A Logic of Planning Agency

Ricardo S. Silvestre

Adjunct Professor of Philosophy, Federal University of Ceará
Campus Cariri, Av. Castelo Branco 150, Juazeiro do Norte-Ce, Brazil
ricardo@lia.ufc.br

Abstract

The purpose of this paper is to give a formal account of a kind of agency so far neglected in the field of philosophical modal logic of action: planning agency. Interestingly enough, planning agency is something to which philosophy of action and other fields of logic had traditionally paid a considerable amount of attention. We therefore present here a logic of planning agency. We shall follow the standard approach of modal logics of agency exemplified by the works of Belnap, Chellas and Pörn. Moreover, since we believe there is a close relation between planning, time and indeterminism, we use the theory of branching time in the semantic characterization of our logic. We present the logic both syntactically and semantically, being the calculus sound and complete with respect to the semantics.

Keywords: Modal logic of action, planning agency, branching time theory.

1. Introduction

Davidson (1980) famously claimed that the mark of agency is intentionality under some description. According to Ascombe (1957), we need to distinguish between at least two different ways in which the concept of intention appears in agentive sentences, as exemplified in the statements below:

(1) Hamlet killed his uncle Claudius intentionally;
(2) Hamlet killed Claudius with the intention of revenging his father’s death.

It seems plausible however that statements of the form (1) might be reduced to statements of the form (2), in the sense that (1) is true iff

(1’) Hamlet killed Claudius with the intention of (thereby) killing Claudius

is also true, making it appear that what we have here is one, and not two kinds of intentionality. Suppose then that Hamlet indeed kills Claudius with the intention of killing him. In addition to that, suppose that Hamlet’s intention is realized a bit differently from how Shakespeare envisaged in his tragedy. Hamlet wants to kill his uncle by shooting at him. However, the bullet he fires misses Claudius by a mile, but the shot stampedes a herd of wild pigs that trample him to death.². In this case, it is at least dubious that Hamlet has killed Claudius intentionally. And the reason for that is that, in order to be so, it must be plausible that Hamlet succeeded in killing Claudius in a manner sufficiently in accordance with whatever plan he had for killing his uncle.

Thus statements of the form (2) sometimes have to refer to a plan, in the case above a plan describing the way through which Hamlet intended to kill Claudius. As an elaboration of this point, consider the following restatement of (2):

(2’) In killing Claudius, Hamlet intended to revenge his father’s death.

¹ This work is supported by CNPq (National Counsel of Technological and Scientific Development, Brazil) and FUNCAP (Fundação Cearense de Amparo à Pesquisa).
² This example is adapted from Davidson (1980) essay 4.
While here there seems to be no loss of meaning in relation to (2), the same does not happen with the statement below:

(3) Veronica mopped the kitchen with the intention of feeding her flamingo afterwards, which definitively is not the same as

(3’) In (by) mopping the kitchen, Veronica intended to feed her flamingo afterwards.

Therefore Bratman (1987) holds that statements of the form (2) are ambiguous between

(4) The agent F’d with the aim or goal of G’ing, and

(5) The agent F’d as part of a plan that incorporated an intention of G’ing.

(3) is an instance of (5), but not of (4). In other words, we need to distinguish intention as an aim or goal of actions and intention as a distinctive state of commitment to future action, a state that results from and subsequently constrains our practical endeavors as planning agents. Following Bratman (1987), I will call the distinctive form of agency involved in these examples planning agency.

Modal logic of action is the field of philosophical logic that tries to advance our grasp of agency with the help of some quite well-known techniques of formal modal logic. One of the key features of this approach is that it abstracts from making any reference, at the level of the logical language, at least, to actions, state changes or moments of time, describing actions simply as a relation between agents and state of affairs. This relation is represented by using a suitable modal operator to say that an agent brings it about that $\alpha$, sees to it that $\alpha$, is able to realize $\alpha$, tries to bring it about that $\alpha$, and the like, where $\alpha$ is a proposition describing a state of affairs. Supposing $\Delta$ is such an operator, $a$ an agent and $\alpha$ a proposition, the composite proposition $a\Delta\alpha$ will then mean “$a$ brings it about that $\alpha$”, “$a$ sees to it that $\alpha$”, etc.

