KRISTER SEGERBERG

Getting Started: Beginnings in the Logic of Action

Abstract. A history of the logic of action is outlined, beginning with St Anselm. Five modern authors are discussed in some detail: von Wright, Fitch, Kanger, Chellas and Pratt.

Is there philosophy of action? The answer, uncontroversial, is yes. Philosophers have always been interested in action, primarily in connection with ethics and with metaphysics of the will. But for a few decades action has been studied for its own sake. Elizabeth Anscombe's book *Intention*, published in 1957, is sometimes seen as the work which inaugurated philosophy of action as an autonomous field of study, and efforts by authors like Donald Davidson, Roderick Chisholm, Georg Henrik von Wright and Alvin Goldman have established it as a viable discipline which is here to stay. Among the questions action philosophers wish to answer are, What is an action? When are two actions the same? What is the relationship between intention, intending and intentional action? How can one describe an action, explain an action, understand an action?

Is there logic of action? In other words, is there a discipline which studies modellings of action which involve formal logic? Certainly the term is around : as early as 1963 von Wright used it in his book *Norm and action* [20]. At the same time, it has to be admitted that the logic of action is not developed to nearly the same degree as — or with anything like the success of — the philosophy of action. This is perhaps because the time for logic in this area is only slowly becoming ripe. Perhaps it is an instance of a universal phenomenon : philosophizing prepares the way for rigorous theorizing, and this is why at the inception of a new discipline more energy goes into the former. Before the gold can be mined (the task of philosophical logic) prospectors must explore the terrain (the task of philosophy). If this view is correct, then there is no difference of purview between philosophy of action and logic of action : the questions are the same, what differs is the technique.

The difference in technique also explains why to date the results of the logic of action, insofar as there are any, have been relatively unimpressive or at least very specialized. Working with logical techniques pushes the requirement of rigour so high that pressures of complexity enforce a very narrow

Studia Logica **51**: 347–378, 1992.

focus. This is a phenomenon well known form other branches of philosophical endeavour. Nontechnical philosophers are naturalists who describe what they see with the naked eye. Logicians examine nature through their microscopes and X–ray cameras : what they see is also an aspect of nature, but a different one.

Thus the logic of action is a subject which is slow in forming. The two central notions, which it is the first priority of the young discipline to analyze, are those of *agency* and *ability*. This essay will try to trace some efforts to do so. No claim to completeness is made. On the contrary, a number of topics and logicians omitted here might well have been included by a different author. In the tradition of philosophical logic the greatest omission is perhaps that of Hector–Neri Castañeda's work : a complete account would have to describe his extensive and fruitful contribution, which, however, is difficult to survey. Also not included here is the work done in computer science which is perhaps where a really comprehensive, formal theory of action will first appear.

1. Prehistory : St Anselm

As we shell see, the logic of action may be said to have many fathers. However, there is but one patron saint, St Anselm (1033 – 1109). The reader need only peruse pp. 121 – 133 of Paul Desmond Henry's *The logic of Saint Anselm* ([7]) to see that Anselm has a powerful claim to this title. What he will find there is an account of Anselm's investigations into the formal properties of the verb *facere* (to do). It is of course very difficult, especially for those of us who are not specialists in mediaeval logic, rightly to appreciate what Anselm is trying to do. For example, Henry summarizes a passage in the words, " 'To do' is to verbs as pronouns are to names : pronouns are name–variables : 'to do' can act as a predicate–variable" ([7] p. 123). This, it is easy to feel, sounds like a brilliant insight. But it is not easy to express it in the language of any modern calculus, at least not without adopting heavy theoretical assumptions which may be foreign to Anselm.

