Causal Reasoning from Almost First Principles (Extended Abstract)

The aim of the paper consists in providing a principle-based description of causal reasoning, of its place in the general picture of reasoning, and its relations to nonmonotonic reasoning and knowledge representation.

1 Causal Theories and their Semantics

A *causal rule* is a rule of the form $a \Rightarrow A$ (a *causes* A) on propositions. A set of causal rules is a *causal theory*. The basic principle of causal reasoning is the following rationality postulate of acceptance for propositions:

Causal Acceptance Principle A proposition A is accepted wrt a causal theory Δ if and only if Δ contains a causal rule $a \Rightarrow A$ such that all propositions in a are accepted.

The two parts of the above principle could be expressed as two independent rationality postulates:

Preservation Principle If all propositions in a are accepted, and a causes A, then A should be accepted.

Principle of Sufficient Reason Any proposition should have a cause for its acceptance.

Preservation states that a causal rule should preserve acceptance of propositions. On a normative reading, it states that existence of reason is sufficient for acceptance. Leibniz' Principle of Sufficient Reason is also a normative principle of reasoning stating that propositions *require* reasons for their acceptance, and such reasons are provided by establishing their causes. The origins of this principle go back to the ancient law of causality (everything has a cause).

Rational Semantics. A *valuation* assigns either 1 ('truth') or 0 ('falsity') to every proposition of the language. If $v(A) = 1$, we say that proposition A is *accepted* in the valuation v. A valuation can be safely identified with its associated set of accepted propositions.

$\Delta(u)$ is the set of propositions that are directly caused by u in Δ: $\Delta(u) = \{A \mid a \Rightarrow A \in \Delta,\ a \subseteq u\}$.

Definition 1. • A *causal model* of a causal theory Δ is a valuation that satisfies the condition

$$v = \Delta(v).$$

• A *rational semantics* of a causal theory is the set of all its causal models.

The notion of a causal model provides precise formal expression of the Causal Acceptance principle.

2 Causal Inference

The underlying *logic* of causal reasoning can be described as follows.

Definition 2. A *causal inference relation* is a causal theory that is closed with respect to

Monotonicity If $a \Rightarrow A$ and $a \subseteq b$, then $b \Rightarrow A$;

Cut If $a \Rightarrow A$ and $a, A \Rightarrow B$, then $a \Rightarrow B$.

Causal inference incorporates two of the three postulates for Tarski consequence, but omits Reflexivity.

$\mathcal{C}(u)$ is the set of propositions caused by u with respect to a causal inference relation: $\mathcal{C}(u) = \{A \mid u \Rightarrow A\}$. This causal operator plays much the same role as the usual derivability operator for consequence relations. For a causal theory Δ, $\Rightarrow_\Delta$ is the least causal inference relation that includes Δ, while $\mathcal{C}_\Delta$ will denote the associated causal operator.

Causal inference constitutes the maximal logic for the rational semantics, though in a different sense from standard completeness (due to the fact that the rational semantics is *nonmonotonic*). More precisely, it has been shown that two causal theories Δ and Γ are equivalent with respect to causal inference if and only if they are *strongly semantically equivalent*: for any set Φ of causal rules, $\Delta \cup \Phi$ has the same rational semantics as $\Gamma \cup \Phi$.

Similarly to other formalisms of nonmonotonic reasoning, justification of accepted propositions is an essential part of causal reasoning. The law of causality leads to a problem known already in antiquity as the *Agrippan trilemma*: if you do not accept infinite regress of causation (or justification), you should accept either uncaused or self-caused propositions. Now, for causal reasoning, there are two kinds of propositions that can play, respectively, these two roles:

Definition 3. • A proposition A is an *axiom* of a causal theory Δ if the rule $\emptyset \Rightarrow A$ belongs to Δ;

• A proposition A is an *assumption* of a causal theory if the rule $A \Rightarrow A$ belongs to it.

In contrast to deductive reasoning, both axioms and assumptions provide reasonable end-points of justification in causal reasoning. Every axiom *must* belong to any causal model, whereas any assumption *can* be accepted when it is consistent with the rest of accepted propositions, but it does not have to be accepted.

Supraclassicality. In order to raise the above abstract formalism to a full-fledged reasoning system, we will require it to subsume classical entailment. Our language now will be a classical propositional language with the usual classical connectives and constants $\{\wedge, \vee, \neg, \rightarrow, \mathbf{t}, \mathbf{f}\}$, $\vDash$ will stand for the classical entailment while Th will denote the associated classical provability operator.

