LOGIC, PLANNING AGENCY AND BRANCHING TIME

RICARDO SOUSA SILVESTRE
Universidade Federal de Campina Grande

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. In doing this we follow the standard approach of modal logics of agency exemplified by the works of Belnap, Chellas and Pörn. Since we believe there is a close relation between planning, time and indeterminism, we use the theory of branching time as a conceptual framework for investigating the basic features of planning agency. Besides introducing a branching-time semantics, we also provide a calculus sound and complete with respect to this 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. Regarding what we might call agentive sentences, that is to say, sentences ascribing an act to an agent, we can distinguish between at least two different kinds of intentionality (Anscombe 1957). For instance, in (1) and (2) below

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

while (1) refers to the fact that Claudius murder was done intentionally by Hamlet, as opposed to accidentally, for example, (2) says more: Hamlet murdered his uncle with the specific intention of revenging his father's death. Someone might object by saying that (1) can be reduced to something very alike (2):

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

clearly (1) is true if and only if (1′) is true. Therefore we have not two, but only one kind of intentionality.

As a reply, suppose that Hamlet really killed 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.[1] In this case, while (1′) is still true, it is at least dubious that Hamlet has killed his uncle intentionally.

Published by NEL — Epistemology and Logic Research Group, Federal University of Santa Catarina (UFSC), Brazil.

Apparently we are entitled to say that Hamlet killed Claudius intentionally only if he succeeded in doing so in a manner sufficiently in accordance with whatever *plan* he envisaged for killing his uncle. Thus we see the import of the notion of plan for ascribing intentionality to an agent.

As an elaboration of this last point, consider the following restatement of (2):

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

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,[2]

which definitively is not the same as

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

While (3) sets Veronica's mopping the kitchen as part of something like a plan that incorporates an intention of feeding her flamingo, (3′) sets the feeding of the flamingo as the very goal of the action of mopping the kitchen. Therefore Bratman (1987) holds that statements of the form (2) are ambiguous between

(4) The agent *F*'d as part of a *plan* that incorporated an intention of *G'*ing.
(5) The agent *F*'d with the aim or goal of *G*'ing, and

(3) is an instance of (4), but not of (5). In other words, there is 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 refer to the distinctive form of agency involved in (3) and (4) as *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 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. It represents agentive sentences simply as a relation between agents and states of affairs. This is invariably done by using a modal operator meant to say that an agent brings it about that α, sees to it that α, is able to realize α, tries to bring it about that α, and the like, where α is a proposition describing a state of affairs. Supposing Δ is such an operator, a is an agent and α is a formula, the composite formula $a\,\Delta\,\alpha$ means "a brings it about that α", "a sees to it that α", etc.[3]

Probably 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 Δ "the door is open".)

(6) John opens the door;
(7) John keeps the door open.[4]

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 it about that a certain state of affairs is the case. In our first example we saw the need of referring to the plan by which Hamlet intended to kill Claudius as well as to the way by which he effectively killed his uncle. Only when we have both is that we can say whether Hamlet killed Claudius intentionally or not as well as 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 Δ "Claudius is dead", say.)

In defense of this so-called modal approach to the analysis of action, one might say that these limitations are nothing more than the side-effect of the generality intrinsic to the 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 power, obligation, belief, etc. in a multi-modal setting.[5]

While we might agree on these alleged advantages, these drawbacks reveal a more serious and deeper problem with the existing logics of action: that they tend to completely ignore the intentionality proper to human action. In support to this claim, Vanderveken (2005) points the failure of most logics of action to see the import of the notion of attempt for the logical analysis of agentive sentences. As our comment on Sergot's criticism shows, this is shown also by the inability of these logics to take the planning aspect of human agency into consideration. Despite its acknowledging importance in the philosophy of action, no effort has been made so far, in philosophical modal logics of action, to logically analyze what we are calling here planning agency.

Interestingly enough, this contrasts sharply with the field of Artificial Intelligence, where the notion of plan has always played a crucial role in traditional logical approaches to action.[6] This is not accidental. The sorts of agents Artificial Intelli-

gence researchers are interested in are artificial agents. By definition, any action an artificial agent performs has to somehow contribute to the achievement of a goal; artificial intelligent agents are essentially planning agents. However, as pointed out by Bratman (1987), we human beings are also planning agents. Therefore perhaps it is the case that we cannot really advance our grasp of agency, to use Belnap's words, unless we somehow account for the planning aspect of agency.