Let us look briefly at another example. Anselm says that *facere esse* and *facere non esse* are affirmative forms and in fact contraries of one another, while *non facere esse* and *non facere non esse* are negative forms and in fact the negations of the former ([7] p. 124). Presumably two forms are contrary if they cannot be true simultaneously, and subcontrary if they cannot be false simultaneously. Moreover, a form and its negation are presumably contradictory. Does *facere esse* imply *non facere non esse* (as we might wish to say), and does *facere non esse* imply *non facere esse*? If so, we would have a Square of Opposition for *facere* as in Fig. 1.1 (which, it should be emphasized, is not actually found in Anselm).

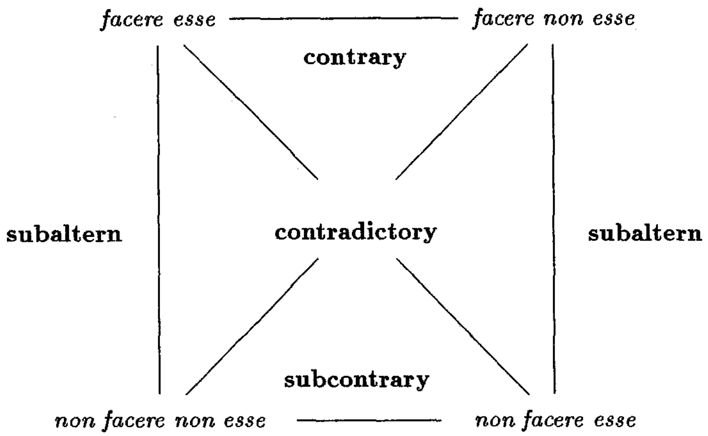

Fig. 1.1

In his translation, Henry uses the following terminology:

facere esse	x does so that p,
facere non esse	x does so that not–p,
non facere esse	x does not so that p,
non facere non esse	x does not so that not–p,

Here 'p' is supposed to be a clause describing a state of affairs, and 'not–p' is short for 'it is not the case that p'; an example of a value of p is 'N is dead'. This translation of course already makes important assumptions, for example, that *facere* is to be understood as a propositional concept. It has the effect of suggesting that Anselm's ideas can be accommodated within some propositional logic by the adoption of a nonclassical propositional operator which we might write as 'does' (leaving the fixed agent x out of the formalism). Using this idea we would obtain the following formalization of Anselm's forms :

facere esse	does p,
facere non esse	does $\neg p$,
non facere esse	$\neg$does p,
non facere non esse	$\neg$ does $\neg p$,

In these terms, the Square of Opposition above becomes as in Fig. 1.2. It is readily seen that in order to guarantee the intended relationship in this square it is enough to postulate that the does–operator satisfies the following axiom :

$$\neg(\text{does } p \wedge \text{does } \neg p).$$

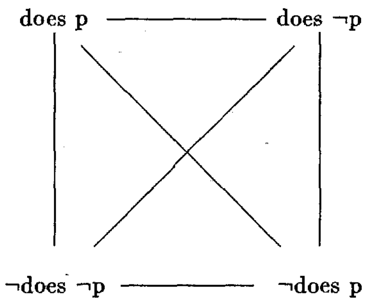

Fig. 1.2

Starting on this path, one would then have to consider Anselm's remark that the affirmative form *facere esse* is sometimes used instead of the negative form *non facere non esse*, and similarly the affirmative form *facere non esse* instead of the negative form *non facere esse* ([7], p.125). The Square of Opposition would collapse if this were always permissible. On the other hand, if it is not always permissible, one would like to know when it is. It is natural to think that a necessary condition for being able to use an affirmative form in place of a negative one is that the agent be active (that he do something). It would be sufficient if he were active with respect to p, for in classical logic,

$$\text{does } p \lor \text{does } \neg p \vdash \neg\text{does } \neg p \supset \text{does } p.$$

However, some stronger grounds would have to be provided in order to account for the examples given by Anselm, such as being able to use 'x does so that there are evils' where the more proper expression would be 'x does not so that evils are not' ([7], p. 125).

