ELEKTRİK, VOL.1, NO.3, October 1993, ©TÜBİTAK EMO

The Logic of Counteraction

Erkan TIN and Varol AKMAN
Department of Computer Engineering and Information Science
Bilkent University, Bilkent, Ankara 06533, Turkey

Abstract

We extend causal theories and study actions in domains involving multiple agents. Causal theories, invented by Yoav Shoham, are based on a temporal nonmonotonic logic and have computationally tractable aspects. Since Shoham's formalism does not provide an adequate mechanism for representing simultaneous actions and specifying their consequences, we introduce the notion of counteractions while preserving the efficiency and model-theoretic properties of causal theories.

1. Introduction

Commonsense knowledge for reasoning about actions has long been a central subject of artificial intelligence (AI) [10, 12]. It has been recognized that the existing formalisms do not provide satisfactory solutions to some fundamental problems, viz. the *frame problem* [7]. Moreover, it has been noticed that reasoning about actions in domains involving many agents is harder than domains involving a single agent.

To illustrate some of the difficulties, Hanks and McDermott [3] state what they call the *temporal projection problem* as follows: Given a description of the current situation, the effects of possible actions, and a sequence of actions to be performed, how can the properties of the world in the resulting situation be predicted? Noticing that this is not a by-product of situation calculus [7], but is independent of the formalism used, they redefine it [4, p. 385]: "Given an initial description of the world (some facts that are true), the occurrence of some events, and some notion of causality (that an event can cause a fact to become true), what facts are true once all the events have occurred?"

Hanks and McDermott [4] applied some of the available logics to the *Yale shooting problem* (YSP) to check whether the expected results are indeed produced. YSP is a paradigm to show how the temporal projection problem arises. At some point in time, a person (Fred) is alive. A loaded gun, after waiting for a while, is fired at Fred. As a result, one expects that Fred would die and the gun would be unloaded. But Hanks and McDermott [4] demonstrate, in the frameworks of default logic [9] and circumscription [5], that unintended minimal models are obtained: The gun gets unloaded during the waiting stage and thus firing the gun does not kill Fred. They also convincingly argue that a solution to the temporal projection problem relies on answers to two questions [4, p. 409]:

1. Given a logical theory that admits more than one model, what are the preferred models of that theory (i.e., what is the preference criterion)?

2. Given a theory and a preference criterion, how do we find the theorems that are true in all 'most preferred' models?

We believe that Shoham's preference criterion [13] provides, in a rather general sense, a satisfactory answer to the first question. Moreover, since Shoham gives an algorithm that computes the true sentences in the models preferred under his preference criterion, the second question is also answered in some sense.

In this paper we will demonstrate how Shoham's formalism can be adapted to obtain intended models in multi-agent domains where simultaneous actions are allowed. To this end, we will introduce the notion of counteractions. It will be shown that this yields intended models while preserving the model-theoretic properties and efficiency of causal theories [15].

2. Terminology

Unless otherwise stated, we follow Shoham's terminology and definitions almost verbatim [12].

2.1. Chronological ignorance

Nonmonotonic logics can be defined by means of a *preference criterion* on the interpretations of (classical or modal) propositional logic or first-order logic. This criterion induces a preference relation over the models. The *logic of chronological ignorance* (CI) is a nonmonotonic logic obtained in this way. The standard monotonic logic on which it is based is called the *logic of temporal knowledge* (TK). The syntax and semantics of the latter are given in the sequel where P is a set of primitive propositions, TV is a set of temporal variables, $TC = \mathcal{Z}$ (integers), and $U = TC \cup TV$.

Well-formed formulae (wff) are defined inductively as follows:

1. If $u_1, u_2 \in U$, then $u_1 = u_2, u_1 \leq u_2$ are wff.
2. If $u_1, u_2 \in U$ and $p \in P$, then $\top(u_1, u_2, p)$ is a wff. ($\top(u_1, u_2, p)$ means that proposition p is *true in the interval* $[u_1, u_2], u_1 \leq u_2$.)
3. If φ_1 and φ_2 are wff, then so are $\varphi_1 \wedge \varphi_2, \neg\varphi_1, \Box\varphi_1$. (Read $\Box\varphi$ as "φ is known.")
4. If φ is a wff and $v \in TV$, then $\forall v \varphi$ is a wff.

We employ the following abbreviations ("$[\neg]$" signifies that $\neg$ may or may not appear):

- $\Diamond$ stands for $\neg\Box\neg$. (Read $\Diamond\varphi$ as "φ is not known to be false.")
- $\Box(t_1, t_2, [\neg]p)$ stands for $\Box[\neg]\top(t_1, t_2, p)$.
- $\Diamond(t_1, t_2, [\neg]p)$ stands for $\Diamond[\neg]\top(t_1, t_2, p)$.
- $\top(t_1, p)$ stands for $\top(t_1, t_1, p)$.

Definition 2.1 A *sentence* is a wff containing no free variables.

Definition 2.2 A *Kripke interpretation* (κ) is a pair $\langle W, M\rangle$ where W is a nonempty universe of possible worlds, and M is a meaning function $M : P \rightarrow 2^{W \times \mathcal{Z} \times \mathcal{Z}}$.

Definition 2.3 A *variable assignment* is a function $VA : TV \rightarrow \mathcal{Z}$.

Definition 2.4 A *valuation function* VF is such that $VF(u) = VA(u)$ if $u \in TV$, and $VF(u) = u$ if $u \in TC$.

$\kappa = \langle W, M\rangle$ and a world $w \in W$ *satisfy a formula* φ *under variable assignment* VA (written $\kappa, w \models \varphi[VA]$) according to the following rules:

1. $\kappa, w \models u_1 = u_2[VA]$ if and only if (iff) $VF(u_1) = VF(u_2)$.
2. $\kappa, w \models u_1 \leq u_2[VA]$ iff $VF(u_1) \leq VF(u_2)$.
3. $\kappa, w \models \top(u_1, u_2, p)[VA]$ iff $\langle w, VF(u_1), VF(u_2)\rangle \in M(p)$.
4. $\kappa, w \models \varphi_1 \wedge \varphi_2[VA]$ iff $\kappa, w \models \varphi_1[VA]$ and $\kappa, w \models \varphi_2[VA]$.
5. $\kappa, w \models \neg\varphi[VA]$ iff $\kappa, w \not\models \varphi[VA]$.
6. $\kappa, w \models \forall v\varphi[VA]$ iff $\kappa, w \models \varphi[VA']$ for all VA' that agree with VA everywhere except possibly on v.
7. $\kappa, w \models \Box\varphi[VA]$ iff $\kappa, w' \models \varphi[VA]$ for all $w' \in W$. (Caveat: Thus we are allowed to write $\kappa \models \Box\varphi[VA]$ without ambiguity.)