We intend here to introduce a modal logic of action which seriously takes into account the planning side of agency. In doing this, we shall follow the standard approach of modal logics of action as described above. First, we want our logic to be as general as possible regarding the philosophical aspects of planning agency; in the same way that modal logic of agency leaves many important questions about agency unanswered, the number of aspects of planning agency taken into consideration in our analysis shall also be limited. Second, by making use of a specific modal operator we shall represent agentive sentences as a relation between agents and states of affairs. 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. Third, following a traditional guideline in philosophical modal logic, we consider here what might be called ideally rational agents. Finally, in order to account for the subtleties of the notion of planning agency we make use of the theory of branching time[7] in the construction of the semantics of our logic.[8]

Our plan for this paper is as follows. In order to give a preliminary description of our attempt to formally analyze planning agentive sentences, we lay down in the next section some basic postulates about planning agency. As a further step, we use the conceptual framework provided by the theory of branching time to elaborate on some key logical features of planning agency. This is done in Section 3. In Section 4 we introduce both semantically and axiomatically the logic of planning agency. In Section 5 we discuss some drawbacks of our approach and elaborate a bit on how they can be sorted out, and finally, in Section 6, we lay down some conclusive remarks.

2. On the Foundations 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. As we have also said, 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 is this relation? And what kind of information shall it encompass?

In order to answer these questions, let us lay down some (hopefully) commonsensical postulates about planning agency:

Postulate 1: A plan is something which might be executed or carried out, either successfully or unsuccessfully, and such execution might last for some extended period of time. Nevertheless, a plan's existence is independent of its being carried out or not.

Here we set four basic features of plans: (1) that they might be executed, in which case (2) the execution is either successful or unsuccessful, (3) that such execution lasts for some period of time and (4) that their existence does not depend on their being carried out.

Postulate 2: Every plan has a goal which it aims at achieving and which might be characterized as a state of affairs.

This postulate says that attached to every plan is a distinguished state of affairs, called the goal of the plan, which the agent who carries out the plan aims at. If I plan, for example, to be in Petrópolis-RJ at May 13th to attend the 16th Brazilian Logic Conference, then the goal of my plan might be characterized by the state of affairs expressed by the statement "At May 13th I am attending the 16th Brazilian Logic Conference in Petrópolis-RJ."[9]

Postulate 3: The way by which the agent plans to achieve his goal might (at least partially) be characterized by a structured set of states of affairs.

Given that I know that the goal of my plan is to be in Petrópolis at May 13th attending the 16th Brazilian Logic Conference, the next step is to know how I shall achieve this goal. Perhaps the first thing I have to do is to get free from my teaching duties for that week, after which I must register at the conference, get ticket planes to Rio de Janeiro, book a hotel room, and so on and so forth. Besides, I know that I have to perform these tasks in a structured way. For instance, I know that I cannot do anything else until I get free from my teaching duties; but once this is done, it does not really matter if I first get the ticket or book the hotel room. Since each one of these tasks can be characterized by a specific state of affairs, namely the states of affairs expressed by the propositions "I have the week off", "I am registered at the conference", "I have ticket planes to Rio", "I have a hotel room booked", etc., the way I plan to achieve my goal might be characterized as such states of affairs organized according to a specific structure. This is what postulate 3 says.

Postulate 4: A plan is successful only if both its goal and the states of affairs which characterize the way the agent plans to achieve his goal are the case.

Trivially, I can say that my plan to get to the conference was successful only if I get to be there according to the manner I planned. In other words, I can say that my plan was successful only if both my goal and the states of affairs which characterize the way I planned to achieve it are the case.

Postulate 5: Each state of affairs which is part of the characterization of the way the agent plans to achieve his goal might in its turn has a structured set of states of affairs as a characterization of the way the agent plans to achieve it.

Postulate 5 says that each one of the states of affairs mentioned in Postulate 4 might, in their turn, be taken as a goal itself—in this case a *sub-goal*—in such a way as to have a set of states of affairs characterizing the way the agent plans to achieve it. For instance, I may plan to register at the conference by 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 and so on and so forth for any level.

The word "might" in this postulate is important for it allows for the possibility of not specifying how a specific state of affairs, member of the way I plan to achieve my goal, shall be achieved. It might be, for instance, that I did not yet think of how to achieve such a sub-goal, or even that it is such a simple task that no plan is required to carry it out. Thus we have that besides being *structured* a plan is also *partial*:

Corollary: A plan is structured and partial.

