知識工学と人工知能 40 − 4
(1 9 8 5. 5. 7)

ON A MECHANICAL REASONING ABOUT CAUSAL RELATIONS

Akira FUSAOKA Kazuko TAKAHASHI
Central Research Laboratory, Mitsubishi Electric Corporation

This paper presents a method for representation and reasoning about causal relations within the framework of modal logic. Two types of interpretations are given to the causal relation " A causes B", progressive causality and regressive causality. The former means " If A has happened, then B will possibly happen as an effect of A," and the latter represents " If B has happened, then A might have happened as a cause of B." We define the models by possible world semantics. For a reasoning method, we have presented a tableaux procedure. It is a procedure to construct a causal tableau which is essentially stands for the model. Several types of reasoning about causality such as causal verification and simulation can be treated by this method in a unified way. The applications to hypothetical reasoning is discussed to illustrate how the causal tableau works.

I. INTRODUCTION

This paper presents a method for representation and reasoning about causal relation, namely, the relation between cause and effect, within the framework of modal logic. The two types of interpretation for the causal relation are defined on the basis of possible world semantics, and the simple reasoning method by using tableaux is presented. Also we discuss about the application of this method to hypothetical reasoning.

Causality, which has been considered to be a problematic notion in the philosophy because of its vagueness, is a very fundamental concept in natural discourse or commonsense reasoning. In fact, a usual human action is itself not logically but causally determined. Compared with the logical (nominal) implication, the causal relation gives so many reasoning methods in various situations of everyday life.

For instance, assume that we know a causal relation

(P1) Eating an ounce of arsenics causes a sudden death.

From (P1), we may suspect arsenics as a candidate for the cause of someone's sudden death. This type of reasoning, the causal analysis or retrodiction, is very often used in expert systems. Or we may imagine that if he ate an ounce of arsenics, he should die suddenly. This hypothetical reasoning is also useful for the thought-experimentation and contingency planning. By virtue of this conceptual importance of causality, many AI systems involve causal reasoning. For example, CASNET[10] utilizes causal relation in medical diagnosis. A qualitative simulation can also be performed on the basis of causal relation if the dynamics of a system is given in the causal form. This paper intends to present a logical system in which such various types of reasoning about causality can easily be realized in a unified way.

Although there have been a lot of works on causal relations, their formal treatments within modal logic return back to Burks.[1] Burks introduces the new modality of the causal necessity to the conventional system of modal logic. However, as is discussed in Simon[9], he defined the causal relation as the (strict) logical implication, so that "X causes Y" should entail its contraposition "not Y causes not X". But this isn't necessarily fit for the ordinary usage of the word " causality ".

For example,

(P2) "If it rains, then the wheat grows."

is logically compatible with

(P3) "If the wheat does not grow, then it doesn't rain ."

But

(P4) "The rain causes the wheat to grow "

is not compatible with

(P5) "Not growing of the wheat causes it not to rain. "

This difference requires us to introduce a causal relation with possible world semantics. In this paper, we attempt to define a causal relation by making a model with a tree-like structure of the possible worlds. The asymmentry between cause and effect is not reflected to the asymmetry of the logical implication but to the asymmetry of the relation which connects the possible worlds.

In the followings, the model structure and basic properties of causal relations are described in section 2. In section 3, tableaux procedure is introduced as a mechanical reasoning method. And in section 4, it is applied to belief contravening hypothesis.

II. CAUSAL RELATION IN POSSIBLE WORLD

2-1. Causal Relation

When we say "A causes B", the following two interpretations may be considered.

(i) If A has happened, then B will possibly happen as an effect of A.

(ii) If B has happened, then A might have happened as a cause of B.

In case of (i), A is considered to be a sufficient condition in the meaning that the occurrence of the cause A is a sufficient premise for the occurrence of B to be possible. We call this relation *the progressive causality*. On the other hand, (ii) is corresponding to necessary condition because it means that the non-occurence of the cause B is sufficient for the non-occurrence of A. We call this relation *the regressive causality*.

