preferred than the current world.

Intervention on desire-causality models

Let $M = \langle \mathcal{F}, \mathcal{A}, \leq \rangle$ be a DC-model based on $\mathcal{S}$.
$M_{\vec{X}=\vec{x}} = \langle \mathcal{F}_{\vec{X}=\vec{x}}.\mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}, \leq \rangle$ is the DC model resulting from an intervention setting the values of variables in $X_1, \ldots, X_n$ to $x_1, \ldots, x_n$:

- $\mathcal{F}_{\vec{X}=\vec{x}}$ is as $\mathcal{F}$ except that, for each endogenous variable X_i in $\vec{X}$, the function f_{X_i} is replaced by a *constant* function f'_{X_i} that returns the value x_i regardless of the values of all other variables.
- $\mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}$ is the updated solution to the updated structural equations.

Truth Conditions

Let $\langle \mathcal{F}, \mathcal{A}, \leq \rangle$ be a DC-model based on $\mathcal{S}$

Truth condition of formulas in $\mathcal{L}$

- $\langle \mathcal{F}, \mathcal{A}, \leq \rangle \models X{=}x$ iff $\mathcal{A}(X) = x$
- $\langle \mathcal{F}, \mathcal{A}, \leq \rangle \models (\vec{X}{=}\vec{x}) \boxdot\!\!\rightarrow \phi$ iff $\langle \mathcal{F}_{\vec{X}=\vec{x}}, \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}, \leq \rangle \models \phi$
- $\langle \mathcal{F}, \mathcal{A}, \leq \rangle \models D(\vec{X}{=}\vec{x})$ iff $\mathcal{A}^{\mathcal{F}} < \mathcal{A}^{\mathcal{F}}_{\vec{X}=\vec{x}}$

X stands for taking sleep-helping pills. Y stands for having a good sleep.
X is an exogenous variable.
$\mathcal{F}_Y$: the value of Y is equal to the value of X

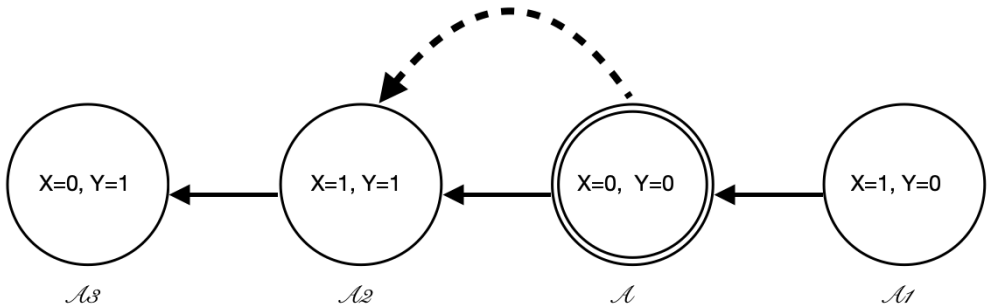

Figure: Robin's desire of pills

- $\mathcal{A}(X) = 0, \mathcal{A}(Y) = 0$
- According to $\mathcal{F}_{X=1}$, $\mathcal{A}_{X=1}(X) = 1$, $\mathcal{A}_{X=1}(Y) = 1$

X stands for taking sleep-helping pills. Y stands for having a good sleep.
X is an exogenous variable.
$\mathcal{F}_Y$: the value of Y is equal to the value of X

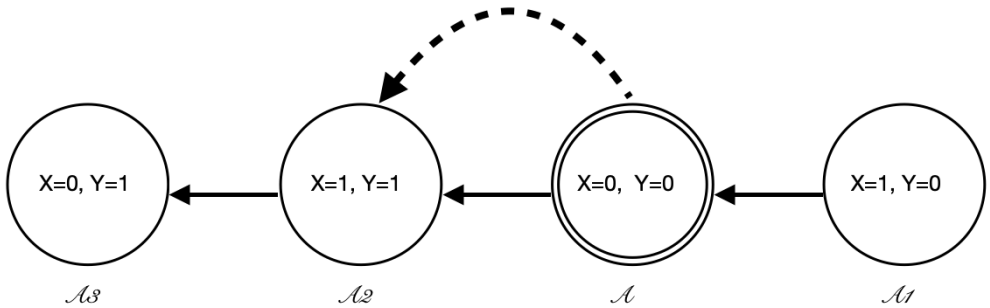

Figure: Robin's desire of pills

- $\mathcal{A}(X) = 0, \mathcal{A}(Y) = 0$
- According to $\mathcal{F}_{X=1}$, $\mathcal{A}_{X=1}(X) = 1, \mathcal{A}_{X=1}(Y) = 1$
- $\mathcal{A} < \mathcal{A}_{X=1}$, therefore $\langle \mathcal{F}, \mathcal{A}, \leq \rangle \models D(X = 1)$

S stands for studying in Netherlands; T stands for paying tuition fee. S is an exogenous variable.
$\mathcal{F}_T$: the value of T is equal to the value of S