With these postulates at hand we can give a preliminary answer to the questions we have posed at the beginning of this section. First of all, in assigning an interpretation to our planning-relation modal operator we are going to *deal exclusively* with those states of affairs referred to in Postulate 3. In other words, the relation we shall formalize with the help of our modal operator is the relation that an agent has with those states of affairs that characterize the way the agent plans to achieve his goal.

Let $\blacktriangleright$ be the syntactical representation of our modal operator. If a is an agent and α is a formula then $a \blacktriangleright \alpha$ is also a formula. Given what we have said in the paragraph above, we have at least two options concerning the meaning of $\blacktriangleright$. The first one is to suppose that the agent is carrying out a plan right now and take $a \blacktriangleright \alpha$ as meaning "agent a brings it about that α as part of the *plan* he is carrying out."[10] The idea is basically that (1) α's being the case shall contribute to the achievement of a goal according to a specific plan and that (2) the agent is right now executing that plan, and as part of that execution, he brings it about that α.

But notice that we might want to speak about (1) without speaking about (2). Being a plan something whose existence is independent of its being carried out (Postulate 1), we might want to speak about the states of affairs which compose the

way the agent plans to achieve his goal without committing ourselves to fact that the agent is right now carrying out that plan. We then have to consider a second meaning, namely one in which $a \blacktriangleright \alpha$ is 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 these two readings of $a \blacktriangleright \alpha$ is that $a \blacktriangleright \alpha \rightarrow \alpha$ holds only regarding the former one. Even though our emphasis shall be exclusively on the *first of these two readings*, we shall afterwards elaborate a bit on how to modify our logic so as to take into account the second meaning.

3. Planning Agency, Branching Time and Historical Necessity

As Belnap (1988, 1992, 2001) has pointed out, agency, branching time and historical modalities are logically related. Even though the arguments he gives are directed towards ordinary agency, which in his case is formalized with the help of his STIT operator,[11] this claim applies with perhaps greater strength in the case of planning agency.

That planning agency requires time for its formalization seems quite clear: as we have laid down in Postulate 1, 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,[12] 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 and conception, has to take into account different incompatible future situations.

The theory of branching time[13] is an attempt to incorporate indeterminism into a logical-semantic framework. According to indeterminism, several simultaneous and incompatible moments of time might follow the same moment in the future of the 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 causally and temporally related 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 is a *tree-like structure* (cf. Fig. 1).

What we have called *history* can be formally defined as a maximal chain of moments of time, representing, as we have said, a possible course of history of our

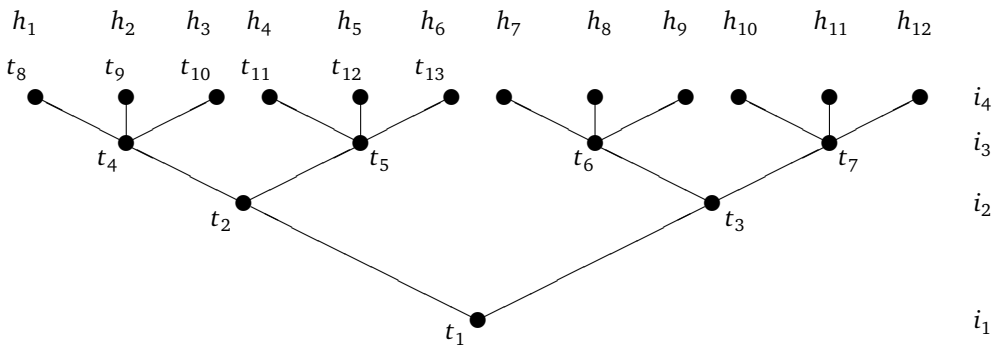

Figure 1:

world. The moments of time belonging to the same history are temporally and causally related to each other. In Figure 1, history h_1 contains (among others) moments t_8, t_4, t_2 and t_1; history h_3 moments t_{10}, t_4, t_2 and t_1; history h_4 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 *alternative* when they belong to histories which have the same past before these moments. In our figure, moments t_8, t_9 and t_{10} are alternative; 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, in which case we say they are *co-instantaneous* to each other.