Perhaps the most well-known drawback of this approach is that it oversimplifies the representation of actions. For instance, although (6) and (7) below mean quite different things, both would have, in most modal logics of action, the same logical form (something like John$\Delta$“the door is open.”)

(7) John opens the door;
(6) John keeps the door open.

As pointed by Sergot and Richards (2001), another serious problem with this approach arises from the fact that “sometimes it is essential to be able to refer to the means by which an agent brings about a state of affairs.” For instance, in the first example we gave, we mentioned the necessity of referring to the plan by which Hamlet intended to kill Claudius as well as to the way by which he did kill his uncle. Only when we have both is that we can say whether Hamlet killed Claudius intentionally and also whether the plan was successful or not. The problem is that even though there are hundreds of different ways by which Hamlet might have killed Claudius, each one possibly corresponding to a different plan, all of them are represented in modal logics of agency in the same way (as Hamlet$\Delta$“Claudius is dead”, say.)

According to some, there is the even more serious problem that the existing logics of action have tended to neglect the intentionality proper to human action. This is evidenced by their failure to see the import of the notion of attempt for the logic of agency, but also by their inability to take into consideration the planning aspect of human agency. Despite its acknowledging importance in the philosophy of action, no effort has been made so far, in

---

3 This example is taken from Wilson (2007).
5 This example is given in Pörn (1970).
6 Vanderveken (2005).
philosophical modal logics of action, to logically analyze what we are calling here planning agency.

In defense of this so-called modal approach to the analysis of action, we might perhaps say that these limitations are nothing more than the side-effect of the generality intrinsic to this approach. Exactly because modal logic of agency aims at being as general as possible in its formal treatment of action, it makes as few commitments as possible regarding agency, which trivially has as consequence a certain degree of limitation in its representational power. Commenting on the fact that his logic does not say at all what an action is, Belnap (2001) replies that this “has the advantage that it permits us to postpone attempting to fashion an ontological theory, while still advancing our grasp of some important features of action...”. Another advantage is that it allows flexibility for the easy combination of agency with a number of other concepts, such as powers, obligations, beliefs, etc., in a multi-modal setting.

However, and despite of this, as we have tried to show at the beginning of this section, it might perhaps be the case that we cannot really “advance our grasp of agency” unless we somehow account for the planning aspect of agency. Therefore, it might be a worthy task to develop a modal logic of action which takes into account the planning side of agency. That is the purpose of this paper. In doing this, we shall follow the standard approach of modal logics of action described above. Also, at this time following a specific family (we might say) of logics of action, we shall use the theory of ramified time to give our semantic account of planning agency.

2. Foundations of the Logic of Planning Agency

As we have said, we shall follow the standard approach of representing agentive sentences as a relation between agents and state of affairs, and of using a modal operator for that purpose. The particularity of our endeavor lies on the nature of this relation. Since we want to deal with planning agency, the agent and the state of affairs should be linked in such a way as to let explicit the planning aspect of the relation. But what precisely shall be this relation? What kind of information shall it encompass?

Let us start with some few remarks on the notion of plan. (1) Every plan has a goal, which we may take as a state of affairs which the plan aims at achieving. (2) A plan is something which might be executed or carried out, either successfully or unsuccessfully, and such execution might last for an extended period of time. (3) A plan is something whose existence is independent of its being carried out; we can conceive, talk about, analyze, judge, etc. a plan independently of its being carried out. (4) The way by which the agent plans to achieve his goal can, at least partially, be characterized by a structured set of states of affairs. (5) A plan is successful if and only if both its goal and the states of affairs which characterize the way the agent plans to achieve his goal are the case. And, finally, (6) each of these states of affairs which are part of the characterization of the way the agent plans to achieve his goal might (but not must) in its turn, have a structured set of states of affairs as a characterization of the way the agent plans to achieve it. The way we shall use these remarks make use of the presupposition that a state of affairs might be characterized as a proposition, statement, formula, or the like.

An example might help us to understand some of these remarks. Suppose I plan to be at Paraty at May 11th to attend the CLE 30/XV EBL/ XIV SLALM conference. The goal of my...