$\kappa = \langle W, M\rangle$ and a world $w \in W$ are a *model* for a formula φ (written $\kappa, w \models \varphi$) if $\kappa, w \models \varphi[VA]$ for any variable assignment VA. A wff is *satisfiable* if it has a model, and *valid* if its negation has no model. It should be noted that if φ is true (respectively, is not true) in $w \in W$, then this is written $\kappa, w \models \varphi$ (respectively, $\kappa, w \not\models \varphi$).

Definition 2.5 *Base formulae* are those wff containing no occurrence of the modal symbols $\Box$ and $\Diamond$.

Definition 2.6 The *latest time point* (ltp) of a base sentence is the latest time point mentioned in it:

1. ltp of $\top(t_1, t_2, p) = t_2$.
2. ltp of $\varphi_1 \wedge \varphi_2 =$ maximum of the ltp's of φ_1 and φ_2.
3. ltp of $\neg\varphi =$ ltp of φ.
4. ltp of $\forall v\varphi =$ minimum of the ltp's of all φ' which result from substituting in φ a time point symbol for all free occurrences of v, or $-\infty$ if there is no such earliest ltp.

Definition 2.7 A Kripke interpretation M is *chronologically more ignorant* than a Kripke interpretation M' (written $M' \subset_{ci} M$) if there exists t_0 such that:

1. For any base sentence φ with ltp $\leq t_0$, if $M \models \Box\varphi$ then also $M' \models \Box\varphi$.
2. There exists a base sentence φ with ltp t_0 such that $M' \models \Box\varphi$ but $M \not\models \Box\varphi$.

Definition 2.8 M is said to be a *chronologically maximally ignorant* (cmi) model of φ if $M \models_{\subset_{ci}} \varphi$, i.e., if $M \models \varphi$ and there is no M' such that $M' \models \varphi$ and $M \subset_{ci} M'$.

Definition 2.9 The *logic of chronological ignorance* (CI) is the nonmonotonic logic obtained by associating the preference relation $\subset_{ci}$ with the monotonic logic of temporal knowledge TK. (As usual, the syntax of the two is identical.)

2.2. Causal theories

Definition 2.10 *Formulae* in CI are those base formulae augmented by the modal symbols.

Definition 2.11 A *theory* in CI is a collection of sentences in CI.

Definition 2.12 *Base sentences* in CI are those sentences not containing any occurrence of the modal symbols.

Definition 2.13 An *atomic base sentence* (abs) is of the form $[\neg]\top(t_1, t_2, p)$.

Definition 2.14 A *causal theory* Ψ is a theory in CI, in which all sentences have the form $\Phi \wedge \Theta \supset \Box\varphi$ where:

1. $\varphi = [\neg]\top(t_1, t_2, p)$.
2. $\Phi = \bigwedge_{i=1}^{n} \Box\varphi_i$, where φ_i is an abs with ltp $t_i < t_1$.
3. $\Theta = \bigwedge_{j=1}^{m} \Diamond\varphi_j$, where φ_j is an abs with ltp $t_j < t_1$.
4. Φ or Θ (or both) may be empty. A sentence in which Φ is empty is called a *boundary condition.* Other sentences are called *causal rules*.
5. There is a time point t_0 (global for Ψ) such that if $\Theta \supset \Box(t_1, t_2, [\neg]p)$ is a boundary condition, then $t_0 < t_1$. (Caveat: Causal theories have *unique* models, viz. any two cmi models of a causal theory agree on the truth value of all sentences $\Box\varphi$, where φ is a base sentence. The proof of this fact is by induction on time, and the basis of the induction is this t_0. Without this condition, causal theories could have multiple models.)
6. It is not permitted to have two sentences in Ψ such that one contains $\Diamond(t_1, t_2, p)$ in its antecedent part and the other contains $\Diamond(t_1, t_2, \neg p)$ in its antecedent part. (Caveat: This is necessitated by the *soundness conditions* of Ψ, which are the set of sentences $\Diamond(t_1, t_2, p) \supset \top(t_1, t_2, p)$ such that $\Diamond(t_1, t_2, p)$ appears in the antecedent part of some sentence in Ψ. Soundness conditions are (implicitly) part of the causal theories. Since soundness conditions force the interpretation of $\Diamond$-sentences as true-by-default, allowing such two sentences could yield causal theories with *multiple* models.)
7. If $\Phi_1 \wedge \Theta_1 \supset \Box(t_1, t_2, p)$ and $\Phi_2 \wedge \Theta_2 \supset \Box(t_1, t_2, \neg p)$ are two sentences in Ψ, then $\Phi_1 \wedge \Theta_1 \wedge \Phi_2 \wedge \Theta_2$ is inconsistent. (Caveat: This condition says that consistent causes cannot bring about inconsistent effects. Without this restriction some causal theories would have multiple models.)

Theorem 2.1 If Ψ is a causal theory, then Ψ has a cmi model. Besides, if M and M' are cmi models of Ψ, and φ is any abs, then $M \models \Box\varphi$ iff $M' \models \Box\varphi$.

Proof. [12, pp. 112–113].

Definition 2.15 A *time-bounded Kripke interpretation* M_t is a structure which can be viewed as an incomplete κ. Like a κ it assigns a truth value to atomic propositions, but only to those whose ltp $\leq t$. The truth value of an arbitrary sentence whose ltp $\leq t$ is also determined in M_t, according to the usual rules. If a sentence φ with ltp $\leq t$ is *satisfied* by M_t, this is denoted $M_t \models \varphi$.