Belnap (1988, 1992) along with Vanderveken (2005) defend the quite intuitive principle 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 α as part of the plan he is carrying out, α 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*.[14] We might say for instance that α is inevitable or historically necessary (in symbols: $\Box\alpha$) at moment t iff α is true at all moments alternatives to t. For instance, α is inevitable or historically necessary at t_8 iff α 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 evaluation belongs: $\Box\alpha$ is true at t in this sense iff α is true at all moments belonging to the instant to which t belongs. In our figure, α is inevitable at t_8 in this stronger sense iff α 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, there is a close relation between planning agency, time and indeterminism. Since the effects of the actions which are part of the execution of a plan are not fully determined, it (the execution of the plan) shall take place in

several different incompatible futures. Now suppose agent a 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 are also the case. Suppose then that a's plan is being carried out at moment t_5, and that the histories compatible with its successful execution are h_1, h_3, h_5, h_8 and h_9. There is in this case a special relevance in those formulas α which are true at those moments t such that, for at least one $h \in \{h_1, h_3, h_5, h_8, h_9\}$, $t \in h$ and $t \in i$, where i is the instant to which t_5 belongs, namely i_3 (they are t_4, t_5 and t_6): they are the ones which are always true considering the success of the plan the agent is carrying out at moment t_5.

Can we then associate these formulas with the states of affairs a brings about at t_5 as part of the plan he is carrying out at t_5? In other words, supposing that α is one of those formulas, can we say that $a \blacktriangleright \alpha$ is true at t_5? Not yet, and the reason for this is twofold. First, among those α's, there are surely those which are true no matter what happens, that is to say, those formulas whose truth is inevitable or historically necessary. Therefore, in order to give an account of the meaning we wish to attribute to $\blacktriangleright$ we have to get rid of these inevitable formulas in our semantic analysis. Second, it is not enough to consider the histories which are compatible with the successful carrying out of the plan. Even though this accounts for the success aspect of agency, it does not account for the fact that α might be true by something else than the a's action; from the point of view of a's agency, α's truth might be an accident.

This second point is sorted out by considering histories that are not only compatible with the successful execution of the plan, but 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. As we have said, moments of time belonging to the same history are temporally and causally related. As one might expect, this causal relation is due also to the actions the agents do at those moments. The idea then is that the moments posterior to moment t which belong to some plan-history of agent a at t are under the control of or are responsive to, as Chellas (1992) would put it, the planning actions which a does at t. Trivially enough, these plan-histories are 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 point is solved, first, by using another modal operator: $\rhd$. If a is an agent and α is a formula then $a \rhd \alpha$ is also a formula. $\rhd$ shall be evaluated with the help of the plan-histories: $a \rhd \alpha$ is true at t iff for all moments of time t' which are co-instantaneous with t and belong to at least one of the plan-histories of a at t, α is true at t'. $a \rhd \alpha$ shall be read as "α is always the case in the successful execution of

the plan the agent is carrying out." However, $a \rhd \alpha$ still does not consider the non-inevitability aspect required by agency. We shall therefore introduce $\blacktriangleright$ as an derived notion: $a \blacktriangleright \alpha =_{\mathrm{df}} a \rhd \alpha \wedge \neg\Box\alpha$, where $\Box$ is the (weak) modality of historical necessity we have spoken about above. In other words, we say that agent a brings it about that α as part of the plan he is carrying out iff α is always the case in the successful execution of the plan the agent is carrying out and α is not historically necessary. In this way, if $a \blacktriangleright \alpha$ is true at moment t we guarantee that a brings it about that α (as part of the plan he is carrying out) and that α is neither inevitable (something which would be true no matter what) nor accidental (something not inevitable, but true due to something else than a's actions.)[15]

4. A Logic of Planning Agency

In addition to $\wedge$, $\neg$ and $\top$ (which shall be used with their usual meanings), the language of the *logic of planning agency* has as primitive logical symbols the modal operators $\rhd$ and $\Box$. If a is an agent and α is a formula, $a \rhd \alpha$ is also a formula; $a \rhd \alpha$ means that α is always the case in the successful execution of the plan the agent a is carrying out. $\Box$ is the modality of inevitability we have mentioned above. However, it is not the stronger notion of historical necessity, but the weaker one defined through the notion of alternative moments of time. The reason for that shall be clear below when we explain the semantic postulates of our logic.

