AD-785 031

SITUATIONS, ACTIONS, AND CAUSAL LAWS

John McCarthy

Stanford University

Prepared for:

Advanced Research Projects Agency

3 July 1963

Ch

1

STANFORD ARTIFICIAL INTELLIGENCE PROJECT
MEMO NO.2

July 3, 1963

SITUATIONS, ACTIONS, AND CAUSAL LAWS

by John McCarthy

Abstract: A formal theory is given concerning situations, causality and the possibility and effects of actions is given. The theory is intended to be used by the Advice Taker, a computer program that is to decide what to do by reasoning. Some simple examples are given of descriptions of situations and deductions that certain goals can be achieved.

The research reported here was supported in part by the Advanced Research Projects Agency of the Office of the Secretary of Defense (SD-183).

1. INTRODUCTION

Although formalized theories have been devised to express the most important fields of mathematics and some progress has been made in formalizing certain empirical sciences, there is at present no formal theory in which one can express the kind of means-ends analysis used in ordinary life. The closest approach to such a theory of which I am aware is made by Freudenthal in Lincos [1].

Our approach to the artificial intelligence problem requires a formal theory. Namely, we believe that human intelligence depends essentially on the fact that we can represent in language facts about our situation, our goals, and the effects of the various actions we can perform. Moreover, we can draw conclusions from the facts to the effect that certain sequences of actions are likely to achieve our goals.

In Programs with Common Sense [2], I discussed the advantages of having a computer program, to be called the Advice Taker that would reason from collections of facts about its problem and derive statements about what it could do. The name Advice Taker came from the hope that its behavior could be improved by giving it advice in the form of new facts rather than by rewriting the program. The reader is referred to that paper for further information about the Advice Taker and to Minsky's paper Steps Towards Artificial Intelligence [3] for a general introduction to the subject.

The first requirement for the Advice Taker is a formal system in which facts about situations, goals and actions can be expressed and containing the general facts about means and ends as axioms. A start is made in this paper on providing a system meeting the following specifications

#1. General properties of causality and certain obvious but until now unformalized facts about the possibility and results of actions are given as axioms.

#2. It is a logical consequence of the facts of a situation and the general axioms that certain persons can achieve certain goals by taking certain actions.

#3. The formal descriptions of situations should correspond as closely as possible to what people may reasonably be presumed to know about them when deciding what to do.

2. SITUATIONS AND FLUENTS

One of the basic entities in our theory is the situation. Intuitively, a situation is the complete state of affairs at some instant of time. The laws of motion of a system determine from a situation all future situations. Thus a situation corresponds to the notion in physics of a point in phase space. In physics, laws are expressed in the form of differential equations which give the complete motion of the point in phase space.

Our system is not intended for the complete description of situations nor for the description of complete laws of motion. Instead, we deal with partial descriptions of situations and partial laws of motion. Moreover, the emphasis is on the simple qualitative laws of everyday life rather than on the quantitative laws of physics. As an example, take the fact that if it is raining and I go outside I will get wet.

Since a situation is a complete state of affairs we can never describe a situation completely, and therefore we provide no notation for doing so in our theory. Instead, we state facts about situations in the language of an extended predicate calculus. Examples of such facts are:

1. raining (s)

   meaning that it is raining in situation s

2. time (s) = 1963.7205

   giving the value of the time in situation s. It will usually prove convenient to regard the time as a function of the situation rather than vice versa. The reason for this is that the numerical value of the time is known and important only where the laws of physics are being used.

3. at(I,home,s) or at(I,home)(s)

   meaning that I am at home in situation s. We prefer and will use the second of the given notations that isolates the situation variable since in most if not all cases we will be able to suppress it completely.

We shall not describe in this memorandum the logical system we intend to use. Basically, it is a predicate calculus, but we shall use the -notation and if necessary conditional expressions as in LISP or ALGOL We shall extend the meaning of the Boolean operators to operate on predicates. Thus by

$$\text{at}(I,\text{home}) \wedge \text{raining}$$

we mean the same as

$$\lambda s.\ \text{at}(I,\text{home})(s) \wedge \text{raining}(s)$$

A predicate or function whose argument is a situation will be called a *fluent*, the former being called a *propositional fluent*. Thus. *raining*, *time*, and *at(I,home)* are all fluents, the first and last being propositional fluents.

The term was used by Newton for a physical quantity that depends on time, and according to my limited understanding of what he meant, the present use of the term is justified.