These two types of causal relation are very often confused with each other because of the symmetry of their logical properties. However, their formalizations should be treated independently, that is, each interpretation and the reasoning method should be given differently and used in the different situation. For instance, the regressive causality is useful to find out the cause from the result, while the progressive causality can be adopted to qualitative simulation by which the behavior of the system is predicted.

We introduce two modal symbols $\xrightarrow{p}$ and $\xrightarrow{r}$ which mean

(i) $A \xrightarrow{p} B$ (progressive causality)
if the cause A has happened, then the effect B will happen

(ii) $A \xrightarrow{r} B$ (regressive causality)
if the effect B has happened, then the cause A happened

2-2. Model Structure

We will introduce two types of causal model, *progressive causal model* and *regressive causal model* , based on the possible world semantics formulated by Kripke[6].

The progressive causal model (G,K,R) is defined as follows. K is a set of possible worlds where a world is a truth value assignment to the atomic formulas . $G \in K$ corresponds to the real world and R is a reflexive and transitive binary relation on K. Hence, our theory is equivalent to $S4$ in modal logic. For $w, w_i \in K$, w_i is said to be *a possible world* of w if wRw_i holds. A formula F is assigned true by the world w (denoted by $w \models F$) in the following way.

(i) $w \models p$ where p is an atomic formula
iff p is true under the truth value assignment for the world w

(ii) $w \models A \wedge B$ *iff* $w \models A$ and $w \models B$

(iii) $w \models \neg A$ *iff* not $w \models A$

(iv) $w \models A \xrightarrow{p} B$
iff for every w_i s.t. wRw_i, (if $w_i \models A$ then there exists $j (i \neq j)$ s.t. w_iRw_j, $w_j \models B$)

Moreover, the standard modal operators are defined as usual.

(v) $w \models \Box A$ *iff* for every w_i s.t. wRw_i, $w_i \models A$

(vi) $w \models \Diamond A$ *iff* there exists w_i s.t. $wRw_i, w_i \models A$

2-3. Fundamental Properties of Progressive Causality

From the definition of progressive causality, the following basic properties hold.

(1) $A \xrightarrow{p} B \supset \Box(A \xrightarrow{p} B)$

(2) $A \xrightarrow{p} B \wedge B \xrightarrow{p} C \supset A \xrightarrow{p} C$

(3) $A \xrightarrow{p} C \wedge B \xrightarrow{p} C \equiv A \vee B \xrightarrow{p} C$

(4a) $\Box(A \supset B) \wedge B \xrightarrow{p} C \supset A \xrightarrow{p} C$

(4b) $A \xrightarrow{p} B \wedge \Box(B \supset C) \supset A \xrightarrow{p} C$

(5) $A \xrightarrow{p} false \equiv \Box\neg A$

(6) $A \xrightarrow{p} B \supset (\Box\neg B \supset \Box\neg A)$

(1) denotes the causal uniformity principle which requires the causal relation to be invariant thoughout the worlds, since causality is a universal law.

(2) denotes a causal chaining (i.e.transitivity of $\xrightarrow{p}$). It requires the transitivity of R, as is shown below.

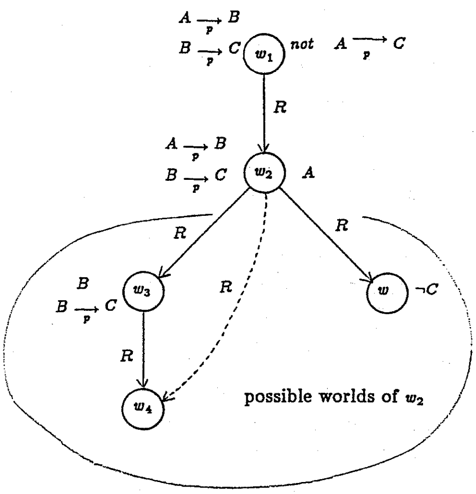

Fig.1 Possible world construction

Assume that there exists a model which falsifies (2). Then there exists a world w_1 such that $w_1 \models A \xrightarrow{p} B \wedge B \xrightarrow{p} C$ and not $w_1 \models A \xrightarrow{p} C$. The latter indicates that there exists a world w_2 such that w_1Rw_2 in which both of the following conditions are satisfied :