An interesting remark is that " 'x does so that p' has the proper sense, 'x does so that p, which was not the case, becomes the case' " and " 'x does so that not–q' has the proper sense, 'x does so that q, which was the case, ceases to be the case' " ([7], p.126). This, as we shall see, is similar to ideas developed by G. H. von Wright almost nine hundred years later.

Thus even a presentation as cursory as the one given here reveals the impressive richness of Anselm's thought. If we could read him correctly, it is quite possible that this would not be of only historical interest but that it would, in Professor Henry's words, "turn out to reveal a rich crop of seminal ideas still relevant in our own day" ([7], p.251). Unfortunately, it is very difficult to approach a thinker of another age or tradition except from the vantage point of one's own age or tradition. Certainly one function of the great classical figures, both scholars and artists, is that later generations can use them as mirrors, seeing themselves, perhaps in an unexpected light.

Thus as we develop the logic of action further, we may well come to realize that insights gained in our time were already known to Anselm. But they are difficult to identify without theory. Let us hope, therefore, that before long a mediaevalist will read up on current logic of action and compare it with that of its patron saint.

2. Foundations : von Wright

Historians are fond of indentifying the fathers of this and that. Insofar as there is one father of the logic of action — insofar as there is a logic of action — that father must be Georg Henrik von Wright. As we shell see, paternity cases could be made also for Frederic B. Fitch and Stig Kanger, for both made distinguished pioneering contributions to the logic of action. However, Fitch only published one paper which, furthermore, had no impact on future development. Kanger never gave much attention to action *per se*, and his original logic of action is incidental, essentially an auxiliary for other purposes. von Wright, on the other hand, has produced an extended investigation into the nature of action. Among his most important works are *Norm and action* ([20]) and *Varieties of goodness* ([21]) from 1963, *An essay in deontic logic and the general theory of action* ([23]) from 1968, the celebrated *Explanation and understanding* ([22]) from 1971 and *Causality and determinism* ([24]) from 1974. More then many other authors von Wright keeps returning to a subject, revising old ideas, always refusing to regard his work as finished. This makes it difficult to give a brief summary of his contribution to the logic of action. Here one particular theme, running through his work, will be highlighted. By necessity, our discussion will omit a great number of other topics.

von Wright's conception of action is expressed in passages like these:

> It would not be right, I think, to call acts a kind or species of events. An act *is* not a change in the world. But many acts may quite appropriately be described as the bringing about or *effecting* ('at will') of a change. To act is, in a sense, to *interfere* with 'the course of nature'.([20], p. 36)

> An act is the bringing about or production at will of a change in the world. ([21], p. 115)

> To act is to interfere with the course of the world, thereby making true something which would not otherwise (i.e. had it not been for this interference) come to be true of the world at that stage of its history. ([24], p.39)

One might think, then, of an act as an event brought about by an agent. Moreover, the realization of a particular event may be identified with an ordered pair of two states of affairs, the initial state and the end–state ([20], p. 27f.). For example, an individual act of opening a certain window may be thought of as an ordered pair $\langle e, i \rangle$ where e is individual event, i is a certain agent, and e is brought about by i; and the event may be thought of as the ordered pair $\langle x, y \rangle$, where x and y are states of affairs and the window is closed in x and open in y.

Opening a window is an example of a kind of action which, in some sense, is particularly simple. Other examples used by vow Wright are: unlocking a door, pulling a trigger, heating a hut. More complex actions are presumably built in some way from simpler actions (and it is a major task for action theory to say how). Thus it makes good sense to begin by analyzing simple actions, and this is what von Wright does. Here we shall focus on the theory he offers in the first part of the chapter "Action logic as a basis for deontic logic" in his book *Practical reason* [26], a collection of papers of his on action and norms. We shall briefly explore the possibilities of providing a formal semantics for this theory true to von Wright's own informal intuitions. Our exploration will take place in two steps.