In our formules we will usually manage to use the fluents without explicitly writing varieblee ropresenting situations. This corresponds to the use of random variables in probability theory without using variables representing points in the sample space even though random variables are supposed to be regarded as functions defined on a sample space.

In fact we shell go further and givenan interpretation of our theory as a sort of modal logic in which the fluente are not regarded ee functions et all.

3. CAUSALITY

In order to express causal laws we introduca the second ordar predicate cause. The statement

$$\text{cause}(\pi)(s)$$

where π is a propositional fluent is intended to mean that the situation s will lead in the future to a situation that satisfies tha fluent π. Thus, cause(π) is itsalf a propositional fluent. As an example of its use we write

$$\forall s.\forall p.\left[\text{person}(p)\wedge\text{raining}\wedge\text{outside}(p)\supset\text{cause}(\text{wet}(p))\right](s)$$

which asserts that a person who is outside when it is raining will get wet. Wa shall make the convention that if π is a fluent then

$$\forall\pi$$

means the same as

$$\forall s.\,\pi(s).$$

With this convention we can writa the pravious statement as

$$\forall\forall p.\text{person}(p)\wedge\text{raining}\wedge\text{outside}(p)\supset\text{causa}(\text{wet}(p)),$$

which suppresses explicit mention of situations. As a second example we give a special case of the law of falling bodies in tha form:

$$\forall\forall t.\forall b.\ \forall t'.\forall h(\text{real}(t)\wedge\text{real}(t')\wedge\text{raal}(h)\wedge\text{body}\ (b)$$
$$\wedge\ \text{unsupported}(b)\wedge\left[\text{height}(b)\ =\ h\right]\wedge\left[\tfrac{1}{2}gt^2<h\ \right]\wedge$$
$$\left[\text{tima}\ =\ t'\right]\supset\text{cause}(\text{height}(b)\ =\ h-\tfrac{1}{2}gt^2\wedge\text{time}=t'=t)$$

The concept of causality is intended to satisfy the two following general laws, which may be taken as axioms:-

C1. $\forall.\ \text{cause}(\pi)\wedge\ [\forall.\pi\supset\rho]\supset\text{causa}(\rho)$

C2. $\forall\ \text{cause}\ (\text{cause}(\pi))\supset\ \text{cause}(\pi)$

C3. $\forall.\text{cause}(\pi_1)\vee\text{causa}(\pi_2)\supset\ \text{cause}(\pi_1\ \dot{\vee}\ \pi_2)$

The fact that we can suppress axplicit mention of situations has the following interesting consequence. Instead of regarding the π's as pradicatas we may regard them as proppsitions and regard causa as a new modal operator. The operator $\forall$ seems then to be equivalent to the N (necessary) operator of ordinary model logic.

Conversely, it would appear that modal logic of necessity might be ragardad as a monadic pradicata calculus where all quantifiars ara over situations.

In the present case of causality, we seam to have our choice of how to proceed. Regarding tha system as a modal logic seems to have the following two advantages.

1. If we usa the predicate calculus interpratation we requira second order predicata calculus in order to handla causa(π)(s), while if we take the modal interpratation we can gat by with first order predicata calculus.

2. We shall want decision procaduras or at laast proof procedures for as much of our system as possible. If we usa the modal approach many problems will involve only substitution of constants for veriables in univeisel statements and will therafore fall into a fairly readily decidabla domain.

Another axample of causality is given by a 2-bit binary counter that counts avery second. In our formalism its behavior may be described by the statement:

$$\forall t \forall x_0 \forall x_1 / \text{ time} = t \wedge \text{bit}00 = x_0 \wedge \text{bit}1 = x_1 \supset \text{cause} \ ($$

$$\text{time} = t+1 \wedge (\text{bit } 0 = x_0 \oplus 1) \wedge (\text{bit } 1 = x_1 \oplus (x_0 \wedge 1)))$$

In this example $\underline{\text{time}}$, $\underline{\text{bit}00}$ and $\underline{\text{bit}11}$ are fluents whila t, x_0 and x_1 ara numerical veriables. The distinction is made clearer if we use the more long-winded statement

$$\forall s \forall t \forall x_0 \forall x1 . \text{time}(s) = t \wedge \text{bit } 0(s) = x_0 \wedge \text{bit}1(s) = x_1 \supset$$

$$\text{cause}(\lambda s'.\text{time}(s') = t+1 \wedge (\text{bit}0(s') = x_0 \oplus 1) \wedge (\text{bit}1(s') = x_1 \oplus (x_0 \wedge 1)))(s)$$