(i) $w_2 \models A$

(ii) for any w, such that w_2Rw, not $w \models C$

On the other hand, $w_1 \models A \xrightarrow{p} B \wedge B \xrightarrow{p} C$ means $w_1 \models A \xrightarrow{p} B$ and $w_1 \models B \xrightarrow{p} C$. From $w_1 \models A \xrightarrow{p} B$ and $w_2 \models A$, there exist a world w_3 such that w_2Rw_3 and $w_3 \models B$. Since $w_3 \models B \xrightarrow{p} C$ also holds, there exists a world w_4 such that $w_4 \models C$. Besides, w_2Rw_4 holds because of the transitivity of R. It is a contradiction. Hence, there is no model which falsify (2). (Fig.1)

(3) is a characteristic property of a sufficient condition. $A \xrightarrow{p} C \wedge B \xrightarrow{p} C$ shows that both A and B cause C independently. Let the worlds be w_1, w_2 in which A, B is true, respectively. Although w_1 and w_2 are not necessarily the same world, $A \vee B$ is true in both of w_1 and w_2. Therefore the formula is equivalent to $A \vee B \xrightarrow{p} C$.

(4a) and (4b) indicate how causal relation is related to (strict) logical implication.

(5) indicates a basic requirement for causal absurdity. The if part represents that nothing happens that causes logical absurdity. The only-if part requires the reflexivity of R, since the model satisfying $\neg A$ in all the worlds except the real world falsifies it if reflexivity is lacking.

(6) shows the requirement on the non-occurrences of cause and effect.

2-4. Regressive Causality

Regressive causal model is defined similarly, since the logical properties which hold in this system are symmetrically treated. In this case we define a model (G, K, R') where R' is a reflexive and transitive binary relation on K. Truth value assignment to $A \xrightarrow{r} B$ is given similarly, that is,

$w \models A \xrightarrow{r} B$

iff for every w_i s.t. $wR'w_i$, (if $w_i \models B$ then there exists $j(i \neq j)$ s.t. $w_iR'w_j$, $w_j \models A$)

Regressive causality satisfies the following properies :

(1) $A \xrightarrow{r} B \supset \Box(A \xrightarrow{r} B)$

(2) $A \xrightarrow{r} B \wedge B \xrightarrow{r} C \supset A \xrightarrow{r} C$

(3) $A \xrightarrow{r} B \quad \wedge \quad A \xrightarrow{r} C \quad \equiv \quad A \xrightarrow{r} B \vee C$

(4a) $\Box(C \supset B) \wedge \quad A \xrightarrow{r} B \quad \supset \quad A \xrightarrow{r} C$

(4b) $A \xrightarrow{r} C \quad \wedge \Box(A \supset B) \supset \quad B \xrightarrow{r} C$

(5) $false \xrightarrow{r} A \quad \equiv \Box\neg A$

(6) $A \xrightarrow{r} B \quad \supset (\Box\neg A \supset \Box\neg B)$

III. CAUSAL TABLEAUX

3-1. Tableaux Procedure

Tableaux procedure gives a mechanical reasoning method to construct a causal tableau which is essentially the model defined in the previous section. There are two types of causal tableau, *progressive causal tableau* and *regressive causal tableau* .

The progressive causal tableau is built in the following way. The tableau consists of two columns : *left column* and *right column*. Left column contains a set of possible worlds and right column is a necessary world. The initially introduced tableau is called *a main tableau* , which corresponds to the real world. The tableau is systematically extended by the following world extension rules.