Definition 2.16 M_t *partially satisfies* Ψ if M_t satisfies all sentences of Ψ whose ltp $\leq t$.

3. Causal Theories and Their Extension

Causal theories contain axioms to reason about the effects of actions [12, p. 108]. Proceeding in time, knowledge about the future is obtained from what is known (and what is not known) about the past. Causal theories have a nice property: All cmi models agree on what is known (Theorem 2.1), i.e., in all cmi models of a causal theory the same abs's are known.

Shoham [12, p. 106] gives the following axiomatization of YSP as a causal theory:

1. $\Box(1, \textit{loaded})$.

2. $\Box(5, fired)$.

3. $\Box(t, loaded) \wedge \Diamond(t, \neg fired) \wedge \Diamond(t, \neg emptied\_manually) \supset \Box(t+1, loaded)$, for all t.

4. $\Box(t, loaded) \wedge \Box(t, fired) \wedge \Diamond(t, air) \wedge \Diamond(t, firing\_pin) \wedge \Diamond(t, no\_marshmallow\_bullets) \wedge \ldots \wedge \Diamond$ *other mundane conditions* $\supset \Box(t+1, noise)$, for all t.

Axioms 1 and 2 are the boundary conditions "it is known that the gun is loaded at 1" and "it is known that the gun is fired at 5," respectively. Axiom 3 is an axiom schema needed for *persistence*. It says that the gun remains loaded unless certain conditions obtain, i.e., "if it is known that the gun is loaded at t, and it is not known that it is fired at t and that it is emptied manually at t, then it is known that the gun is loaded at $t+1$." Axiom 4 is again an axiom schema. It is a causal rule stating that firing a loaded gun causes a noise unless certain conditions obtain. Causal theories can only contain axioms, not axiom schemata with time variables (Definition 4.7 in [12, pp. 109–110]). Therefore, the axiom schemata 3 and 4 above must be replicated by replacing t with time points from 1 to 5.

Shoham's algorithm steps through each axiom—checking whether the antecedent of the axiom is satisfied—and computes the base sentences known in all cmi models of this causal theory. It produces the expected abs's:

$$\top(1, loaded), \top(2, loaded), \ldots, \top(5, loaded), \top(5, fired), \top(6, noise).$$

If the size of a finite causal theory is n, the algorithm yields the abs's known in all cmi models of the causal theory in time $O(n\log n)$ (Theorem 4.4 in [12, p. 114]).

Now we examine causal theories to see whether they can be used in computing the intended models when simultaneous actions are introduced. In the blocks world of Figure 1, there is a block initially located at the middle of the table (*center*). Two operations, *push_left* and *push_right*, can be executed in this world. Executing *push_left* (respectively *push_right*) moves the block to *left* (respectively *right*). It is assumed that the forces applied on the block are of equal magnitude when these operations are performed simultaneously. We do not consider some real world features such as friction and energy consumption.

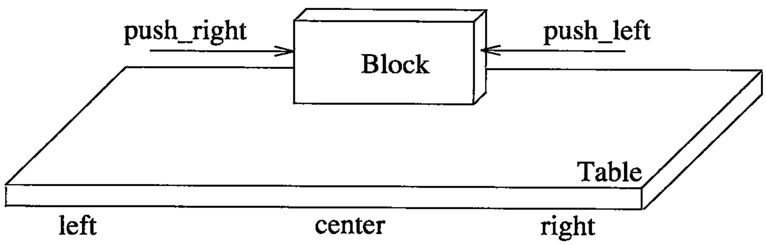

Figure 1. Simultaneous actions on a block.

Assuming that the block is at *center* at time 1, and *push_left* and *push_right* are simultaneously executed at time 1, we have the following causal theory:

1. $\Box(1, center)$.

2. $\Box(1, push\_left)$.

3. $\Box(1, push\_right)$.

4. $\Box(1, center) \wedge \Diamond(1, \neg push\_left) \wedge \Diamond(1, \neg push\_right) \supset \Box(2, center)$.

5. $\Box(1, center) \wedge \Box(1, push\_left) \wedge \Diamond(1, \neg push\_right) \supset \Box(2, left)$.

6. $\Box(1, center) \wedge \Box(1, push\_right) \wedge \Diamond(1, \neg push\_left) \supset \Box(2, right)$.

As a result, Shoham's algorithm outputs $\top(1, center), \top(1, push\_left), \top(1, push\_right)$. No other abs is known in the cmi models of this causal theory. But this is strange! Since *push_left* and *push_right* are executed simultaneously, $\top(2, center)$ must also be obtained.

This problem can be resolved by introducing additional axioms such as "if it is known that the block is at *center*, and that *push_right* and *push_left* are simultaneously performed, then it is known that the block remains at *center*" and "if it is known that the block is at *center*, and that no *push_right* or *push_left* are performed, then it is known that the block remains at *center*." Clearly, in more complex domains, the number of such axioms can grow quickly [10]. There must be a way of resolving this problem with a persistence axiom.

Definition 3.1 The *set of counteractions* is the set of actions that prevent each other from being operative when performed simultaneously.

Note that the set of counteractions is not unique; there may be many of them. A set of counteractions contain actions which are 'counteractive' with respect to the features of the domain under consideration.

Definition 3.2 Let $\Pi(t_1) = \{\Diamond(t_1, \neg p_i) : 1 \le i \le n\}$ where each p_i is a counteraction with respect to other p_j's, $j \ne i$. Letting M be the unique cmi model of a causal theory, we write $M \models \Pi$ iff $M \models \Diamond(t_1, \neg p_i)$ for all $\Diamond(t_1, \neg p_i) \in \Pi$, or $M \models \Box(t_1, p_i)$ for all $\Diamond(t_1, \neg p_i) \in \Pi$. Otherwise, we write $M \not\models \Pi$. (Due to simultaneity, the time parameter will usually be dropped. Thus, we are allowed to write Π instead of $\Pi(t_1)$.)

To return to our last example, axiom 4 is replaced with the axiom below, where $\Pi = \{\Diamond(1, \neg push\_left), \Diamond(1, \neg push\_right)\}$ (relaxing the notation slightly, Π will be used as if it were a function over its members):

4$'$. $\Box(1, center) \wedge \Pi(\Diamond(1, \neg push\_left), \Diamond(1, \neg push\_right)) \supset \Box(2, center)$.