With respect to a given state of affairs A, an agent might produce it, sustain it, destroy it, or suppress it. Production and destruction are positive, in a certain sense : they bring change. Sustaining and suppression, on the other hand, are negative : they prevent change. von Wright uses two new operators $\mathcal{B}$ and $\mathcal{S}$ to formalize these notions. With von Wright let us read for $\mathcal{B}A$:

"the agent produces the state that A", or

"the agent brings it about that A", or

"the agent makes it so that A.

Similarly, for $\mathcal{S}A$, let us read :

"the agent sustains the state that A", or

"the agent prevents the state that A from vanishing".

In these terms we may express the other two notions as well: for $\mathcal{B}\neg A$ we may read :

"the agent destroys the state that A",

and for $\mathcal{S}\neg A$:

"the agent suppresses the state that A".

Suppose that something happens on some occasion (the initial point). When would we say that the agent had produced A? Obviously, necessary conditions include that A did not obtain at the initial point, that A did obtain at the end–point, and that the change is due to the agent. Similarly, *mutatis mutandis*, with sustaining, destruction and suppression.

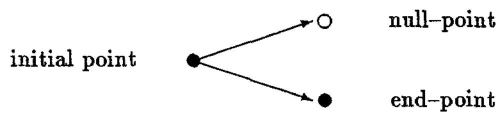

Fig. 2.1

Here it is convenient to use a graphic representation related to (but marginally different from) one occasionally used by von Wright himself (see [23], p. 51f.). In the previous paragraphs there are three "points" that are at issue: the initial point, the end–point and the point which would have resulted if the course of nature had not been tampered with (von Wright never names the last point, but a name might be useful — let us call it the *null–point*). Think of these three points as in Fig. 2.1. Using that figure as a template, we can readily represent the four concepts mentioned above as in Fig. 2.2.

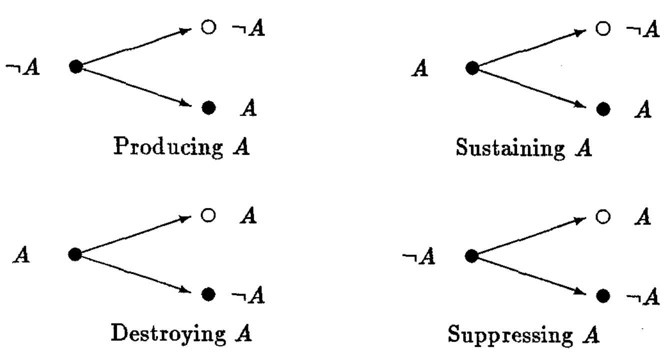

Fig. 2.2

If we were to try to regiment these ideas — as a first step towards a formal semantics — we might proceed as follows. Let U be a given space of *points*. They may be thought of as "total states of affairs"; this seems to be how von Wright usually thinks of his "occasions". Usually, but not always :

> In the cases where the state either comes to be or ceases to be it is presupposed that the "occasion" has a certain duration, beginning with a "phase" when the state of affairs is absent (present) and ending in a "phase" when the state is present (absent). ([26], p. 174)

In this passage it is "phase" rather then "occasion" that is rendered by our "point".

Let $F = \{F_x : x \in U\}$ be a family of functions which we shall call *change functions* in U. By means of these functions we shall try to represent (the events corresponding to) the agent's action. As a first abstraction we assume that there are only two *options* (courses of action) open to the agent : to do something or to do nothing. Let us represent these alternatives by the natural numbers 1 and 0 respectively. Now, if x is the initial point of the agent's action and if $s \in \{0,1\}$, then $F_x(s)$ is meant to be its end–point. Thus, writing $2 = \{0,1\}$, we make the formal requirement that F_x be a function $2 \to U$. In order words, $F_x(1)$ gives us the point at which the actual action of the agent lands us, while $F_x(0)$ gives us the null–point, the point at which we would have landed, had it not been for the agent's action. This is clearly in accord with von Wright's view of action as interference with the course of nature.