World Extension Rules

[left column]

(PL1) If S is a set of formulas appeared in the left column, then we can add every formula A, such that $\Box(S \supset A)$ to the same column. (select only the relevant ones)

(PL2) If A such that $A \xrightarrow{p} B$ appears in the left column then we can create a new *auxiliary column* by putting B to it.

(PL3) [or-split]

If $A \vee B$ appears in the left column, then the column splits into two alternatives : either put A or put B. We call them *alternative columns.*

(PL4) [and-split]

If A_1, A_2 appear in the left column where $A_1 \xrightarrow{p} B_1$ and $A_2 \xrightarrow{p} B_2$ hold, then three new auxiliary columns can be created ; put $B_1, B_2, B_1 \wedge B_2$ to each new column. If $A_1, A_2, \dots, A_n$ appear in the left column where $A_1 \xrightarrow{p} B_1$, $A_2 \xrightarrow{p} B_2$, ..., $A_n \xrightarrow{p} B_n$ hold, then 2^{n-1} new auxiliary columns can be created.

[right column]

(PR1) If S is a set of formulas appeared in the right column, then we can add every formula A , such that $\Box(S \supset A)$. (select the relevant ones)

(PR2) If $\neg B$ appears in the right column such that $A \xrightarrow{p} B$ holds, then $\neg A$ is added.

(PR3) If $\neg B_1, \neg B_2, \dots, \neg B_n$ appear in the right column, such that $A_1 \xrightarrow{p} B_1$, $A_2 \xrightarrow{p} B_2$, ..., $A_n \xrightarrow{p} B_n$, then add $A_1, A_2, \dots, A_n$.

Note that right column is not splitted.

An auxiliary column is said to be *closed* if one of the following conditions is satisfied.

(i) A complementary pair of a formula appears

(ii) A formula P appears in it, while $\neg P$ appears in the right column of the main tableau

Alternative column is said to be *closed* if some auxiliary column of the alternative column is closed. The main tableau is said to be closed if one of the following conditions is satisfied.

(i) A complementary pair of a formula appears in the right column

(ii) All the alternative columns are closed

The construction of the regressive causal tableau is completely symmetric with that of the progressive causal tableau. In this case, the roles of the left column and the right column are turned over, that is, left column shows the necessary world and right column shows a set of possible worlds.

3-2. Reasoning Method by Causal Tableaux

Several types of causal reasoning can be executed by means of causal tableaux.

(1) Progressive causal verification (Forward chaining)

When a cause C is given and its effect E can be predicted, we prove the effect will possibly happen from the cause C and show the process which leads to the effect. It can be applied to the verification of behavioral correctness, that is, the system behaves as it is intended to.

In order to prove that $C \xrightarrow{p} E$ is valid, we try to construct its countermodel by using the progressive causal tableau. If the main tableau is not closed, we obtain a countermodel. Otherwise, $C \xrightarrow{p} E$ is proved to be valid.

[Example — Wheat Growing Problem I]

We will consider a closed world model determined by six elements ; fertilizer, rainfall, wheat-crop, demander, demand and supply relation, and price. There are two causal relations.

(i) Small rain-fall or lack of fertilizer causes the decrease of the amount of wheat crop

(ii) A small wheat crop or an increase of demander causes the wheat price to rise

We can formalize them in the followings :

$\neg F \vee \neg R \xrightarrow{p} \neg W \qquad \neg W \vee \neg N \xrightarrow{p} \neg P$

where

F : enough fertilizer is given

R : there is enough rain-fall

W : wheat-crop increases

N : population(demander) decreases

S : supply of the wheat is larger than the demand

P : price of the wheat falls

In addition, we assume the following logical property holds.

$\Box(S \supset P)$

It represents the fact that the supply of the wheat is larger than the demand implies the price of the wheat falls.

We will show that small fertilizer may cause the situation that the supply of the wheat is less than the demand, which is represented by $\neg F \xrightarrow{p} \neg S$.