In this case however we can rewrite the statement in the form

$$\forall s . \text{cause}(\lambda s' . [\text{time}(s') = \text{time}(s)+1] \wedge [\text{bit}0(s') = \text{bit}0(s) \oplus 1] \wedge$$

$$[\text{bit}1(s') = \text{bit}1(s) \oplus (\text{bit } 0(s) \wedge 1)]) \ (s)$$

Thus we see that the suppression of explicit mention of the situations forced us to introduce the auxiliary quantities t, x_0 and x_1 which are required because we can no longer use functions of two different situations in the same formula. Nevertheless, the s-suppressed form may still be worthwhile since it admits the modal interpretation.

The time as a fluent satisfies certain axioms. The fact that there is only one situation corresponding to a given value of the time may be expressed by the axiom

$$\text{T1.} \forall \pi \forall \rho \forall t . \ \text{cause}(\text{time} = t \wedge \pi) \wedge \text{cause}(\text{time} = t \wedge \rho) \supset$$

$$\text{cause } (\text{time} = t \wedge \pi \wedge \rho)$$

Another axiom is

$$\text{T2.} \quad \forall t . \text{real}(t) \wedge t > \text{time} \supset \text{cause}(\text{time} = t)$$

4. ACTIONS AND THE OPERATOR can

We shall regard the fact that a person performs a certain action in a situation as a propositional fluent. Thus

$$\text{moves (person, object, location) (s)}$$

is regarded as asserting that person moves object to location in the situation s. The effect of moving something is described by

$$\forall\forall p\ \forall o\ \forall l.\ \text{moves}\ (p,o,l) \supset \text{cause}\ (\text{at}(o,l))$$

or in the long form

$$\forall s\ \forall p\ \forall o\ \forall l.\text{moves}\ (p,o,l)(s) \supset \text{cause}\ (\lambda s'.\text{at}\ (o,l)(s'))(s)$$

In order to discuss the ability of persons to achieve goals and to perform actions we introduce the operator can.

$$\text{can}(p,\pi)\ (s)$$

asserts that the person p can make the situation s satisfy. We see that can (p,π) is a propositional fluent and that like cause, can may be regarded either as a second order predicate or a modal operator. Our most common use of can will be to assert that a person can perform a certain action. Thus we write

$$\text{can}(p,\ \text{moves}\ (p,o,l))\ (s)$$

to assert that in situation s, the person p can move the object o to location l.

The operator can satisfies the axioms

K1. $\forall\forall\pi\ \forall\rho\ \forall p.\ [\text{can}\ (p,\pi) \wedge (\pi \supset \rho) \supset \text{can}(p,\ \rho)$

K2. $\forall\forall\pi\ \forall p_1\ \forall p_2.\ [\sim \text{can}(p_1,\ \pi) \wedge \text{can}(p_2,\ \sim\pi)]$

K3. $\forall\forall p\ \forall\pi\forall\rho\ \ [\text{can}(p,\pi) \vee \text{can}\ (p,\ \rho) \supset \text{can}(p,\pi \vee \rho)]$

Using K1 and

$$\text{can}(p,\ \text{moves}\ (p,\ o,\ l))$$

and

$$\forall\forall\rho\ \ \forall o\ \forall l.\ \text{moves}\ (p,\ o,\ l) \supset \text{cause}\ (\text{at}(o,l))$$

we can deduce

$$\text{can}\ (p,\ \text{cause}\ (\text{at}\ (o,l)))$$

which shows that the operators can and cause often show up in the same formula.

The ability of people to perform joint actions can be expressed by formulas like

$$\text{can}(p_1,\ \text{can}\ (p_2,\ \text{marry}\ (p_1,p_2)))$$

which suggests the commutative axiom

K4. $\forall\forall\ p_1\ \ \forall p_2\ \ \forall\pi.\ \text{can}(p_1,\ \text{can}(p_2,\pi)) \supset \text{can}\ (p_2,\ \text{can}(p_1,\pi))$

A kind of transitivity is expressed by the following:-

<u>Theorem</u> - From

1) can(p, cause(π))

and

2) $\forall. \pi \supset$ can(p, cause(ρ))

it follows that

3) can(p , cause(can(p, cause(ρ))))

<u>Proof</u> - Substitute can(p, cause(ρ)) for ρ in axiom C1 and substitute cause (π) for π and cause(can(p, cause(ρ))) for ρ in axiom K1. The conclusion then follows by propositional calculus.