---

7 As for instance in Carmo and Pacheco (2001), Jones (2003) and Gelati et al. (2002).
10 Traditionally, in philosophy, a plan is considered to be a mental entity, whose existence is of course not dependent on the actions of the agent. However, our analysis does not depend on this assumption.
plan in this case might be characterized by the state of affairs expressed by the statement “At May 11th I am attending the CLE 30/XV EBL/ XIV SLAM conference in Paraty.” This is what (1) says. But how shall I achieve this? Well, I have to register at the conference, get ticket planes to Rio de Janeiro, be free from my teaching duties for that week, etc. This list of tasks I intend to perform in order be in Paraty at May 11th might be said to be the way I plan to achieve my goal. What (4) then says is that each one of these tasks might be characterized as a state of affairs, in this case the states of affairs expressed by the statements “I am registered at the conference”, “I have ticket planes to Rio”, “I have the week off”, etc. And trivially, in order to say that my plan was successful, both these states of affairs as well the goal shall be the case, vice-versa. That is what (5) says. Now, each one of these states of affairs, might, in their turn, be taken as a goal itself, in such a way as to have a set of states of affairs characterizing the way I plan to achieve it. For instance, I may plan to register at the conference going to the conference web-site, filling up the registration form, getting the payment form, filling it up and sending a check by mail, etc. And this might be done for each one of the members of the first set of states of affairs, as well as for the members of this “second level” set of states of affairs or for any level whatsoever. This is what (6) says. The term “might” in this phase is important, for it allows for the possibility that I do not have yet any idea of how, for example, I shall get my ticket planes (perhaps because I did not think of that yet.) This is the same to say that a plan is potentially partial. Another consequence of (4) and (6) is that a plan is structured.

Even though this sheds some light on the notion we want to explain, we have not answered the two questions posed on the first paragraph of this section. More specifically, we have not yet said which interpretation we are going to give to the modal operator we shall use to represent our planning relation. Basically what we are going to do is to deal exclusively with those states of affairs referred to in (4) and (6). In other words, the relation we shall formalize with the help of a modal operator is one that an agent has with those states of affairs that characterize the way the agent plans to achieve his goal.

Let ▶ be the syntactical representation of our modal operator, so that if a is an agent and α is a formula, a▶α is also a formula. Given what we have said in the paragraph above, we have at least two options concerning the meaning of ▶. The first is to suppose that the agent is carrying out a plan right now, and take a▶α as meaning “agent a brings it about that α as part of the plan he is carrying out.” The idea is first that (i) α’s being the case shall contribute to the achievement of a goal according to a specific plan, and second that (ii) the agent is right now carrying out that plan, and as part of that execution, he brings it about that α. But notice that we might want to speak about (i) without speaking about (ii). Being a plan something whose existence is independent of its being carried out, we might want to speak about the states of affairs which compose the way the agent plans to achieve his goal without committing ourselves with fact that the agent is right now carrying out the plan. We then have to consider a second meaning, in which case a▶α shall be read as “in order to carry out his plan, agent a must bring it about that α.” From a logical point of view, the main difference between the two readings is that ⊢ a▶α → α shall hold only in the former. Even though our emphasis here shall be on the first of these two readings, we shall elaborate a bit afterwards on how to modify our logic so as to take the second interpretation into account.

As Belnap (1988) pointed out, agency, branching time and historic modalities are logically related. The arguments he gives are of course directed towards ordinary agency, which in his case is formalized with the help of his STIT operator. However, this claim

11 We decided to use “brings it about that” instead of “sees to it that” in our description of the meaning of ▶ because the latter exhibits a clear intentional character, whereas the former may refer as well to unintentional actions (Hilpinen 1997). Even though we do think plans cannot be dissociated from intentionality, we also think that using an expression with less philosophical commitments is more in accordance with our generality guideline.
applies with perhaps greater strength in the case of planning agency. That planning agency requires time for its formalization seems quite clear: a plan is something whose execution lasts for a considerable extended period of time. Therefore, both in its conception as in its execution the time factor has to be considered. In addition to that, any analysis of the notion of plan should be in conformity with indeterminism. Since no action is fully determined, its execution might have different incompatible future effects. But since the execution of a plan presupposes the realization of several actions, its continuation, both in terms of execution as in terms of conception, has to take into account different incompatible future situations.