At first we put $\neg F$ in the left column and S in the right column of the main tableau. And we will attempt to construct the progressive causal tableau according to the world extension rules.

left column	right column
¬F ¬F ∨ ¬R	S P
¬W ¬N ∨ ¬W	
¬P	

Fig.2 Wheat growing problem I (Progressive causal verification)

Since the main tableau is closed, $\neg F \xrightarrow{p} \neg S$ is proved to be valid.

(2) Simulation

Simulation means the analysis how the system behaves from the given condition. It is executed by using only left column of the progressive causal tableau. At first, the formulas which are satisfied in the initial state are put in the left column of the main tableau and possible worlds are created by the world extension rules. Each auxiliary column shows a possible behavior of the system.

We can apply this method to qulitative reasoning[2][3][5] which is a methodology that predicts and explains the behavior of mechanics of the physical systems in qualitative terms.

(3) Regressive causal verification (Backward chaining)

When an event E is observed and its cause C is suspected as a cause of E, we prove that C is actually the cause of E. The proof of the validity of $C \xrightarrow{r} E$ is done by means of the regressive causal tableau, symmetrically with the case of forward chaining.

(4) Causal analysis

Causal analysis is a process which infers a cause from an effect. It corresponds to the backward simulation, which is executed by using only right column of the regressive causal tableau . At first, the given formulas are put in the right column of the main tableau. And then possible worlds are created by the world extension rules for the regressive causal tableau .

In the process of this backward simulation, auxiliary column terminates if one of the following conditions is satisfied.

(i) There is no causal relations to extend the tableau, namely, causal relations are exhausted.

(ii) The world which already appeared in the column appears again. (It means that the system is cyclic)

Alternative column terminates if all the auxiliary columns of the alternative column terminate, and the main tableau terminates if all the alternative column terminate. Each auxiliary column shows the possibilities of the behavior of the system, and a set of formulas appeared in the auxiliary column except in the real world is the candidate for the causes of the effect.

[Example — Wheat Growing Problem II]

We will consider again a closed world model previously defined. We assume the follwing two causal relations.

(i) If the amount of wheat crop has increased, there were the rain-fall and enough fertilizer

(ii) If the wheat price has fallen, there were a large wheat crop and a decrease of population

We can formalize them in the followings :

$$F \wedge R \xrightarrow{r} W \qquad W \wedge N \xrightarrow{r} P$$

Similarly to the prior example, we assume the following logical property holds.

$$\Box(S \supset P)$$

We will show the procedure to find the cause if the price of wheat rises.

left column	right column
	P
	W ∧ N W N
	F ∧ R F R

Fig.3 Wheat growing problem II (Causal analysis)

Since no more worlds can be created, the procedure terminates, and all the formulas but P appeared in the tableau are candidates for the causes.

IV. APPLICATION

In this section an application to fault diagnosis on the basis of belief contravening hypothesis is presented in order to illustrate how the causal tableau is used to more sophisticated reasoning.

A belief contravening hypothesis is a supposition or assumption standing in a logical confliction with accepted beliefs or known facts. [8]

For example,

" Assume that Lincoln had been defeated by Douglas in the Presidential election in 1860 "

Since we believe, and indeed know that Lincoln won in 1860, it is a belief contravening hypothesis.

To make a belief contravening hypothesis, it is necessary to reject or modify some beliefs so that the ultimate residue is logically compatible with the hypothesis in question. This reasoning method is formulated by the causal tableaux as follows.

Assume that a belief contravening hypothesis F has happened in the circumstances of a set of beliefs $B_1, B_2, \ldots, B_n$. It means that $B_1 \wedge B_2 \wedge \ldots \wedge B_n \wedge F$ is inconsistent. We try to reject a belief $B_i (1 \leq i \leq n)$ and retain the others, and attempt to construct the progressive causal tableau so that $B_1 \wedge \ldots \wedge B_n \wedge F$ (except B_i) is consistent. If it can be constructed, then it is a model, which means that B_i is the cause of the fault.

Belief contravening hypothesis can be used in the wide range of applications[4] which include contingency planning, thought-experimentation and so on. In the followings, we apply it to the problem of fault diagnosis .