This axiom says that "if it is known that the block is located at *center* at 1, and either it is known that the block is simultaneously pushed left and right at 1 or it is not known that the block is simultaneously pushed left and right at 1, then it is known that the block is located at *center* at 2."

We replace axiom 4 by axiom 4$'$ in the causal theory above. By axioms 2 and 3, *push_left* and *push_right* are known to be performed at 1. Therefore, Π is satisfied. By axiom 1 it is known that the block is located at *center* at 1. Hence, by axiom 4$'$ it is known that the block stays at *center* at 2. Axioms 5 and 6 fail. In all cmi models of the causal theory $\top(1, center), \top(1, push\_left), \top(1, push\_right), \top(2, center)$ are known.

There may be many sets of counteractions, depending on the world features. Assume that in our blocks world, we have not only *push_left* and *push_right*, but also *push_forward* and *push_backward* which can be applied with the same magnitude of force as the others. Allow three sets of counteractions: $\Pi_1 = \{\Diamond(1, \neg push\_left), \Diamond(1, \neg push\_right)\}$, $\Pi_2 = \{\Diamond(1, \neg push\_forward, \Diamond(1, \neg push\_backward)\}$, and $\Pi_3 = \Pi_1 \cup \Pi_2$. In this case, axiom 4$'$ can be rewritten as:

4$''$. $\Box(1, center) \wedge \Pi_1 \wedge \Pi_2 \supset \Box(2, center)$.

(Caveat: Π_3 is omitted since Π_1 and Π_2 together provide the same interpretation as the case where Π_3 is added.)

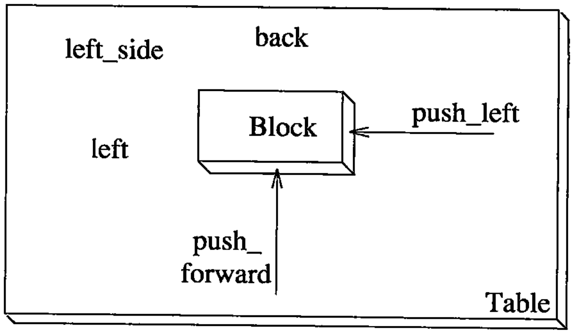

Figure 2. Top view of the blocks world.

Actions in a set of counteractions mutually delete their effects when performed simultaneously. This, however, is a special case. In fact, there are cases in which actions belonging to different sets of counteractions, when executed simultaneously, modify their effects or mutually destroy some of their effects, but not all of them. Consider a scenario in which only $push\_left$ and $push\_forward$ are performed simultaneously. Again assuming that the block initially rests at $center$, if these operations are applied individually, then the block would move left or forward, respectively. However, simultaneously executing them causes the block to move to a place, say $left\_side$, between $left$ and $back$ (Figure 2). Then, the axiom to reason about the effects of these actions will be:

$$\Box(1, center) \wedge \Box(1, push\_left) \wedge \Box(1, push\_forward) \wedge \Diamond(1, \neg push\_right) \wedge \Diamond(1, \neg push\_backward) \supset \Box(2, left\_side).$$

We now embed the notion of counteractions in causal theories. The result can be regarded as a broader class of causal theories although the extended theories can solve exactly the same class of problems as Shoham's causal theories.

Definition 3.3 An *extended causal theory* ξ is a theory in CI containing sentences of the form $\Phi \wedge \Theta \wedge \Omega \supset \Box\varphi$ where

1. $\varphi = [\neg]\top(t_1, t_2, p)$.
2. $\Phi = \bigwedge_{i=1}^{n} \Box\varphi_i$, where φ_i is an abs with ltp $t_i < t_1$.
3. $\Theta = \bigwedge_{j=1}^{m} \Diamond\varphi_j$, where φ_j is an abs with ltp $t_j < t_1$.
4. $\Omega = \bigwedge_{k=1}^{r} \Pi_k$, where Π_k is a set of counteractions with ltp $t_k < t_1$.
5. Φ or Θ or Ω (or all) may be empty. A sentence in which Φ and Ω are empty is a *boundary condition.* Other sentences are *causal rules.*
6. There is a t_0 (global for ξ) such that if $\Theta \supset \Box(t_1, t_2, [\neg]p)$ is a boundary condition, then $t_0 < t_1$.
7. It is not permitted to have two sentences in ξ such that one contains $\Diamond(t_1, t_2, p)$ in its antecedent part and the other contains $\Diamond(t_1, t_2, \neg p)$ in its antecedent part.
8. If $\Phi_1 \wedge \Theta_1 \wedge \Omega_1 \supset \Box(t_1, t_2, p)$ and $\Phi_2 \wedge \Theta_2 \wedge \Omega_2 \supset \Box(t_1, t_2, \neg p)$ are two sentences in ξ, then $\Phi_1 \wedge \Theta_1 \wedge \Omega_1 \wedge \Phi_2 \wedge \Theta_2 \wedge \Omega_2$ is inconsistent.

Extended causal theories are always consistent and have cmi models. In all cmi models of a given extended causal theory the same base sentences are known. Furthermore, these known base sentences can be computed efficiently. The following two theorems make these points more precise.

Theorem 3.1 If ξ is an extended causal theory, then ξ has cmi models and in all these cmi models the same abs's are known.

Proof. Appendix.

In order to motivate Theorem 3.1, we should show that extended causal theories are not just special cases of causal theories. Consider the following argument, advanced by one of the referees of this paper. Since $M \models \Pi$ (where $\Pi = \{\Diamond\varphi_1, \ldots, \Diamond\varphi_n\}$ is a set of counteractions) is equivalent to $M \models (\bigwedge_{i=1}^{n} \Diamond\varphi_i) \vee (\bigwedge_{i=1}^{n} \Box\neg\varphi_i)$, an 'extended' rule $\Phi \wedge \Theta \wedge \Pi \supset \Box\varphi$ is equivalent to two 'non-extended' rules:

1. $\Phi \wedge \Theta \wedge (\bigwedge_{i=1}^{n} \Diamond\varphi_i) \supset \Box\varphi$.
2. $\Phi \wedge \Theta \wedge (\bigwedge_{i=1}^{n} \Box\neg\varphi_i) \supset \Box\varphi$.