In order to discuss the achievement of goals requiring several consecutive actions we introduce <u>canult(p,π)</u> whitch is intended to mean that the person p can ultimately bring about a situation satisfying π. We connect it with <u>can</u> and <u>cause</u> by means of the axiom

KC1. $\forall. \forall p \, \forall \pi. \, \pi \vee$ can(p,cause(canult(p,π))) $\supset$ canult(p,π)

This axiom partially corresponds to the LISP-type recursive definition

canult(p,π) = $\pi \vee$ can(p, cause(canult(p,π)))

We also want the axiom

KC2. $\forall \forall p \forall \pi.$ cause(canult(p,π)) $\supset$ canult(p,π)

5. EXAMPLES

1. The Monkey can get the Bananas

The first example we shall consider is a situation in which a monkey is in a room where a bunch of bananas is hanging from the ceiling too high to reach. In the corner of the room is a box, and the solution to the monkey's problem is to move the box under the bananas and climb onto the box from which the bananas can be reached.

We want to describe the situation in such a way that it follows from our axioms and the description that the monkey can get the bananas. In this memorandum we shall not discuss the heuristic problem of how monkeys do or even might solve the problem. Specifically, we shall prove that

$$\text{canult}(\text{monkey},\ \text{has}(\text{monkey},\ \text{bananas}))$$

The situation is described in a very oversimplified way by the following seven statements:-

H1. $\forall \forall u.\ \text{place}(u) \supset \text{can}(\text{monkey}, \text{move}(\text{monkey},\ \text{box}, u))$

H2. $\forall \forall u\ \forall v\ \forall p\ \text{move}\ (p, v, u) \supset \text{cause}(\text{at}(v, u))$

H3. $\forall\ \text{can}(\text{monkey}, \text{climbs}(\text{monkey}, \text{box}))$

H4. $\forall\ \forall u \forall v \forall p.\ \text{at}(v, u) \wedge \text{climbs}(p, v) \supset \text{cause}(\text{at}(v, u) \wedge \text{on}(p, v))$

H5. $\forall\ \text{place}(\text{under}(\text{bananas}))$

H6. $\forall\ \text{at}(\text{box},\ \text{under}(\text{bananas})) \wedge \text{on}(\text{monkey}, \text{box}) \supset \text{can}(\text{monkey}, \text{reach}(\text{monkey}, \text{bananas}))$

H7. $\forall\ \forall p\ \forall x.\ \text{reach}(p, x) \supset \text{cause}(\text{has}(p, x))$

The reasoning proceeds as follows: From 1 and 5 by substitution of under(bananas) for u and PC (propositional calculus) we get

1) $\text{can}(\text{monkey},\ \text{move}(\text{box},\ \text{under}(\text{bananas})))$

Using 1) and H2 and axiom C1, we get

2) $\text{can}(\text{monkey},\ \text{cause}(\text{at}(\text{box},\ \text{under}(\text{bananas}))))$

Similarly H3 and H4 and C1 give

3) $\text{at}(\text{box})\ \text{under}(\text{bananas})) \supset \text{can}(\text{monkey}, \text{cause}(\ \text{at}(\text{box}, \text{under}(\text{bananas})) \wedge \text{on}(\text{monkey}, \text{box})))$

Then H6 and H7 give

4) $\text{at}(\text{box}, \text{under}(\text{bananas})) \wedge \text{on}(\text{monkey}, \text{box}) \supset \text{can}(\text{monkey}, \text{cause}\ (\text{has}(\text{monkey}, \text{bananas})))$

Now, Theorem 1 is used to combine 2), 3) and 4) to get

5) $\text{can}(\text{monkey}, \text{cause}(\text{can}(\text{monkey}, \text{cause}(\text{can}(\text{monkey}, \text{cause}(\text{has}(\text{monkey}, \text{bananas}))))))$

Using KC1, we reduce this to

$$\text{canult}(\text{monkey},\ \text{has}(\text{monkey},\ \text{bananas})),$$

2. AN ENDGAME

A simple situation in a two person game can arise where player p_1 has two moves, but whichever he chooses player p_2 has a move that will beat him. This situation may be described as follows:-