Fault diagnosis , which is an analysis of causes when some failure of a system happens, is well performed by the causal analysis if the dynamics of the system is specified in detail enough to handle failures and exceptional cases. However, since the persons surrounding a system usually have knowledges about only its normal behaviors, the failure of the system appears to them as a belief contravening hypothesis . Even in this case the cause for failure can be found out by rejecting some rules for normal function and inferring the failure part of the system, as is actually performed in the following example.

[Example — Simple plant controller]

We will consider a simple plant controller system.(Fig.4-1) It consists of two components of boiler and temperature controller which together form a feedback loop. The boiler has five temperature states, T1,T2,T3,T4 and T5, which indicate the state of the temperature t of water as follows.

$T1\,(t < a)$; $T2\,(t = a)$; $T3\,(a < t < b)$

$T4\,(t = b)$; $T5\,(b < t)$

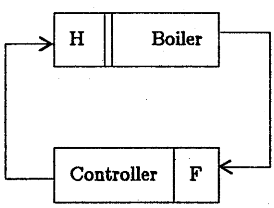

Fig.4-1 Simple plant controller system

The system is designed to keep the temperature between a and b by the following control rules ; if the temperature decreases to the degree less than a, then the switch is *on*, and it increases to the degree higher than b, then switch is *off*. If the switch of the boiler is *on*, then it causes the temperature of the water to increase, and if the switch is *off*, then it causes the temperature to decrease.

This is formulated in the followings :

Control Rule

Cr1 $\Box(T1 \supset on)$ Cr2 $\Box(T2 \supset on)$

Cr3 $\Box(T4 \supset off)$ Cr4 $\Box(T5 \supset off)$

Heat Transmission

H1 $\Box(on \supset increase)$

H2 $\Box(off \supset decrease)$

Frame Axioms

F1 $\Box(T1 \lor T2 \lor T3 \lor T4 \lor T5)$

F2 $\Box(increase \lor decrease \lor constant)$

F3 $\Box(on \lor off)$

Physical Causalities

P1 $(T1 \land decrease) \lor (T2 \land decrease) \xrightarrow{p} T1$

P2 $(T1 \land increase) \lor (T3 \land decrease) \xrightarrow{p} T2$

P3 $(T2 \land increase) \lor (T4 \land decrease) \xrightarrow{p} T3$

P4 $(T3 \land increase) \lor (T5 \land decrease) \xrightarrow{p} T4$

P5 $(T4 \land increase) \lor (T5 \land decrease) \xrightarrow{p} T5$

Assume that the water initially in the state of $T1$ becomes to be boiled in the state $T5$, although temperature is believed to be kept between a and b. We will find out the cause for the failure by a belief contravening hypothesis . Note that some control rule or heat transmission are more questionable than the frame axioms and physical causalities. Therefore we retain the frame axioms and physical causalities, while we examine doubtful rules.

(i) We try to reject Cr2, that is , we assume that $T2 \supset off$ does not work correctly. (Fig.4-2)

As $\neg T5$ appears in all the alternative columns, $T1 \xrightarrow{p} \neg T5$ is proven, which shows the failure of $T1 \xrightarrow{p} T5$. Therefore, we cannot conclude that the rule should be rejected. It is similar as in the case of Cr1 and H1.

(ii) Assume that the control rule Cr3 is rejected, that is, $T4 \supset off$ does not work correctly. (Fig.4-3)

In this case, the tableau gives a model which satisfies $T1 \xrightarrow{p} T5$. Therefore we conclude that it causes the failure of the system if the rule works incorrectly. Hence, it may be a cause. The case of Cr4 and H2 are in the similar situation.

V. CONCLUDING REMARKS

We have presented a mechanical method for representation and reasoning about causal relations. Since the causality is deeply related to the notion of time, we may extend the discussion on the basis of temporal logic or tense logic. In this case, " A causes B " is represented by " B regularly follows A ". Although this approach will give a similar result to this paper, we have not attempted it, because the time independent treatment of causal relations appears to give a simpler, and therefore, more practical reasoning method, and also because some cases are known in which it is not necessarily clear that whether " A causes B " is compatible with " B regularly follows A ".