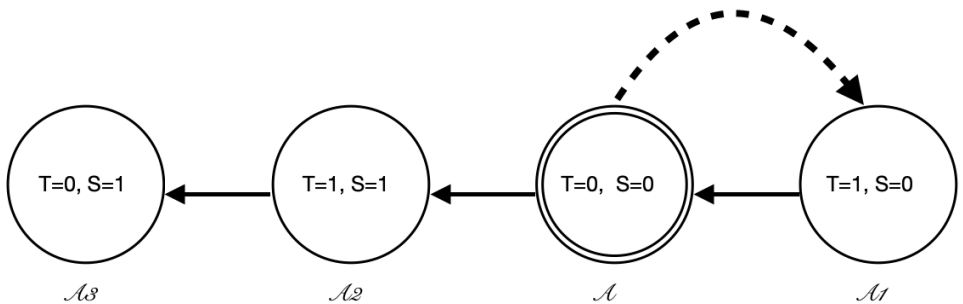

Figure: Robin's desire of tuition

- $\mathcal{A}(S) = 0, \mathcal{A}(T) = 0;$
- According to $\mathcal{F}_{T=1}$, $\mathcal{A}_{T=1}(T) = 1$, $\mathcal{A}_{T=1}(S) = 0$

S stands for studying in Netherlands; T stands for paying tuition fee. S is an exogenous variable.
$\mathcal{F}_T$: the value of T is equal to the value of S

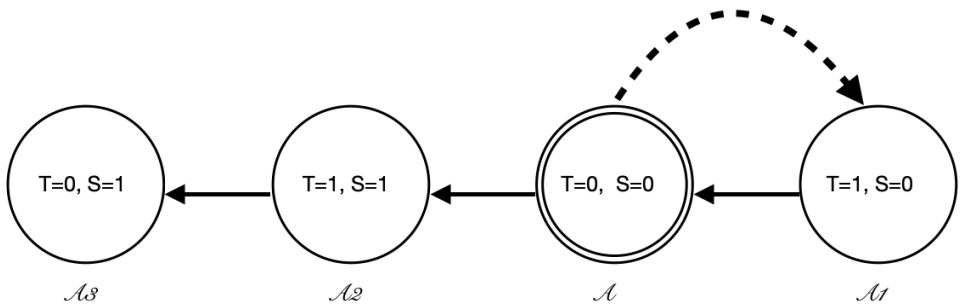

Figure: Robin's desire of tuition

- $\mathcal{A}(S) = 0, \mathcal{A}(T) = 0$;
- According to $\mathcal{F}_{T=1}$, $\mathcal{A}_{T=1}(T) = 1, \mathcal{A}_{T=1}(S) = 0$
- $\mathcal{A}_{S=1} < \mathcal{A}$, therefore $\langle \mathcal{F}, \mathcal{A}, \leq \rangle \models \neg D(T = 1)$

Expected Outcomes

Valid formulas of Logic ID

- $\vec{X}{=}\vec{x} \boxdot\!\!\rightarrow D(\vec{Y}{=}\vec{y}) \rightarrow D(\vec{X}{=}\vec{x} \wedge \vec{Y}{=}\vec{y}))$, if $\vec{X}$ and $\vec{Y}$ are disjoint
- $(\vec{X}{=}\vec{x} \boxdot\!\!\rightarrow \vec{Y}{=}\vec{y}) \wedge D(\vec{X}{=}\vec{x}) \rightarrow D(\vec{X}{=}\vec{x} \wedge \vec{Y}{=}\vec{y})$, if $\vec{X}$ and $\vec{Y}$ are disjoint

Axiomatization

Axioms of Logic ID:

- (A modified version of) Axioms for counterfactuals based on causal models in Halpern (2000)
- Reduction axioms for counterfactuals
- Axiom for the interaction between desire and conditionals

Axioms of Logic ID:

- (A modified version of) Axioms for counterfactuals based on causal models in Halpern (2000)
- Reduction axioms for counterfactuals
- Axiom for the interaction between desire and conditionals