The theory of branching time\(^{12}\) is an attempt to incorporate indeterminism into a logic-semantic framework. According to the indeterminism, several simultaneous and incompatible moments of time might follow the same moment in the future of this world. Therefore, any moment of time can belong to several “histories” representing possible courses of the world, with the same past and present but with different historic continuations of that moment. In the theory of branching time, a moment is a possible complete state of the world at a certain instant. There is a countable a set of moments of time related causally and temporally through a relation of anteriority/posteriority. Due to indeterminism, this relation is partial, which makes the future to be ramified. However, there is the requirement that all moments be preceded by a common past moment. This guarantees that the past is unique. Consequently, the set of moments of time has the formal structure of a tree-like frame:

![Tree-like frame diagram]

What we called \textit{history} is then defined a maximal chain of moments of time, which represents a possible course of history of our world. In the figure above, history \(h_1\), for instance, contains (among others) moments \(t_8, t_4, t_2\) and \(t_1\); history \(h_2\) contains moments \(t_{10}, t_4, t_2\) and \(t_1\); history \(h_3\) contains moments \(t_{11}, t_5, t_2\) and \(t_1\), and so on and so forth. Two (or more) moments of time are said to be \textit{alternative} when they belong to histories which have the same past before these moments. In our figure, moments \(t_8, t_0\) and \(t_{10}\) are alternative, for they represent how the world could be immediately after moment \(t_4\). An instant is a member of a partition of the set of all moments of time containing exactly one moment of each history, and such that its members respect the temporal order of histories. In our figure, \(t_4, t_5, t_6\) and \(t_7\), for instance, are all members of instant \(i_3\).

Belnap (1988) and Vanderken (2005) defend that no agent could act so as to bring about an inevitable fact. Inevitable facts exist no matter what we do. Therefore, if an agent brings it about that \(\alpha\) as part of the plan he is carrying out, \(\alpha\) cannot represent an inevitable state of affairs. Inside the framework of the logic of branching time, we can define several notions of inevitability or historical necessity\(^{13}\). We might say for instance that \(\alpha\) is inevitable or historically necessary (in symbols: \(\square\alpha\)) at moment \(t\) iff \(\alpha\) is true at all moments alternatives to

---

\(^{12}\) Thomason (1970), Belnap and Green (1994).

\(^{13}\) Vanderveken (2005), Thomason (1984).
For instance, $\alpha$ is inevitable or historically necessary at $t$ iff $\alpha$ is true at $t_8$, $t_9$ and $t_{10}$. A stronger notion of inevitability is one which considers the instant to which the moment of the evaluation belongs: $\forall t \in \alpha$ is true at $t$ in this sense iff $\alpha$ is true at all moments belonging to the instant to which $t$ belongs. In our figure, $\alpha$ is inevitable at $t_6$ in this stronger sense iff $\alpha$ is true at all moments belonging to $i_4$, namely $t_8$, $t_9$, $t_{10}$, $t_{11}$, $t_{12}$, etc.

As we have said before, there is a close relation between planning, time and indeterminism. Since a plan is something whose execution lasts for an extended period of time, both its conception and its execution shall have to deal with the time factor. Moreover, since the effects of the actions which are part of the execution of the plan are not fully determined, in terms of branching time theory the execution of the plan shall take place in several different incompatible futures. Now suppose that an agent is carrying out a plan at a particular moment of time. Given this specific plan, there are histories which are compatible with its successful carrying out, and histories which are not. One should recall that the success of the execution of a plan implies not only that the state of affairs representing its goal is the case, but that all those states of affairs which characterize the way the agent plans to achieve the goal, and are such that he should bring it about that they are the case in the execution of the plan, are also the case. Now, suppose a plan is being carried out at moment $t_4$, and that the histories compatible with its successful execution are $h_1$, $h_3$, $h_5$, $h_8$, $h_9$. What then can we say about those formulas $\alpha$ which are true at moments $t_4$, $t_5$ and $t_6$, which are the moments $t$ such that, for at least one he $\{h_1$, $h_3$, $h_5$, $h_8$, $h_9\}$, $t \in h$ and $t \in i$, where $i$ is the instant to which $t_5$ belongs, namely $i_3$? Trivially that $\alpha$ is always the case considering the success of the plan the agent is carrying out at moment $t_5$.