As we have said, $\blacktriangleright$ shall be introduced as an abbreviation from $\rhd$ and $\Box$: $a \blacktriangleright \alpha =_{\mathrm{df}} a \rhd \alpha \wedge \neg\Box\alpha$. Besides $\blacktriangleright$, we also define a modality of historical possibility: $\Diamond\alpha =_{\mathrm{df}} \neg\Box\neg\alpha$. $\vee$, $\rightarrow$ and $\bot$ 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 $\rhd$:

Axioms for $\Box$

N1. $\Box(\alpha \rightarrow \beta) \rightarrow (\Box\alpha \rightarrow \Box\beta)$
N2. $\Box\alpha \rightarrow \alpha$
N3. $\Diamond\alpha \rightarrow \Box\Diamond\alpha$

Axioms for $\rhd$

P1. $a \rhd \alpha \rightarrow \alpha$
P2. $\Box\alpha \rightarrow a \rhd \alpha$
P3. $a \rhd \alpha \rightarrow \Diamond\alpha$
P4. $a \rhd (\alpha \rightarrow \beta) \rightarrow (a \rhd \alpha \rightarrow \alpha \rhd \beta)$
P5. $a \rhd \alpha \rightarrow \Box(a \rhd \alpha)$

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 $\langle T, \prec, A, P, \| \|, V \rangle$ where

(i) T is a non-empty set of moments of time;

(ii) $\prec$ 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 $t_1 \prec t_3$ and $t_2 \prec t_3$, then either $t_1 = t_2$ or $t_1 \prec t_2$ or $t_2 \prec t_1$ (*no backward ramification condition*);

A *history* is a maximal chain defined on $\langle T, \prec \rangle$. We call H the set of all histories. Let h, $h' \in H$ and $t \in T$. h and h' *share the same past in* t (in symbols: $h \cong_t h'$) iff, for all $t' \prec t$, $t' \in h$ and $t' \in h'$. The set of *alternatives* $\mathscr{T}$ is a partition of T such that $t, t' \in T$ belong to the same partition (which might be referred either as $\mathscr{T}_t$ or $\mathscr{T}_{t'}$) iff, for all h, $h' \in H$, $h \cong_t h'$ iff $h \cong_{t'} h'$.

(iii) A is a non-empty set of agents.

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 the agent does in the execution of the plan he is carrying out at that moment. We shall refer to the histories which are members of a plan-history as *plan-histories*.

(iv) $P : A \times T \to \mathscr{P}(H)$ 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 $h \in P(a, t)$, $t \in h$; (*success condition*)
2. $P(a, t) \neq \emptyset$; (*non-contradiction condition*)
3. If $h' \in P(a, t)$, then for all $h \in H$ such that $t \in h$, $h \cong_t h'$; (*historical relevance condition*)
4. If $\mathscr{T}_t = \mathscr{T}_{t'}$, then $P(a, t) = P(a, t')$. (*alternative plans condition*)

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

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

In the logic of branching time, since the logical language has future modalities, the evaluation relation $\Vdash$ (which, given a model M, says whether a formula α is true at M) has as parameter, besides a moment of time, also a history. Therefore, our relation $\Vdash$ shall have, besides moments of time, also histories as parameters. The definition of $\Vdash$ for $\wedge$, $\neg$ and $\top$ is done in the usual way. Below we have the truth-definitions of $\Box$ and $\rhd$:

$M \Vdash_{t,h} \Box\alpha$ iff for all $t' \in T$ such that $t' \in \mathscr{T}_t$ and all h' such that $t' \in h'$,
$M \Vdash_{t',h'} \alpha$

$M \Vdash_{t,h} a \rhd \alpha$ iff for all $h' \in P(\|a\|, t)$ and $t' \in h'$ such that $\mathscr{T}_t = \mathscr{T}_{t'}$, $M \Vdash_{t',h'} \alpha$.

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 on the conditions we have imposed on function P and answer why we have used the weak modality of historical necessity instead of the stronger one. 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. It is this condition which guarantees the truth of P1. In order to obtain the other reading for $\blacktriangleright$ we have mentioned at the end of Section 2 (something like "in order to carry out his plan, agent a must bring it about that α") we would have just to drop the success condition, on the semantic level, and axiom P1, on the syntactic side. This of course would give rise to a different logic of planning agency.

The *non-contradiction condition* says simply that set of plan-histories cannot be empty. If it could, by vacuity we would have formulas such as $a \rhd \bot$ as satisfiable. A consequence of this condition is that $\neg a \rhd \bot$ is a tautology.

In order to explain and justify the two other conditions, let us elaborate a bit further on the notion of planning agent using our branching time semantic framework. First of all, as we know, 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, from the point of view of that moment of time, be considered any more as possible courses of the history of our world. 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. Neither h_7, nor h_8, ..., nor h_{12} are possible from the standpoint of t_2; nor from the standpoint of any moment posterior to t_2.