Definition 4. A causal inference relation is *supraclassical* if it satisfies the following additional rules:

(Strengthening) If $b \Rightarrow C$ and $a \vDash B$, for every $B \in b$, then $a \Rightarrow C$;

(Weakening) If $a \Rightarrow B$ and $B \vDash C$, then $a \Rightarrow C$;

(And) If $a \Rightarrow B$ and $a \Rightarrow C$, then $a \Rightarrow B \wedge C$;

(Truth) $\mathbf{t} \Rightarrow \mathbf{t}$;

(Falsity) $\mathbf{f} \Rightarrow \mathbf{f}$.

Definition 5.
- A *classical causal model* of a causal theory Δ is a consistent valuation that satisfies

$$v = \mathrm{Th}(\Delta(v)).$$

- A *supraclassical rational semantics* of a causal theory is the set of all its classical causal models.

Supraclassical causal inference constitutes an adequate logic for the supraclassical rational semantics.

3 Summary of the Rest of the Paper

The above formalism of the causal calculus has been shown to cover significant parts of nonmonotonic reasoning such as abduction and diagnosis, logic programming, and reasoning about action and change. The rest of the paper provides an overview of some of the key applications of this formalism both for representation of other formalisms of causal reasoning and a number of formalisms of nonmonotonic reasoning in AI.

Pearl's Structural Equation Models. Pearl's approach to causal reasoning can be viewed as an instantiation of the above theory via a modular translation of structural equations as causal rules. By this representation, each structural equality $v_i = f_i(pa_i, u_i)$ is translated as a causal rule

$$PA_i = pa_i, U_i = u_i \Rightarrow V_i = f_i(pa_i, u_i).$$

In the special case when all the relevant variables are Boolean, a Boolean structural equation $p = F$ (where F is classical logical formula) produces in this sense two causal rules

$$F \Rightarrow p \quad \text{and} \quad \neg F \Rightarrow \neg p.$$

Given this translation, Pearl's causal worlds correspond precisely to classical causal models of the associated causal theory that are (classical) *worlds*.

Defaults in causal reasoning. Default logic can be represented in the causal calculus by defining defaults as a special kind of assumptions that satisfy the following principle:

A default is an assumption that is accepted whenever it is not refuted.

In the framework of the rational semantics, the above principle boils down to

Default Bivalence *For any causal model v and any default assumption A, either $A \in v$ or $\neg A \in v$.*

Default negation and logic programming. Logic programming can be represented as a species of causal reasoning in which negative literal are defaults:

(Default Negation) $\neg p \Rightarrow \neg p$, for any propositional atom p.

The principle of sufficient reason is reduced in such systems to the necessity of explaining only positive facts. In this setting, a general program rule

$$\mathbf{not}\, d, c \leftarrow a, \mathbf{not}\, b \qquad (*)$$

is interpreted as the following causal rule:

$$d, \neg b \Rightarrow \bigwedge a \rightarrow \bigvee c.$$

Then a stable semantics of a logic program will coincide with the rational semantics of its translation. Moreover, any causal rule can be identified with some program rule under this interpretation. Accordingly, any causal theory in which negated atoms are defaults is reducible to a logic program, and vice versa.

4 Conclusions

The causal calculus has been shown to provide a formal basis for reasoning and problem-solving in many areas and applications of AI. Moreover, due to deep and natural connections of causes with reasons and explanations, causal reasoning brings with it the promise of Explainable AI, an approach to artificial intelligence that is not only practically successful but is also susceptible to rational explanation and justification.

The theory of causal reasoning described in this paper poses, however, a lot of questions and challenges for a general theory of reasoning. To begin with, being a nonmonotonic formalism, it is based on a unidirectional connection between the language (of causal rules) and its (rational) semantics, and this should force us to reconsider the basic notions associated with denotational approaches, such as truth and meaning of language expressions. It also puts into question the very possibility, or even desirability, of constructing causal reasoning or its semantics bottom up from propositional atoms.

In a more general perspective, the miracle of resurrection of causal reasoning in artificial intelligence and other important fields of science confirms once again that causation should be viewed as an essential part of our reasoning, a kind of reasoning that has deep, though almost forgotten, roots in human history. Our inferential approach to causation provides also natural connections of our theory with a general philosophical approach of inferentialism (Brandom).