Truth–conditions are now straightforward. Assume that we have a valuation V, assigning to each propositional letter a subset of U. Then we may define, in the usual manner, a truth–value with respect to a point for each Boolean formula A; let us write $\models_x A$ if A is true at x and $\not\models_x A$ otherwise. Thus if P is a propositional letter we have

$$\models_x P \text{ iff } x \in V(P).$$

For complex Boolean formulae we have conditions

$$\models_x \neg A \text{ iff } \not\models_x A, etc.$$

Let F be a family of change functions in U. As truth–conditions for the more complicated formulae we adopt the following. Suppose $s \in 2$. Then,

$$\begin{aligned} s \models_x BA &\text{ iff } s = 1 \text{ and } \not\models_x A \text{ and } \not\models_{x'} A \text{ and } \models_{x''} A, \\ s \models_x SA &\text{ iff } s = 1 \text{ and } \models_x A \text{ and } \not\models_{x'} A \text{ and } \models_{x''} A, \end{aligned}$$

where $x' = F_x(0)$ and $x'' = F_x(1)$. These are the crucial conditions, and they are sufficient if we note that Boolean combinations of non–Boolean formulae are handled in the same way as Boolean combinations of Boolean ones. For convenience let us also add the condition that if A is Boolean, then, $\models_x A$ iff, for any $s \in 2$, $s \models_x A$.

We now have a concept of validity: A is *valid*, in this sense, if, for all $s \in 2$, $s \models_x A$, for all $\langle U, V \rangle$ and all F in U. This is not the concept von Wright has in mind, but it goes some way towards capturing it. For example, $\mathcal{B}p \supset p$ is not valid (cf. [26], p. 195f.). On the other hand, $\mathcal{B}p \supset \neg p$ is, which seems all for the good if the horse–shoe is read as material implication.

However, matters are more complicated than this. For one thing, there may be several actions open to the agent. For another, there are often more agents than one to consider. These aspects come together in the context of omission, another important concept of logic of action, and one notoriously difficult to analyse.

Embarking on the second step of our exploration, we now proceed to improve on the restricted modelling just given. Let us assume that there are n agents (they might be identified with the natural numbers $0, 1, \ldots, n-1$). Each agent has some options $0, 1, \ldots, m$ or $0, 1, 2, \ldots$ — finitely or at most denumerably many — where 0 represents passivity ("doing nothing" — always an option). The totality of what the agents do on a certain occasion may be represented by an n–ary vector $(s_0, \ldots, s_{n-1})$ where each s_i is an element of $\mathcal{N}$, the set of natural numbers. Let us write $\oslash$ for the null–vector $(0, \ldots, 0)$, which represents universal passivity ("everybody doing nothing"). A change function F_x now becomes a function $\mathcal{N}^n \to U$. As before, if x is the initial point, then $F_x(s)$ is the end–point of the event taking place if s is the vector representing the totality of what the agents do at x. Moreover, $F_x(\oslash)$ represents the end–point of the event taking place if at x all agents had remained passive and the course of Nature had not been interfered with. (Small technical point: Suppose that the agent i has only finitely many options $0, 1, \ldots, m$. Then we agree that, for all points x and vectors s, if $s_i > m$, then $F_x(s) = F_x(s')$, where s' is the vector just like s except that $s'_i = 0$.)

The question is whether this semantics is sufficiently subtle to formalize von Wright ideas. Let us go over what he actually says.