P	φ	for φ an instance of a propositional tautology
MP	From φ_1 and $\varphi_1 \to \varphi_2$ infer φ_2	
Generalization	From ϕ infer $\vec{X} = \vec{x} \boxdot\!\!\rightarrow \phi$	
$\mathbf{A_1}$	$\vec{X}{=}\vec{x} \boxdot\!\!\rightarrow Y{=}y \to \neg \vec{X}{=}\vec{x} \boxdot\!\!\rightarrow Y{=}y'$	for $y, y' \in \mathcal{R}(Y)$ with $y \neq y'$
$\mathbf{A_2}$	$\bigvee_{y \in \mathcal{R}(Y)} \vec{X}{=}\vec{x} \boxdot\!\!\rightarrow Y{=}y$	
$\mathbf{A_3}$	$\big(\vec{X}{=}\vec{x} \boxdot\!\!\rightarrow (Y{=}y) \wedge \vec{X}{=}\vec{x} \boxdot\!\!\rightarrow (Z{=}z)\big) \to \vec{X}{=}\vec{x}, Y{=}y \boxdot\!\!\rightarrow (Z{=}z)$	
$\mathbf{A_4}$	$\vec{X}{=}\vec{x}, Y{=}y \boxdot\!\!\rightarrow (Y{=}y)$	
$\mathbf{A_5}$	$\big(\vec{X}{=}\vec{x}, Y{=}y \boxdot\!\!\rightarrow (Z{=}z) \wedge \vec{X}{=}\vec{x}, Z{=}z \boxdot\!\!\rightarrow (Y{=}y)\big) \to \vec{X}{=}\vec{x} \boxdot\!\!\rightarrow (Z{=}z)$	for $Y \neq Z$
$\mathbf{A_6}$	$(X_0 \rightsquigarrow X_1 \wedge \cdots \wedge X_{k-1} \rightsquigarrow X_k) \to \neg(X_k \rightsquigarrow X_0)$	
$\mathbf{A_7}$	$U{=}u \leftrightarrow \vec{X}{=}\vec{x} \boxdot\!\!\rightarrow U{=}u$	for $U \in \mathcal{U}$ with $U \notin \vec{X}$
$\mathbf{A_\neg}$	$\vec{X}{=}\vec{x} \boxdot\!\!\rightarrow \neg\varphi \leftrightarrow \neg \vec{X}{=}\vec{x} \boxdot\!\!\rightarrow \varphi$	
$\mathbf{A_\wedge}$	$\vec{X}{=}\vec{x} \boxdot\!\!\rightarrow (\varphi_1 \wedge \varphi_2) \leftrightarrow (\vec{X}{=}\vec{x} \boxdot\!\!\rightarrow \varphi_1 \wedge \vec{X}{=}\vec{x} \boxdot\!\!\rightarrow \varphi_2)$	
$\mathbf{A_{\boxdot\!\!\rightarrow}}$	$\vec{X}{=}\vec{x} \boxdot\!\!\rightarrow (\vec{Y}{=}\vec{y} \boxdot\!\!\rightarrow \varphi) \leftrightarrow \vec{X'}{=}\vec{x'}, \vec{Y}{=}\vec{y} \boxdot\!\!\rightarrow \varphi$	with $\vec{X'}{=}\vec{x'}$ the subassignment of $\vec{X}{=}\vec{x}$ for $\vec{X'} := \vec{X} \setminus \vec{Y}$

Axiomatization

$\mathbf{A_8}$	$\vec{X}{=}\vec{x} \to \neg D(\vec{X}{=}\vec{x})$	
$\mathbf{A_9}$	$(\vec{X} = \vec{x} \boxdot\!\!\to Z = z) \to (\vec{X} = \vec{x} \boxdot\!\!\to D(\vec{Y} = \vec{y})) \leftrightarrow (\vec{X} = \vec{x}, Z = z \boxdot\!\!\to D(\vec{Y} = \vec{y}))$	if $\vec{Y} = \vec{y}$ is a full assignment of $\mathcal{U} \cup \mathcal{V}$ and $Z \notin \vec{X}$
$\mathbf{A_{10}}$	$(\vec{X}{=}\vec{x} \boxdot\!\!\to \vec{Y}{=}\vec{y}) \to (D(\vec{X}{=}\vec{x}) \leftrightarrow D(\vec{X}{=}\vec{x} \wedge \vec{Y'}{=}\vec{y'}))$	with $\vec{Y'} = \vec{y'}$ the sub-assignment of $\vec{Y}{=}\vec{y}$ for $\vec{Y'} := \vec{Y} \setminus \vec{X}$
$\mathbf{A_{11}}$	$(\vec{X}{=}\vec{x} \boxdot\!\!\to D(\vec{Y}{=}\vec{y})) \wedge D(\vec{X}{=}\vec{x}) \to D(\vec{X'}{=}\vec{x'}, \vec{Y}{=}\vec{y})$	with $\vec{X'} = \vec{x'}$ the sub-assignment of $\vec{X}{=}\vec{x}$ for $\vec{X'} := \vec{X} \setminus \vec{Y}$

Desire for propositions in complex forms

Desire for propositions in complex forms

Semantics of counterfactuals with complex antecedent

Briggs (2012): $\chi \boxdot\!\!\rightarrow \phi$ holds iff ϕ holds at every submodels generated by interventions on M corresponding to states that verify χ

Desire for propositions in complex forms

Semantics of counterfactuals with complex antecedent

Briggs (2012): $\chi \boxdot\!\!\rightarrow \phi$ holds iff ϕ holds at every submodels generated by interventions on M corresponding to states that verify χ

Semantics of desire for complex propositions

Briggs (2012): $D(\chi)$ holds iff every assignment in submodels generated by interventions on M corresponding to states that verify χ is preferred.

$$(D(\phi_1 \vee \phi_2)) \leftrightarrow (D\phi_1 \wedge D\phi_2 \wedge (D(\phi_1 \wedge \phi_2))$$

THANK YOU!