This seems to substantiate the claim that extended causal theories are but syntactic sugar! While this argument makes sense, it does not tell the whole story. There can be many other actions that one may wish to include, but then the non-extended axioms would become long and clumsy. Our approach is more elegant and natural because we can represent the information about counteractions as a separate axiom. Thus adding information about further, previously unknown counteractions would amount to extending the counteractions set, rather than modifying the non-extended axioms. In other words, while there is a syntactic equivalence between Shoham's framework and ours, it is the semantic notion of counteractions that is novel with our theory.

Furthermore, in widening simple causal theories to extended causal theories, there is no loss of efficiency:

Theorem 3.2 If ξ is an extended causal theory of size n, then the unique set of abs's known in any cmi model of ξ can be computed in time $O(n\log n)$.

Proof. Appendix.

The basic argument behind Theorem 3.2 is simple: We can always build a cmi model of an extended causal theory by going forward in time and adding the knowledge implied by what is already known and what is not known about the past.

4. Simultaneity of Cause and Effect

4.1. Circular situations

In causal theories, the $\Box$-conditions on the antecedent part of a causal rule denote causes while the consequent part denotes their effect. Under this interpretation, having simultaneous temporal propositions on both sides of causal sentences results in *circular causation* [12, p. 179]: $\Box(t, p_i) \supset \Box(t, p_{i+1})$, $i = 1, \ldots, n-1$, where it is assumed that $p_n = p_1$. Furthermore, sentences of the form $\Box(t_1, p) \supset \Box(t_2, p)$, where $t_1 < t_2$, are allowed in causal theories. Does this mean that p causes itself? There can be sentences of the form $\Diamond(t_1, p) \supset \Box(t_2, p)$, where $t_1 < t_2$. Is this rendered as "if $\neg p$ is not known at t_1, then p is known at t_2 for no reason"? Through soundness conditions, one can write sentences like $\Diamond(t_1, p) \supset \top(t_1, p)$. Shoham [12, p. 118] assumes that "the soundness conditions are implicitly part of the causal theory itself, and are omitted simply for reasons of economy of expression." Moreover, the boundary between $\Box$- and $\Diamond$-conditions in Shoham's account becomes obscure if $\Box$-conditions in a causal rule strictly denote the causes.

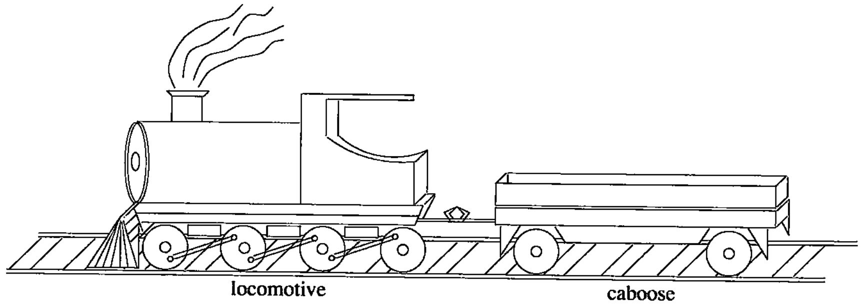

Figure 3. Cause-effect distinction.

As an example, consider Figure 3. Assume that the causal theory contains the following:

1. $\Box(4, \mathit{locomotive\_moves}) \supset \Box(4, \mathit{caboose\_moves})$.
2. $\Box(4, \mathit{caboose\_moves}) \supset \Box(4, \mathit{locomotive\_moves})$.

Looking only at the syntactic forms of these rules, one can say that they permit circular causation. But now add $\top(4, \mathit{locomotive\_moves})$ to the causal theory. Then, $\top(4, \mathit{locomotive\_moves})$ and $\top(4, \mathit{caboose\_moves})$ will be the only sentences known in all cmi models of the causal theory. In this case, if one investigates the cause of the motion of the locomotive and the caboose, one may identify the motion of the locomotive as the cause of the motion of the caboose although $\Box(4, \mathit{caboose\_moves}) \supset \Box(4, \mathit{caboose\_moves})$ implies self-causation.

There are additional problems [12, p. 179]:

> "One might have a set of sentences $\Box(t, p_i) \supset \Box(t, p_{i+1})$, $i = 1, \ldots, n-1$, $p_n = \neg p_1$. This would destroy the independence of the past from the future in general, and the 'unique'-model property in particular. Or, as another example, one might have sentences of the form $\Box(t, p) \supset \Box(t, \neg p)$, which would have a similarly detrimental effect."

By placing some restrictions on the sentences in the definition of causal theories, these problems can be eliminated. In the former case, it is possible to impose a restriction on the sentences, similar to one in the definition of the original causal theories. Recall that consistency of the causal theories is maintained by condition 5, Definition 2.14.

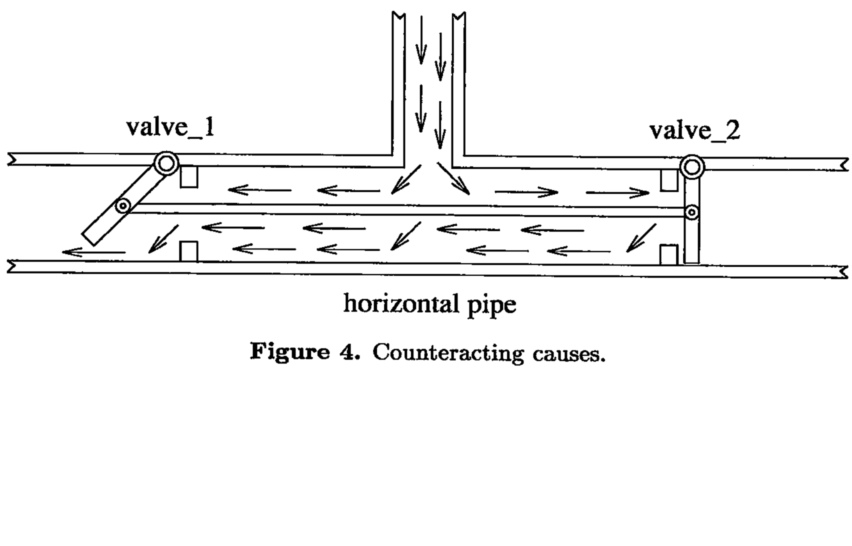

Figure 4. Counteracting causes.