> The productive action Bp can be performed by a given agent on a given occasion only on condition that the state of affairs that p is absent and remains absent unless some agent interferes and produces it.([26], p. 170)

> ...a on o omits ...to produce the state that p if o affords an opportunity for producing this state but a does not produce it. ([26], p. 171)

> That a on o omits to produce a certain state of affairs presupposes, we have said, an opportunity for producing it. Then it may happen that another agent b "seizes the opportunity" and produces the state on the occasion in question. ...The omission logically presupposes that the state *is* absent but not that it stays absent. ([26], p. 172)

von Wright draws a distinction between an opportunity *simpliciter* and an opportunity *for a specific agent.* The conditions laid down in the first of the three quotations are meant to define the former concept. In our formal semantics they would be rendered as follows : occasion x *offers an opportunity simpliciter* for producing A if

$$\not\models_x A \text{ and } \not\models_y A \text{ and } \models_z A,$$

where $y = F_x(\varphi)$ and $z = F_x(t)$, for some t. But this concept is not enough for the purpose of giving truth–conditions for $\mathcal{B}_i A$ ("agent i brings it about that A") :

> On any given occasion, the agent either performs or omits to perform the action for which there is an opportunity. It is supposed that the opportunity qualifies as an opportunity *for that agent* ([26], p. 174; italics in the original).

It is clear that the concept of an opportunity for an agent is stronger than that of opportunity *simpliciter.* However, within our (still limited) formalism it does not seem possible to explicate the extra strength. One attempt would be this : x *offers an opportunity for i to produce A* if

$$\not\models_x A \quad \text{and } \not\models_{F(\varphi)} A \text{ and there is some } k \text{ such that, for any vector } s, \text{ if } s_i = k, \text{ then } \models_{F(s)} A.$$

This is quite a strong concept of opportunity. An example of a weaker concept is this :

$$\not\models_x A \quad \text{and } \not\models_{F(\varphi)} A \text{ and there are some vectors } s \text{ and } t \text{ such that } s \text{ and } t \text{ are identical except that } s_i \neq t_i \text{ and } \models_{F(s)} A \text{ but } \not\models_{F(t)} A.$$

One may think of still other conditions. None of them would seem to bring out exactly what von Wright has in mind. The difficulty is with causality: for production it is not enough that A becomes true, what is required is that A becomes true because of the agent's action. This difficulty is felt not only with regard to the concept of opportunity but also if one wants to give truth–conditions for $\mathcal{B}_i$. One natural candidate is this :

$$s \models_x \mathcal{B}_i A \quad \text{iff } x \text{ offers an opportunity for } i \text{ to produce } A \text{ and } \models_{F(s)} A.$$

But nothing in this definition precludes that A has become true quite independently of i's action.

The problem recurs in connexion with omission. von Wright introduces a new operator (which we will render by a bar) which operates on expressions of type $\mathcal{B}A$ and $\mathcal{S}A$. Thus for $\bar{\mathcal{B}}_i A$, read

"the agent i omits to bring it about that A",

and for $\bar{\mathcal{S}}_i A$, read

"the agent i omits to sustain the fact that A",

The quotations above suggest that, if i has an opportunity to produce A, then he nevertheless omits to do so if one of the following cases occurs :

(i) i does nothing,

(ii) i does something, but A does not result,

(iii) i does something, and A results, but that A results was not due to what i did.

Here we have the same difficulty as before when it comes to formalizing (iii). Thus, to give adequate truth–conditions for either $\bar{\mathcal{B}}_i$ or $\bar{\mathcal{S}}_i$ is not possible within our present formalism. It might be possible, though, in a technical sense to adapt our type of modelling to fit syntactical system of the kind indicated in von Wright's text. These systems would all have *modus ponens* and Replacement of Provable Equivalents as inference rules, and their axioms would include all tautologies and as well as all instances of the schemata

1. $\neg(oA \wedge o'A)$, if $o, o' \in \{\mathcal{B}_i, \bar{\mathcal{B}}_i, \mathcal{S}_i, \bar{\mathcal{S}}_i\}$ and $o \neq o'$,
2. $\neg(oA \wedge o'\neg A)$, if $o, o' \in \{\mathcal{B}_i, \bar{\mathcal{B}}_i, \mathcal{S}_i, \bar{\mathcal{S}}_i\}$
3. $\mathcal{B}_i A \vee \bar{\mathcal{B}}_i A \vee \mathcal{S}_i A \vee \bar{\mathcal{S}}_i A \vee \mathcal{B}_i \neg A \vee \bar{\mathcal{B}}_i \neg A \vee \mathcal{S}_i \neg A \vee \bar{\mathcal{S}}_i \neg A$

and perhaps also all instances of the schemata

4. $\mathcal{B}_i A \supset \neg A$,

5. $\mathcal{S}_i A \supset A$.