Can we then associate those formulas with the states of affairs the agent brings it about at $t_5$ as part of the plan he is carrying out at $t_5$? In other words, can we say that, in this case, $\alpha \triangleright t_5$ is true at $t_5$? Not yet, and the reason for that is twofold. First, among those $\alpha$’s, there are surely those which are true no matter what happens, that is to say, those whose truth is inevitable or historically necessary. Therefore, in order to give an account of the meaning we wish to attribute to $\triangleright$, we have to get rid of these inevitable formulas in the semantic analysis of $\triangleright$. Second, it is not enough to consider just the histories which are compatible with the successful carrying out of the plan. Even though this accounts for the success aspect of agency (if $\alpha$ is true in all $t$-co-instantaneous moments belonging to those histories, then it is because the truth of $\alpha$ is guaranteed by the successful execution of the plan), it does not take into consideration that $\alpha$ might be true by something else than an act of agent $a$. From the point of view of $a$’s agency, $\alpha$’s truth might be an accident.

The second of these problems is sorted out by considering not only histories that are compatible with the successful execution of the plan, but those ones which are under the control of agent $a$ in the execution of the plan he is carrying out at moment $t$. We shall call such histories the plan-histories of $a$ at moment $t$. Moments of time belonging to the same history are temporally and causally related. As one might expect, this causal relation is also due to the actions the agents do at those moments. The idea then is that the moments posterior to moment $t$ and which belong to some plan-history of agent $a$ at $t$ are under the control of, or responsive to (as Chellas would put) the planning actions which $a$ does at $t$. Trivially enough, these plan-histories shall also be compatible with the successful execution of the plan $a$ is carrying out (even though there might be compatible histories which are not under the control of $a$.)

The first problem is solved, first, by using another modal operator, say $\triangleright$, to be evaluated with the help of the plan-histories: $\alpha \triangleright t_5$ is true at $t_5$ iff for all moments of time $t'$ which are $t$-co-instantaneous with $t$ and belong to at least one of the plan-histories of $a$ at $t$, $\alpha$ is true at $t'$. $\alpha \triangleright \alpha$ shall be read as “$\alpha$ is always the case in the successful execution of the plan
the agent is carrying out.” However, a\(\triangleright\alpha\) still does not consider the non-inevitable aspect required by agency. We therefore shall define a\(\triangleright\alpha\) as a\(\triangleright\alpha\land\neg\Box\alpha\), where \(\Box\) is the (stronger) modality of historical necessity we showed above. In other words, we say that agent \(a\) brings it about that \(\alpha\) as part of the plan he is carrying out iff \(\alpha\) is always the case in the successful execution of the plan the agent is carrying out, and \(\alpha\) is not historically necessary. In this way, if a\(\triangleright\alpha\) is true at moment \(t\) we guarantee that \(a\) brings it about that \(\alpha\) (as part of the plan he is carrying out) and that \(\alpha\) is neither inevitable (something which would be true no matter what) nor accidental (something which is not inevitable, but is true due to something else than \(a\)’s actions.)^{14}

3. The Logic of Planning Agency

Here we shall present the language, semantics and calculus of the logic of planning agency (which contemplates only the propositional case.) In addition to \(\land, \neg\) and \(\top\), we shall have the modal operators \(\triangleright\) and \(\Box\), as primitive logical constants. If \(a\) is an agent and \(\alpha\) is a formula, a\(\triangleright\alpha\) means that \(\alpha\) is always the case in the successful execution of the plan the agent is carrying out. \(\Box\) is the modality of inevitability we mentioned above. However, in contrast to what was suggested, it is not the stronger, but the weaker historical necessity modality, which we defined through the notion of alternative moments of time. The reason for that shall be clear when we present the semantic postulates of our logic.

As we have said, \(\triangleright\) shall be introduced as an abbreviation from \(\triangleright\) and \(\Box\): a\(\triangleright\alpha\) \(\equiv\) a\(\triangleright\alpha\land\neg\Box\alpha\). Besides \(\triangleright\), we also define a modality of historical possibility: \(\Diamond\alpha\equiv\neg\Box\neg\alpha\). \(\lor\) and \(\rightarrow\) and \(\perp\) are defined in the usual way. In addition to all tautologies of the propositional calculus, our axiomatic has two sets of axioms, one for \(\Box\) and other for \(\triangleright\):

Axioms for \(\Box\)
\begin{align*}
N1. & \Box(\alpha\rightarrow\beta)\rightarrow(\Box\alpha\rightarrow\Box\beta) \\
N2. & \Box\alpha\rightarrow\alpha \\
N3. & \Diamond\alpha\rightarrow\Box\Diamond\alpha
\end{align*}