[REFERENCES]

[1] Burks,A.W., " Chance, Cause, Reason ," Chicago Press, 1977.

[2] De Kleer, J. and Brown, J.S., " The Origin, Form and Logic of Qualitative Physical Laws, " IJCAI83, pp.1158-1169, 1983.

[3] Forbus, K.D., " Qualitative Reasoning about Physical Process, " IJCAI81, pp.326-330, 1981.

[4] Ginsberg, M.L., "Counterfactuals," Department of Computer Science, Stanford University, CA, 1984.

[5] Kuipers, B., " Getting the Envisionment Right, " Proceedings of National Conference on Artificial Intelligence, AAAI82, pp.209-212.

[6] Kripke,S.A., " A Completeness Theorem in Modal Logic, " The Journal of Symbolic Logic, Vol.24, No.1, pp.1-14, 1959.

[7] McCarthy,J. and Hayes,P.J., " Some Philosophical Problems from the Standpoint of Artificial Intelligence, " Machine Intelligence 4, 1966.

[8] Rescher,N., " Hypothetical Reasoning, " North-Holland, 1964.

[9] Simon,H.A., " Models of Discovering, " D.Reidel Publishing Corp., 1977.

[10] Weiss, S.M. et al., " A Model-Based Method for Computer-Aided Medical Decision-Making ," Artificial Intelligence, Vol.11, No.2, pp.145-172, 1978.

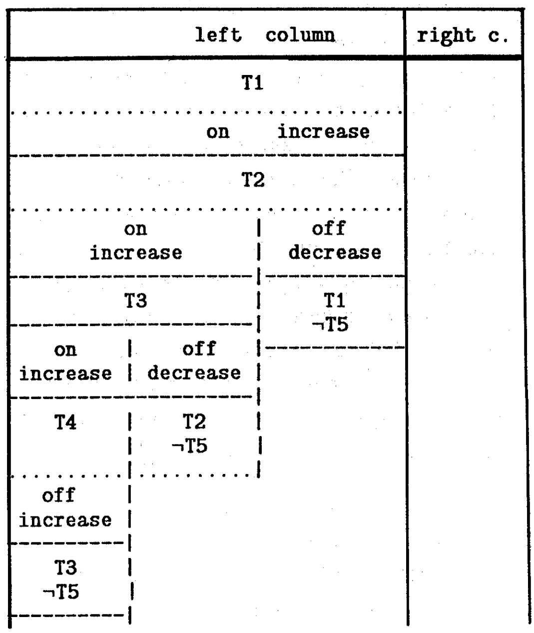

Fig.4-2 Belief contravening hypothesis (rejecting Cr2)

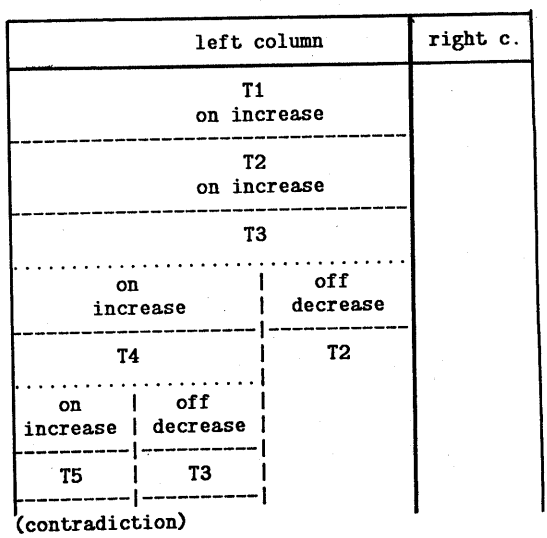

Fig.4-3 Belief contravening hypothesis (rejecting Cr3)