Second, the plan agent a is carrying out at some moment, say t_{11}, was obviously conceived or created at a moment t' anterior to t_{11}; in symbols: $t' \prec t_{11}$. Now suppose that the plan a is carrying out at t_{11} was formulated at t_2. Given what we have said in the above paragraph, it does not make sense to include neither h_7, nor h_8, ..., nor h_{12} in $P(a, t_{11})$, for these histories are not possible any more from the point of view of t_2.

Third, there is a moment of time between t_2 and t_{11}, namely t_5. Adopting the quite reasonable assumption that an ideally rational agent reviews his plan always when he has opportunity to do so, when t_5 becomes the present moment it does 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. So, if a does review his

plan at t_5, these histories trivially shall not belong to the corresponding new plan-history. To sum up then, we can assume that the plan an agent a is carrying out at moment t was formulated or at least reviewed at the moment immediately anterior to t, which in its turn impose some restrictions on the histories that can belong to $P(a,t)$.

Going then to our remaining conditions, the *historical relevance 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_1 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 the formulation or reviewing of the plan, that is, the moment immediately anterior to t, can belong to $P(a,t)$.

The *alternative plans condition* says that at all alternative moments, a should carry out the same plan. The reason for this 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 or reviewed, 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, other implication of the above considerations is 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 $\blacktriangleright$ shall be those which are inevitable at those alternative moments. Therefore, the weak historical necessity is enough for our purposes.

5. Drawbacks and Further Developments

There is one common criticism to the sort of approach we are following here which deserves special attention. It is mentioned in Elgesem 1997 and involves the requirement found in Pörn 1970, Belnap 1988, 1992, Chellas 1992 and Vanderveken 2005 that no agent can act so as to bring about an inevitable fact. As we have seen, in our logic this requirement is formalized by defining our distinguished planning agency operator $a \blacktriangleright \alpha$ as $a \rhd \alpha \wedge \neg\Box\alpha$.

In order to support his claim, Elgesem (1997) gives the following example.

> Suppose my one-year-old boy is in the process of learning to eat by himself. Sometimes he succeeds in getting the food into his mouth with the spoon,

> and sometimes not. Suppose he succeeds at some point during the meal, i.e. he brings it about that he has food in his mouth. During the whole of this meal, I am watching him to make sure that he gets fed. So if he does not succeed in getting the food into his mouth, I put the food into his mouth anyway. In this situation, it seems to be the case that there is no relevant alternative where it is not true that he gets food in his mouth. Now, in the case where he hits his mouth with the spoon, it must be right to say that he brings it about that he has food in his mouth. This is the case despite the existence of a reliable back-up system which guarantees that the goal is satisfied in any case. (Elgesem 1997, p. 10)

In other words, even though it is in some sense necessary that the food gets into the child's mouth, the cases where he himself hits his mouth with the spoon are genuine cases where he brings it about that he has food in his mouth.

According to Elgesem, the traditional approach conflates two ideas: avoidability and the agent's activity being instrumental in the production of the result. The real point of requiring that no agent can bring it about inevitable states of affairs is in fact that the agent be instrumental in the production of the result of his action. What is important is that the corresponding state of affairs be the case due to *his* efforts, be it an inevitable state of affairs or not.

While we think this is a serious criticism regarding agency in general, we think it can be properly replied in the case of planning agency as follows. First of all, we are dealing with what we have called ideally rational agents. This makes it reasonable to suppose that these agents know enough about the states of affairs they are concerned about to know whether they are historically necessary in the weak sense or not. Second, as we have seen, it is reasonable to suppose that an ideally rational agent reviews his plans at every moment of time. As a consequence of these two points, we have the equally reasonable assumption that an ideally rational agent does not consider in his plans inevitable states of affairs. Roughly put, the ideal character of such an agent—which we suppose involves some sort of economy principle—prevents him from considering as part of his plan a state of affairs which is the case no matter what he does.[16]

Despite of this, the logical system introduced in this paper has some serious limitations. First, we allow for the realization of at maximum one plan per agent at a time; there is no way an agent might carry out two plans simultaneously. Second, even though our logic is a multi-agent system, there is no consideration whatsoever of the agents mutually cooperating to achieve a common goal; or mutually cooperating to each one achieve his respective goal. We could of course have things like $a \blacktriangleright (b \blacktriangleright \alpha)$, which could be read as "as part of the *plan* he is carrying out, a brings it about that agent b brings it about that α as part of the *plan* b is carrying out." This however would seem much more like persuasion than cooperation.