Axioms for \(\triangleright\)
\begin{align*}
P1. & a\triangleright\alpha\rightarrow\alpha \\
P2. & \Box\alpha\rightarrow a\triangleright\alpha \\
P3. & a\triangleright\alpha\rightarrow\Diamond\alpha \\
P4. & a\triangleright(\alpha\rightarrow\beta)\rightarrow(a\triangleright\alpha\rightarrow a\triangleright\beta) \\
P5. & a\triangleright\alpha\rightarrow\Box(a\triangleright\alpha)
\end{align*}

As rules of inference we have modus ponens and necessitation (\(\alpha/\Box\alpha\)). The definition of the relation of deduction is done in the usual way.

For the semantics, a model is a sextuple \(<T, <, A, P, \parallel, V, >\) where
\begin{enumerate}
\item \(T\) is a non-empty set of moments of time;
\item \(<\) is a non-reflexive, transitive and asymmetric relation between moments of time such that, for any \(t_1, t_2, t_3 \in T\), if \(t_3\) is such that \(t_1 < t_3\) and \(t_2 < t_3\), then either \(t_1 = t_2\) or \(t_1 < t_2\) or \(t_2 < t_1\) (no backward ramification condition);
\item \(A\) is a non-empty set of agents.
\end{enumerate}

\(^{14}\) This would reply the criticism made by Elgesem (1997) about Belnap’s and Pörn’s logic, which, in our view, is due to a misunderstanding about the theory of branching time.
A plan-history is a set of histories containing, for a given agent and a given moment of time, all histories that are under the control of, or responsive to, the actions of the agent in the execution of the plan he is carrying out at that moment. We shall also refer as plan-histories the histories which are members of a plan-history.

(iv) \( P: A \times T \rightarrow \mathcal{R}(H) \)\(^{15} \) is a function which, for each agent \( a \) and moment \( t \), gives the plan-history of \( a \) at \( t \). \( P \) satisfies the following conditions:

1. For at least one \( h \in P(a,t) \), \( t \in h \). \( \quad \) (success condition)
2. \( P(a,t) \neq \emptyset \); \( \quad \) (non-contradiction condition)
3. If \( h' \in P(a,t) \), then for all \( h \in H \) such that \( t \in h, h \cong h' \); \( \quad \) (historical relevance condition)
4. If \( T_t = T_{t'} \), then \( P(a,t) = P(a,t') \); \( \quad \) (alternative plans condition)

(v) \( \llbracket \| \rrbracket \) is a function which maps the (names of) agents to elements of \( A \).

(vi) \( V: T \times P \rightarrow \{ \text{True, False} \} \) is a function which, given a moment of time, maps propositional symbols to truth-values.

In the logic of ramified time, since the logical language has future modalities, the evaluation relation \( \models \) (which, given a model \( M \), says whether a formula \( \alpha \) is true at \( M \)) has as parameter, besides a moment of time, also a history. Therefore, our relation \( \models \) shall have, besides moments of time, also histories as parameters. The definition of \( \models \) for \( \land, \neg \) and \( \top \) is done in the usual way. Below we have the truth-definitions of \( \Box \) and \( \triangleright \): \( M \models_{\neg, \Box} \alpha \) sss for all \( t' \in T \) such that \( t' \in T_t \) and all \( h' \) such that \( t' \in h', M \models_{\neg, \Box} h' \alpha \)

\( M \models_{\neg, \Box} \neg \alpha \) sss for all \( h' \in P(l(t),t) \) and \( t' \in h' \) such that \( T_t = T_{t'} \), \( M \models_{\Box} h' \alpha \).

The notions of validity in \( M \) and consequence relation are defined in the usual way. The calculus showed above is sound and complete with respect to this semantics.

Let us now comment the conditions we have imposed on function \( P \) and answer why we have used the weaker instead of the stronger modality of historical necessity. The success condition says that for at least one of the plan-histories \( h \) of \( a \) at \( t \), \( h \) is such that \( t \) belongs to it. This guarantees that the plan is in fact being carried out by \( a \) at \( t \), that is to say, that \( a \) brings it about the state of affairs related to the plan he is carrying out at \( t \). This is the condition which guarantees the truth of \( P1 \). In order to obtain the other reading for \( \triangleright \) we have mentioned at the beginning of Section 2 (something like “in order to carry out his plan, agent \( a \) must bring it about that \( \alpha' \)”) we have just to drop the success condition, on the semantic level, and axiom \( P1 \), on the syntactic side. This shall of course give rise to a different logic of planning agency. The non-contradiction condition simply says that set of plan-histories cannot be empty. If it could, by vacuity we would have formulas such as \( a \triangleright \bot \) as satisfiable. A consequence of this condition is that \( \neg a \triangleright \bot \) is a tautology.