In addition there would be a set of reduction schemata :

> ...no set of rules for the distribution of the action operators over molecular compounds of potential action results can claim to be *the* "correct" rules. ...It seems, however, extremely natural, maybe even compelling, to take the view that the actions of producing and sustaining, and the corresponding omissions, should, when applied to molecular compounds of states of affairs, be dissolvable *some way or other* into molecular compounds of atomic or elementary cases of productive and sustaining actions and their omissions. ([26], p. 179f.)

One set of schemata — "the simplest, and perhaps also most natural distribution principles" ([26], p. 180) — is as follows :

6. $\mathcal{B}_i(A \wedge B) \equiv (\mathcal{B}_i A \wedge \mathcal{B}_i B)$,

7. $\mathcal{B}_i(A \vee B) \equiv (\mathcal{B}_i A \wedge \mathcal{B}_i B) \vee (\mathcal{B}_i A \wedge \bar{\mathcal{B}}_i B) \vee (\bar{\mathcal{B}}_i A \wedge \mathcal{B}_i B)$,

8. $\mathcal{S}_i(A \wedge B) \equiv (\mathcal{S}_i A \wedge \mathcal{S}_i B)$,

9. $\mathcal{S}_i(A \vee B) \equiv (\mathcal{S}_i A \wedge \mathcal{S}_i B) \vee (\mathcal{S}_i A \wedge \bar{\mathcal{S}}_i B) \vee (\bar{\mathcal{S}}_i A \wedge \mathcal{S}_i B)$.

Notice that it follows from (1) – (3) that the eight disjuncts of (3) are pairwise exlusive as well as jointly exhaustive.

The preceding system may be compared with an earlier calculus, devised by von Wright in his paper "Handlungslogik" (1974), reissued in 1980 under the title "Elemente der Handlungslogik"[25]: subsequent developments in English are found in [26]. Here von Wright is concerned to find a new way to represent the logical structure of an action proposition :

> As "atoms" of this logic of action we may regard variables; I will denote them by 'p', 'q', etc. In the logic of action these variables don't stand for arbitrary propositions such as "it is raining" or "Mannheim is a city in Germany", as they do in propositional logic; here they rather stand for verbs or verb phrases that denote human actions. Thus, e.g., "write", "read", "walk" or "steal" might stand for 'p' ([25], p. 23; my translation)

Verbs can be combined much as propositions can: one might read or write, read and write, or read and not write. Thus in von Wright's symbolism, '$[A]i$' may be read
"i A's" or perhaps
"i is A–ing",
if 'i' names an agent and 'A' is built from variables by connectives looking just like the ordinary propositional connectives. This is an interesting touch, and St Anselm naturally comes to mind. However — unfortunately, one is almost tempted to say — the interesting, verbal interpretation is combined with one that is more standard (cf. [26], p. 108):

"i makes it so that A".

The latter, propositional interpretation is of course related to the one we have already analysed above; in fact, it may be seen as a somewhat cruder version, one that does not distinguish production and sustaining, destruction and suppression. Omitting the reference to the agent i, let us choose the symbolism 'dA' with this reading in mind. We can describe von Wright's calculus in [25] as follows: the smallest logic, by definition closed under *modus ponens*, that contains all ordinary tautologies and also all instances of the following schemata :

1. $d\neg A \supset \neg dA$,
2. $d\neg\neg A \equiv dA$,
3. $d(A \wedge B) \equiv (dA \wedge dB)$,
4. $d(A \vee B) \equiv (dA \wedge d\neg B) \vee (d\neg A \wedge dB) \vee (d\neg A \wedge d\neg B)$.