4.2. Counteracting causes

Consider the following scenario adapted from [14, p. 108]. There is a horizontally positioned pipe whose two ends are controlled with two valves (Figure 4). These are connected in such a way that if one is opened at one time, the other is closed at the same time, and vice versa. The opening of one valve and closing of the other occur simultaneously. The pipe also has a top-inlet continuously supplying high pressure water into it. Hence, the pipe, with the valves directing water flow in one direction at a time, functions as a two-way watering system. Now let the state of the first valve's being open be represented by p while that of the second by q. Then, at any time either $p \wedge \neg q$ or $\neg p \wedge q$. Additionally, let the flow of the water through the first (respectively the second) valve be represented by r (respectively s). Then, a causal theory will contain:

1. $\Box(t,p) \wedge \Theta_1 \supset \Box(t+1,r)$, for all t, where $\Theta_1 = \bigwedge_{i=1}^{m} \Diamond(t_{i_1}, t_{i_2}, a_i)$.
2. $\Box(t,q) \wedge \Theta_2 \supset \Box(t+1,s)$, for all t, where $\Theta_2 = \bigwedge_{j=1}^{n} \Diamond(t_{j_1}, t_{j_2}, b_j)$.

Assume that the states of the valves are causally related to a common cause (e.g., if there is a possibility for an agent which pushes only the first valve and closes it, this action causes the first valve to close and the second valve to open). In this case, the causal theory above might contain the following sentences where the pushing of the first valve is represented by u:

1. $\Box(t,u) \wedge \Theta_1 \supset \Box(t+1,\neg p)$, for all t.
2. $\Box(t,u) \wedge \Theta_2 \supset \Box(t+1,q)$, for all t.

If $\Theta_1 = \Theta_2$, one cause produces more than one effect. This suggests that causal rules can represent multiple effects.

If the two changes have separate causes, the situation is easy. For example, let the first valve be open and the second closed. If there is an agent pushing the first valve to close it, there may be another agent pulling the second valve to open it. Then, closing of the first valve can be attributed to the pushing of it, and opening of the second valve to the pulling of it. Or, it may well be the case that one agent pushes the first valve while another pushes the second valve. In this case, there are two causes, namely pushing of the first and the second valves, that intervene with each other. Although each cause separately has the efficacy to produce its effect(s), they now prevent the changes that they will bring about. Since each one prevents the other from being operative, these two can be termed *counteracting causes*, following von Wright [16, pp. 75–77]. These two causes must be involved in the related causal sentences either in the form $\Box(t,u) \wedge \Diamond(t,\neg v)$ or in the form $\Box(t,v) \wedge \Diamond(t,\neg u)$ where u and v denote pushing of the first and the second valves, respectively.

5. What Else is Needed?

Assume that we fire a loaded gun. Absence of the air in the environment prevents hearing a loud noise after the firing. Moreover, a loud noise would not be inferred if the gun had no firing pin, or if the gun was immersed in the water, or if the bullets were made from marshmallow, or if the gun powder was moist, and so on. But how does one deal with the totality of these conditions, with the so-called *qualification problem* [7]?

What makes something a real bullet is its having a part with the property of being explosive. Then, a marshmallow bullet or a bullet lacking explosives would not be a real bullet. For the gun powder, being immersed in water or orange juice is not the crux of the problem. What matters is that it becomes moist and that being moist directly affects the explosiveness property. Therefore, the condition $\Diamond(t, \neg gun\_powder\_moist)$, when added to axiom 4 (Section 3), should suffice to handle the effects of all liquids that make the gun powder moist.

Besides, firing the gun is not necessary for the explosion of the bullet. If the bullet is in direct contact with something hot, it may again explode and produce noise. Then, all primitive propositions sharing the same property (e.g., being hot) can cause the bullet to explode. However, propositions are weak for representing relations in this respect. Using first-order logic seems to be a better approach in this direction [2, 11].

The qualification problem is in fact twofold [5]. First, there is the problem of fully specifying the characteristics of the qualifications in a realistically complex environment. For example, if one would like to predict noise when a loaded gun is fired, it is more appropriate to talk about the existence of a *general* environment carrying sound rather than a *specific* environment, namely air. This calls for a categorization and close analysis of the properties of the related entities in a particular context of reasoning.

The second aspect of the qualification problem is computational. It has been thought that qualifications may become infinite, showing the impossibility of an effective computational model for intelligent action. Naur, for one, sees the formalism introduced by Shoham as a non-solution to the qualification problem. His objection, excerpted below, centers around the nature of $\Diamond$-conditions [8]:

> "Roughly these lines [he is referring to axiom 4, Section 3] state that if at time t the gun is loaded, fired, surrounded by air, provided with proper firing pin, and loaded with bullets not made of marshmallow, and in addition other mundane conditions are also satisfied, then at time $t+1$ a noise will be heard. That this is a non-solution, however, is made visible most prominently by the appearance of 'other mundane conditions' clause. This clearly will have to take care of the rest of the world. But the world cannot be captured in terms of predicates."

Naur's objection seems to do with the idea of *context* in AI [1]. One can immediately ask whether all the things in the rest of the world are relevant to the process of inferring noise. No matter how large this set of qualifications, the context of reasoning limits this set, leaving the rest as *background assumptions*. One simply cannot talk about a context capturing the whole world, as McCarthy observes [6, p. 1034]: "Whenever we write an axiom, a critic can say it is true only in a certain context. With a little ingenuity, the critic can usually devise a more general context in which the precise form of the axiom does not hold. [...] We encounter Socratic puzzles over what the concepts mean in complete generality and encounter examples that never arise in life. There simply is not a most general context."