1) $\mathrm{can}(p_1,m_1) \wedge \mathrm{can}(p_1,m_2) \wedge (m_1 \vee m_2)$

2) $[m_1 \supset \mathrm{cause}(\pi_1)] \wedge [m_2 \supset \mathrm{cause}(\pi_2)]$

3) $V.\, \pi_1 \vee \pi_2 \supset [\mathrm{can}(p_2, n_1) \wedge \mathrm{can}(p_2,n_2) \wedge (n_1 \vee n_2)]$

4) $V.\, (\pi_1 \wedge n_1) \vee (\pi_2 \wedge n_2) \supset \mathrm{cause}(\mathrm{win}(p_2))$

We would like to be able to draw the conclusion

3) $\mathrm{canult}(p_2,\mathrm{win}(p_2))$

We proceed as follows: From 1) and 2) we get

4) $\mathrm{cause}(\pi_1) \vee \mathrm{cause}(\pi_2)$

and we use axiom C3 to get

5) $\mathrm{cause}(\pi_1 \vee \pi_2)$

Next we weaken 3) to get

6) $V.\, \pi_1 \supset \mathrm{can}(p_2, n_1)$ and

7) $V.\, \pi_2 \supset \mathrm{can}(p_2, n_2)$

and then we use K1 to get

8) $V.\, \pi_1 \supset \mathrm{can}(p_2, \pi_1 \wedge n_1)$ and

9) $V.\, \pi_2 \supset \mathrm{can}(p_2, \pi_2 \wedge n_2)$

The propositional calculus gives

10) $V.\, \pi_1 \vee \pi_2 \supset \mathrm{can}(p_2, \pi_1 \wedge n_1) \quad \mathrm{can}(p_2, \pi_2 \wedge n_2)$

and using K3 we get

11) $V.\, \pi_1 \vee \pi_2 \supset \mathrm{can}(p_2, (\pi_1 \wedge n_1) \vee (\pi_2 \wedge n_2))$

which together with 4) and K1 gives

12) $V.\, \pi_1 \vee \pi_2 \supset \mathrm{can}(p_2, \mathrm{cause}(\mathrm{win}(p_2)))$

which together with 5) and C1 gives

13) $\mathrm{cause}(\mathrm{can}(p_2, \mathrm{cause}(\mathrm{win}(p_2))))$

Using the axioms for canult we now get

14) $\mathrm{canult}(p_2, \mathrm{win}(p_2))$.

6. NOTE

After finishing the bulk of this memorandum I came across *The Syntax of Time Distinctions* 4 by A.N. Prior. Prior defines model operators P and F where

$P(\pi)$ means 'it has been the case that π' and,

$F(\pi)$ means 'it will be the case that π'

He subjects these operators to a number of axioms and rules of inference in close analogy to the well-known [5] modal logic of possibility. He also interprets this logic in a restricted predicate calculus where the variables range over times. He then extends his logic to include a somewhat undetermined future and claims (unconvincingly) that this logic cannot be interpreted in predicate calculus.

I have not yet made a detailed comparison of our logic with Prior's, but here are some tentative conclusions.

1. The causality logic should be extended to allow inference about the past.

2. Causality logic should be extended to allow inference that certain propositional fluents will always hold.

3. cause(π) satisfies the axioms for his $F(\pi)$ which means that his futurity theory possesses, from his point of view, non-standard models. Namely, a collection of functions $p_1(t), p_2(t)$ may satisfy his futurity axioms and assign truth to $p(1) \wedge \sim(Fp)(0)$. In our system this is okay because something can happen without being caused to happen.

4. If we combine his past and futurity axioms, our system will no longer fit his axioms and

PF1. $p \supset \sim F(\sim P(p))$

PF2. $p \supset \sim P(\sim F(p))$

since we do not wish to say that whatever is, was always inevitable.

REFERENCES

1. Freudenthal, H.A., "LINCOS, Design of a Language for Cosmic Intercourse Part I," Amsterdam (1960)

2. McCarthy, J. "Programs with Common Sense", Proceedings Symposium on Mechanization of Thought Processes, Her Majesty's Stationery Office, London, England; 1959

3. Minsky, M.L., "Steps Toward Artificial Intelligence", Proceedings of the IRE, Vol. 29, No. 1, January 1961

4. Prior, A.N "The Syntax of Time Distinctions" Franciscan Studies (1958)

5. von Wright, G.H. "An Essay in Modal Logic", Amsterdam (1951)