Third, even though Postulates 2, 3 and 5 establishes a plan as something structured, the system we presented here deals only with one level of this structure. As we have said, our analysis takes into consideration only the states of affairs referred to in Postulate 3, that is to say, the states of affairs that characterize the way the agent plans to achieve his goal. There is no way, for instance, to represent the goal of the plan. This, we might concede, is a very serious drawback. Second, there is no way either to represent the states of affairs mentioned in Postulate 5, that is to say, the states of affairs which characterize the way the agent plans to achieve each one of the sub-goals mentioned in Postulate 3; nor is there a way to represent the states of affairs which characterize the way the agent plans to achieve each one of these 'sub-sub-goals', and so on and so forth.

As an attempt towards filling in this gap as well as solving the first mentioned drawback, the logical language of such a fully structured logic of planning agency could be as follows. First of all, modal operator $\blacktriangleright$ would be indexed by a formula representing a specific goal. For example, $a \blacktriangleright_\beta a$ would mean something like "agent a brings it about that α as part of the *plan* he is carrying out having as goal the state of affairs expressed by β." If, for instance, $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the formulas representing the states of affairs that characterize the way the agent plans to achieve his goal expressed by β, then the whole plan would be represented by formulas $a \blacktriangleright_\beta \alpha_1, a \blacktriangleright_\beta \alpha_2, \ldots, a \blacktriangleright_\beta \alpha_n$. If in its turn, $\varphi_1, \varphi_2, \ldots, \varphi_m$ are the formulas representing the states of affairs that characterize the way a plans to achieve the goal expressed by α_1, for instance, then this sub-plan, as we could call it, would be represented by $a \blacktriangleright_{\alpha_1} \varphi_1, a \blacktriangleright_{\alpha_1} \varphi_2, \ldots, a \blacktriangleright_{\alpha_1} \varphi_m$.

Finally, in order to properly characterize our plans, especially their structured aspect, we need to represent, inside our logical language, future states of affairs. Therefore we need future modalities. We postpone to a future work the development of a logic which combines our modal planning agency operator with future modalities, as well as the development of a fully structured logic of planning agency as sketched in the above paragraph.

6. Conclusion

We have presented in this paper a first attempt to provide a formal account of planning agency. We followed the standard approach in philosophical modal logic of action of representing agentive sentences as a relation between agents and state of affairs. Besides providing a logical system composed by language, calculus and semantics, which was built upon the semantic framework of the theory of branching time, we also provided a sketch of a theory of planning agency which might serve as philosophical guideline for further developments in the logic of planning agency.[17]

References

Allen, J. & Hendler, J. 1990. *Readings in Planning*. San Mateo: Morgan Kaufmann.

Anscombe, E. 1957. *Intention*. Oxford: Basil Blackwell.

Belnap, N. & Green, M. 1994. Indeterminism and the Thin Red Line. In J. Tomberlin (ed.) *Philosophical Perspectives*, vol. 8, Atascadero: Ridgeview, pp. 365–88.

Belnap, N. & Perloff, M. 1988. Seeing to it that: a canonical form for agentives. *Theoria* **54**: 175–99.

———. 1992. The way of the agent. *Studia Logica* **51**: 463–84.

Belnap, N.; Perloff, M.; Xu, M. 2001. *Facing the future: Agents and choices in our indeterminist World*. Oxford: Oxford University Press.

Bratman, M. 1987. *Intentions, Plans, and Practical Reasoning*. Cambridge, MA: Harvard University Press.

Carmo, J. & Pacheco, O. 2001. Deontic and action logics for organized collective agency modeled through institutionalized agents and roles. *Fund. Inform.* **48**: 129–63.

Chellas, B. F. 1992, Time and Modality in the Logic of Agency. *Studia Logica* **51**: 485–518.

Davidson, D. 1980. *Essays on Actions and Events*. Oxford: Oxford University Press.

Elgesem, D. 1997. The modal logic of agency. *Nordic J. Philos. Logic* **2**: 1–46.

Gelati, J.; Governatori, G.; Rotolo, A.; Sartor, G. 2002. Declarative power, representation, and mandate: A formal analysis. In: T. Bench-Capon, A. Daskalopulu and R. Winkels (eds.) *Legal Knowledge and Information Systems*. Amsterdam: IOS Press, pp. 41–52.

George W. 2007. Action. In E. N. Zalta (ed.) *The Stanford Encyclopedia of Philosophy*. URL: http://plato.stanford.edu/entries/action.

Goldman, A. 1970. *A Theory of Human Action*. Englewood Cliffs, N. J: Prentice-Hall.