Thus, another source of difficulty arises: Finding out the factors relevant to the phenomenon under consideration. If, for example, causation is at stake, a bear jumping in the North Pole cannot have causal connections with hearing a loud noise when a gun is fired in the South Pole. This approach, being that of *causal determinism*, requires some restrictions on the range of possibly relevant factors to the phenomenon. In particular, the relevant factors must be limited at least to the spatio-temporally neighboring ones. These assumptions allow one to deal with less number of factors in that particular context since there might be still causally related factors which cannot be enumerated a priori. This, however, requires a well-defined notion of context to be embedded in the causal theories [1, 6].

Finally, it is unreasonable to explicitly determine all consequences of an action. For example, if we push a block on a table to the left, all other things on the block, if any, move along with the block. Some of them may fall down onto the table, or even to the floor. Certain parts of the table are uncovered and covered as a result of this action. In fact, for any given action there are essentially an infinite number of possible consequences that might arise depending on the context in which the action occurs. Of course, determining the consequences of a number of actions occurring simultaneously is harder than that of a single action. All these remain as fruitful problems for future research.

6. Conclusion

We have shown that axiomatization of the effects of actions in causal theories needs to be extended to obtain a concise theory in case of simultaneous actions. We have found out that causal theories call for a syntactic entity to be used in these cases. Counteractions, which enable one to write concise descriptions for simultaneous actions and their effects, have been embedded in causal theories. We have shown that this destroys neither the unique-model property of causal theories nor their computational efficiency.

We have also indicated various problems with temporal propositions and investigated the use of causal theories in treating simultaneous cause and effect. It has turned out that causal theories must be elaborated further to eliminate, or at least soften, their limitations vis-à-vis these problems.

7. Acknowledgments

We are grateful to two anonymous referees of *Elektrik* for comments and criticism which improved the paper considerably. The second author wishes to thank the Scientific and Technical Research Council of Turkey (TÜBİTAK) for partial support (Grant No. TBAG–992).

We offer this paper as a tribute to Teo Grünberg (Middle East Technical University, Ankara).

References

[1] V. Akman and E. Tın, "What is in a context?" in *New Trends in Communication, Control, and Signal Processing*, ed. E. Arıkan, Volume 2, Elsevier, Amsterdam, 1990, pp. 1670–1676.

[2] J. Bell, "Extended causal theories," *Artificial Intelligence* **48** (1991), pp. 221–224.

[3] S. Hanks and D. V. McDermott, "Default reasoning, nonmonotonic logics, and the frame problem," *Proceedings AAAI-86*, Philadelphia, PA, 1986, pp. 328–333.

[4] S. Hanks and D. V. McDermott, "Nonmonotonic logic and temporal projection," *Artificial Intelligence* **33** (1987), pp. 379–333.

[5] J. McCarthy, "Circumscription—A form of nonmonotonic reasoning," *Artificial Intelligence* **13** (1980), pp. 27–39. Addendum: *Artificial Intelligence* **13** (1980), pp. 171–172.

[6] J. McCarthy, "Generality in artificial intelligence," *Communications of the ACM* **30** (1987), pp. 1031–1035.

[7] J. McCarthy and P. J. Hayes, "Some philosophical problems from the standpoint of artificial intelligence," in *Machine Intelligence*, eds. B. Meltzer and D. Michie, Volume 4, Edinburgh University Press, 1969, pp. 463–502.

[8] P. Naur, Review No. 8901–0013, *ACM Computing Reviews* **30** (1989).

[9] R. Reiter, "A logic for default reasoning," *Artificial Intelligence* **13** (1980), pp. 81–132.

[10] Y. Shoham, "Ten requirements for a theory of change," *New Generation Computing* **3** (1985), pp. 467–477.

[11] Y. Shoham, "Temporal logics in AI: Semantical and ontological considerations," *Artificial Intelligence* **33** (1987), pp. 89–104.

[12] Y. Shoham, *Reasoning About Change: Time and Causation from the Standpoint of Artificial Intelligence*, MIT Press, Cambridge, MA, 1988.

[13] Y. Shoham, "Chronological ignorance: Experiments in nonmonotonic temporal reasoning," *Artificial Intelligence* **36** (1988), pp. 279–331.

[14] E. Sosa, ed., *Causation and Conditionals*, Oxford University Press, 1975.

[15] E. Tın, *Computing with Causal Theories*, M.S. Thesis, Department of Computer Engineering and Information Science, Bilkent University, Ankara, 1990.

[16] G. H. von Wright, *Causality and Determinism*, Columbia University Press, 1974.

Appendix: Proofs

Theorem 3.1

In Section 3., it has been shown that for any 'extended' rule (with one set of counteractions) of an extended causal theory ξ, one could write equivalent 'non-extended' rules in Shoham's causal theories. We now generalize this transformation.

Let $\Phi \wedge \Theta \wedge \Omega \supset \Box\varphi$ be a rule in an extended causal theory ξ where $\Omega = \bigwedge_{k=1}^{r} \Pi_k$. Also let $\Pi_i = \{\Diamond\varphi_{i_1}, \ldots, \Diamond\varphi_{i_{n_i}}\}$ for $1 \leq i \leq r$. One should write 'non-extended' rules with combinations of sets of $\Diamond$- and $\Box$-sentences in all sets of counteractions of Ω. This means that for each 'extended' rule in ξ with r sets of counteractions, 2^r 'non-extended' (vanilla-flavor) rules should be written:

1. $\Phi \wedge \Theta \wedge (\bigwedge_{j=1}^{n_1} \Diamond\varphi_{1_j}) \wedge (\bigwedge_{j=1}^{n_2} \Diamond\varphi_{2_j}) \wedge \ldots \wedge (\bigwedge_{j=1}^{n_r} \Diamond\varphi_{r_j}) \supset \Box\varphi$.
2. $\Phi \wedge \Theta \wedge (\bigwedge_{j=1}^{n_1} \Box\neg\varphi_{1_j}) \wedge (\bigwedge_{j=1}^{n_2} \Diamond\varphi_{2_j}) \wedge \ldots \wedge (\bigwedge_{j=1}^{n_r} \Diamond\varphi_{r_j}) \supset \Box\varphi$.

... ...

2^r. $\Phi \wedge \Theta \wedge (\bigwedge_{j=1}^{n_1} \Box\neg\varphi_{1_j}) \wedge (\bigwedge_{j=1}^{n_2} \Box\neg\varphi_{2_j}) \wedge \ldots \wedge (\bigwedge_{j=1}^{n_r} \Box\neg\varphi_{r_j}) \supset \Box\varphi$.