The plan agent \( a \) is carrying out, say, at moment \( t_{11} \) was obviously conceived or created at a moment \( t' \) anterior to \( t_{11} \) (in symbols: \( t' < t_{11} \)) \( P(a,t_{11}) \) contains all histories which are responsive to the actions \( a \) does in the execution of the plan he is carrying out at \( t_{11} \). We have also said that a history is a maximal chain of moments of time representing a possible course of the history of our world. But depending on the moment we are considering, some histories cannot be considered any more, from the point of view of that moment of time, as possible courses of the history. For instance, from the point of view of moment \( t_2 \), that is, if \( t_2 \) were the present moment, only histories which pass through \( t_2 \), namely \( h_1, \ldots, h_6 \), are still possible courses of the history of the world. Neither \( h_7, \ldots, h_{12} \) are possible any more if \( t_1 \) were the present moment. An obvious corollary of this is that none of these histories could be under the control of the actions agent \( a \) does in the execution of his plan. Consequently,

\(^{15} \mathcal{R}(A) \) is the powerset of set \( A \).
supposing $a$ formulates his plan at $t_5$, these histories ($h_7, h_8, \ldots, h_{12}$) could not belong to its corresponding set of plan-histories.

Now, suppose that the plan which agent $a$ is carrying out at $t_{11}$ was indeed formulated at $t_2$. Then neither $h_4$ nor $h_5$ nor $h_6$ nor $h_7$ shall be part of this plan. Note that there is a moment of time between $t_2$ and $t_{11}$, namely $t_5$. Adopting the quite reasonable assumption that $a$ reviews his plan always when he has opportunity to do it, when $t_5$ becomes the present moment it will not make sense to keep the plan as it was formulate at $t_2$, for there are histories such as $h_1$, $h_2$ and $h_3$ which might belong to the plan formulated at $t_2$ (since they were possible at $t_2$) but which are not possible any more at $t_5$. That this is the case is trivially seen by observing that $h_1$, $h_2$ and $h_3$ cannot be under the control of the actions $a$ does at $t_5$. If he does review his plan at $t_5$, these histories trivially will not belong to the corresponding new plan-history. Therefore, adopting this principle of plan reviewing, we can conclude that the plan which $a$ is carrying out at $t_{11}$ was conceived (or had its last review made) at $t_5$. Generalizing the reasoning, we can say that the plan an agent is carrying out at moment $t$ was formulated (or reformulated) at the moment immediately anterior to $t$.

These observations have important implications for the formal structure of our semantics. The first one is formalized by the historical relevance condition. This condition says that if $h'$ belongs to the plan-histories of $a$ at $t$, then this history and all the histories which pass through $t$ should share the same past at $t$. In other words, only those histories which share the same past at $t$ with the histories passing through $t$ can belong to $P(a,t)$. In our figure, $h_4$ cannot belong to $P(a,t_{11})$, for instance. This is trivially a consequence of the fact that only those histories which are possible courses of the history of the world at the time of formulation of the plan, that is, the moment immediately anterior to $t$, can belong to $P(a,t)$. The second implication is taken into account by the alternative plans condition. This condition says that at all alternative moments, $a$ should carry out the same plan. The reason for that is that the moment immediately anterior to all these alternative moments is the same. Since this is the moment where the plans the agent is executing at those alternative moments were formulated, and supposing that an agent does not have the ability to formulate more than one plan at the same moment, we can conclude that the plan an agent is executing at a moment $t$ is the same as the one he is executing at all moments alternatives to $t$.

Finally, these observations imply that the non-inevitability aspect we should guarantee in our account of planning agency shall be confined to the alternative moments of the moment of evaluation. If $P(a,t)$ can contain only histories passing through alternative moments of $t$, then the inevitable formulas we should get rid of in the evaluation of $\bigtriangledown$ shall be those which are inevitable at those alternative moments. Therefore, the weaker historical necessity is enough for our purpose.

References