Hilpinen, R. 1997. On action and agency. In: E. Ejerhed and S. Lindström (eds.) *Logic, Action and Cognition: Essays in Philosophical Logic*. Dordrecht: Kluwer Academic Publishers, pp. 3–7.

Horty, J. F. & Belnap, N. 1995. The deliberative stit: A study of action, omission, ability and obligation. *J. Philos. Logic* **24**: 583–644.

Jones, A. J. I. 2003. A logical framework. In J. Pitt (ed.) *Open Agent Societies: Normative Specifications in Multi-Agent Systems*. Chichester: Wiley, Chapter 3.

Pollack, M. 1992. The uses of plan. *Artificial Intelligence* **57**: 43–68.

Pörn, I. 1970. *The Logic of Power*. Oxford: Blackwell P.

———. 1977. *Action Theory and Social Science: Some Formal Models*. Dordrecht: Reidel.

Segerberg, K. 1989. Bringing it about. *J. Philos. Logic* **18**: 327–47.

———. 1992. Getting started: Beginnings in the logic of action. *Studia Logica* **51**: 347–58.

Sergot, M. & Richards, F. 2001. On the representation of action and agency in the theory of normative positions. *Fund. Inform.* **48**: 273–93.

Thomason, R. 1970. Indeterminist time and truth-value gaps. *Theoria* **36**: 264–81.

———. 1984. Combinations of Tense and Modality. In: D. Gabbay and F. Guenthner (eds.) *Handbook of Philosophical Logic: Extensions of Classical Logic*. Dordrecht: Reidel, pp. 135-65.

Vanderveken, D. 2005. Attempt, success and action generation: a logical study of intentional action. In D. Vanderveken (ed.) *Logic, thought and action*. Dordrecht: Springer, pp. 315–42.

RICARDO SOUSA SILVESTRE
Departamento de Filosofia
Universidade Federal de Campina Grande
R. Aprígio Veloso 882
58429-900 Campina Grande, PB
BRAZIL
ricardoss@ufcg.edu.br

Resumo. O propósito desse artigo é fornecer um tratamento formal para um tipo de ação até o momento negligenciada nas lógicas modais filosóficas da ação: ação em plano. Ao fazer isso nós seguimos a abordagem padrão nas lógicas modais da ação exemplificados pelos trabalhos de Belnap, Chellas and Pörn. Como nós acreditamos que existe uma relação forte entre plano, tempo e indeterminismo, nós usamos a teoria do tempo ramificado para investigar as características básicas da ação em plano. Além de introduzir uma semântica do tempo ramificado, nós também apresentamos um cálculo correto e completo com relação a essa semântica.

Palavras-chave: Lógica modal da ação, ação em plano, teoria do tempo ramificado.

Notes

[1] This example is adapted from Davidson 1980, essay 4.
[2] This example is taken from Wilson 2007.
[3] Examples of such approach are Belnap & Perloff 1988 and 1992, Chellas 1992, Vanderveken 2005, Elgesem 1997, Horty & Belnap 1995, Pörn 1970 and 1977, Segerberg 1989 and 1992.
[4] This example is given in Pörn 1970.
[5] As for instance in Carmo & Pacheco 2001, Jones 2003 and Gelati et al. 2002.
[6] See Pollack 1992 and Allen 1990, for instance.
[7] See Thomason 1970 and Belnap & Green 1994.
[8] Among the modal logics of action which also use, in their semantics, the theory of branching time we can mention Belnap & Perloff 1988, Belnap & Perloff 1992, Chellas 1992 and Vanderveken 2005.
[9] We are here of course making use of the basic assumption that states of affairs can be characterized or expressed by propositions, statements, formulas and the like.
[10] We have decided to use the expression "bring it about that" instead of "see 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.
[11] Belnap & Perloff 1988 and 1992.
[12] Here we are adopting an indeterminist view of actions which we assume is consensual enough to require argumentation.
[13] See Thomason 1970 and Belnap & Green 1994.

[14] See Vanderveken 2005 and Thomason 1984.
[15] This would reply the criticism made by Elgesem (1997) about Belnap's (1988, 1992) and Pörn's logic (1970), which in our view is due to a misunderstanding about the theory of branching time.
[16] Notice that the ideal character we are assuming for our agent does not require logical omniscience; in particular it does not require knowledge of historically necessary states of affairs in the strong sense.
[17] Work partially supported by CNPq (National Counsel of Technological and Scientific Development of Brazil), public notice MCT/CNPq Nº 03/2009. A previous version of this paper has appeared in CLE e-Prints (Online), v. 8.