Application of this transformation to all 'extended' rules in an extended causal theory results in a causal theory in Shoham's sense. ¿From this point on, the proof smoothly follows from Shoham's fundamental result (Theorem 2.1).

Theorem 3.2

The algorithm below organizes the sentences of ξ in ascending order of their ltp's. Then, the sentences are examined to see if their antecedents are satisfied. If so, their consequents are 'marked' accordingly. Two symbols, $\oplus$ and $\ominus$, are used for this purpose.

Algorithm:

1. Let L be the list of all sentences in ξ.

2. Let C be a list. Gather all abs's appearing in L into C by dropping negation signs.

3. Sort L in ascending order by the ltp of the antecedent parts of the sentences in it. Also sort C in ascending order by the ltp of the base sentences.

4. Remove duplicates of any abs in C. Let all members be unmarked.

5. If L is empty, then halt. The abs's known in the cmi models of ξ are the atomic sentences marked $\oplus$ in C, and the negations of the ones marked $\ominus$ in C.

6. Remove the first sentence of L and let this be $\Phi \wedge \Theta \wedge \Omega \supset \Box(t_1, t_2, [\neg]p_i)$. For each conjunct $\Diamond(t_{i_1}, t_{i_2}, [\neg]p_i)$ of Θ, determine how $\top(t_{i_1}, t_{i_2}, p_i)$ is marked in C by performing a binary search on C. If one of the following conditions is true:

   a. $\Box(t_{i_1}, t_{i_2}, p_i)$ is a conjunct of Φ and $\top(t_{i_1}, t_{i_2}, p_i)$ is not marked $\oplus$,

   b. $\Box(t_{i_1}, t_{i_2}, \neg p_i)$ is a conjunct of Φ and $\top(t_{i_1}, t_{i_2}, p_i)$ is not marked $\ominus$,

   c. $\Diamond(t_{i_1}, t_{i_2}, p_i)$ is a conjunct of Θ and $\top(t_{i_1}, t_{i_2}, p_i)$ is marked $\ominus$,

   d. $\Diamond(t_{i_1}, t_{i_2}, \neg p_i)$ is a conjunct of Θ and $\top(t_{i_1}, t_{i_2}, p_i)$ is marked $\oplus$,

   then go to 5;

   else for each conjunct Π_j of Ω, check for each $\Diamond(t_{ji_1}, t_{ji_2}, \neg p_{ji}) \in \Pi_j$ how $\top(t_{ji_1}, t_{ji_2}, p_{ji})$ $(1 \leq i \leq m)$ is marked in C by performing a binary search on C. If one of the following conditions is true for any Π_j:

   $\Diamond(t_{jk_1}, t_{jk_2}, \neg p_{jk}) \in \Pi_j$ and $\Diamond(t_{jl_1}, t_{jl_2}, \neg p_{jl}) \in \Pi_j$ where

   a. $\top(t_{jk_1}, t_{jk_2}, p_{jk})$ is marked $\oplus$ while $\top(t_{jl_1}, t_{jl_2}, p_{jl})$ is unmarked,

   b. $\top(t_{jk_1}, t_{jk_2}, p_{jk})$ is unmarked while $\top(t_{jl_1}, t_{jl_2}, p_{jl})$ is marked $\oplus$,

   c. $\top(t_{jk_1}, t_{jk_2}, p_{jk})$ is marked $\ominus$ while $\top(t_{jl_1}, t_{jl_2}, p_{jl})$ is unmarked,

   d. $\top(t_{jk_1}, t_{jk_2}, p_{jk})$ is unmarked while $\top(t_{jl_1}, t_{jl_2}, p_{jl})$ is marked $\ominus$,

   e. $\top(t_{jk_1}, t_{jk_2}, p_{jk})$ is marked $\ominus$ while $\top(t_{jl_1}, t_{jl_2}, p_{jl})$ is marked $\oplus$,

   f. $\top(t_{jk_1}, t_{jk_2}, p_{jk})$ is marked $\oplus$ while $\top(t_{jl_1}, t_{jl_2}, p_{jl})$ is marked $\ominus$,

   then go to 5;

   else mark $\top(t_1, t_2, p)$ $\oplus$ if the consequent of the sentence removed from L is $\Box(t_1, t_2, p)$, and $\ominus$ if it is $\Box(t_1, t_2, \neg p)$. Go to 5.

Complexity:

- Step 1: $O(1)$ (initialization).
- Step 2: $O(n)$ (collection).
- Step 3: $O(n\log n)$ (sorting).
- Step 4: $O(n)$ (removing duplicates and marking).
- Step 5: $O(1)$ (checking if L is empty).

- Step 6: $O(n)$. Labels of all the abs's in a sentence removed from L are checked once. This can be done at most n times. Determination of the label of each abs requires a binary search. Hence, label checking is completed in $O(n\log n)$. A new labeling can be done in time $O(\log n)$ since it also requires a binary search. There can be at most n new labeling operations during the execution of the algorithm, to be completed in time $O(n\log n)$.

Hence, the total time complexity of the algorithm is $O(n\log n)$.

Karşı-hareketlerin Mantığı

Erkan TIN ve Varol AKMAN

Bilgisayar Mühendisliği ve Enformatik Bilimleri Bölümü
Bilkent Üniversitesi, Bilkent, Ankara 06533, Türkiye

Özet

Bu makalede nedensel teoriler geliştirilmekte ve birden fazla robotun bulunduğu alanlarda hareket konusu üzerinde çalışılmaktadır. Yoav Shoham tarafından ortaya konan nedensel teoriler, temporel ve tekdüze olmayan bir mantık üzerine kurulmuş olup hesaplanması basit özelliklere sahiptirler. Shoham'ın formelizmi, aynı anda yapılan hareketlerin gösterimi ve bunların sonuçlarının belirtilmesi için uygun bir mekanizma sağlamamaktadır. Bundan dolayı, bu makalede nedensel teorilerin verimliliklerini ve model teorik özelliklerini bozmaksızın bir karşı-hareketler mantığı ortaya konulmaktadır.