One of von Wright's main interests in the paper is omission, and one pleasing feature of this calculus is that two different kinds of omission can be distinguished: it can be proved that '$d\neg A$' is a stronger condition than '$\neg dA$'. But each expresses a kind of non–doing.

For a later elaboration of this logic, see the author's "A topological logic of action" (1985).

3. A Syntactic Start : Fitch

1963 was an important year in the history of the logic of action as it saw the publication not only of von Wright's first works on action but also of F. B. Fitch's article "A logical analysis of some value concepts" [4]. While von Wright was to return to it again and again, Fitch never seems to have published anything else on this topic. Moreover, his paper has had no apparent impact on later work. This is surprising, for Fitch's approach in this paper is very fresh and quite remarkable for its day.

Fitch is interested in a number of intentional notions, some of which are 'doing', 'believing', 'knowing', 'desiring', 'ability to do', 'obligation to do' and 'value for'. Thus, in his highly compressed paper the scope is much wider than that of the logic of action. The following quotation explains what Fitch tries to do :

> First of all, we assume that striving, doing, believing and knowing are two–termed relations between an agent and a possible state of affairs. It is convenient to treat these possible states of affairs as propositions, so if I say that a strives for p, where p is a proposition, I mean that a strives to bring about or realize the (possible) state of affairs expressed by the proposition p. Similarly, if I say that a does p, where p is a proposition, I mean that a brings about the (possible) state of affairs expressed by the proposition p. We do not even have to restrict ourselves to *possible* states of affairs, because impossible states of affairs can be expressed by propositions just as well as can possible states of affairs. ... So we treat all these concepts as two–termed relations between an agent and a proposition. ([4], p.136)

Fitch's ideas are presented within the framework provided by the modal logic S4 supplemented with propositional quantifiers. That is to say, on this basis he is able to define notions of agency and ability in terms of the unary propositional operator 'striving for' (unary because the agent is kept fixed) and the binary propositional operator 'causes'. The new operators are supposed to satisfy certain conditions. Then the unary propositional operators 'does' and 'can do' (again the agent is kept fixed) are introduced by definition :

$$(\text{does } A) = \exists Q(\text{striving for } (A \wedge Q) \wedge (\text{striving for } (A \wedge Q) \text{ causes } A)),$$

$$(\text{can do } A) = \exists Q(\text{striving for } (A \wedge Q) \text{ causes } A),$$

where '$=$' is the strict biconditional of S4.

The idiom employed by Fitch is rather special (and perhaps this explains why later authors have ignored his paper). In order to make his ideas more directly comparable to current discussion, we propose the following reformulation of his theory in a language without propositional quantifiers (this means of course that we will neglect certain features of Fitch's theory). Let us add 'str', '$\Rrightarrow$', 'does' and 'can do' as four new primitive operators to the classical propositional calculus (without propositional quantifiers); these operators cannot nest. The following informal readings are suggested :

str A : the agent strives for it to be the case that A;
$A \Rrightarrow B$: that A partially causes that B;
does A : the agent sees to it that A;
can do A : the agent is able to see to it that A.

Consider the smallest classical logic (deducibility relation $\vdash$) that satisfies the following conditions. (We write "$A \dashv\vdash B$" if $A \vdash B$ and $B \vdash A$.) First four conditions that guarantee Replacement of Provable Equivalents :

(CS) if $A \dashv\vdash A'$, then str $A \dashv\vdash$ str A',
(CPC) if $A \dashv\vdash A'$ and $B \dashv\vdash B'$, then $A \Rrightarrow B \dashv\vdash A' \Rrightarrow B'$,
(CD) if $A \dashv\vdash A'$, then does $A \dashv\vdash$ does A',
(CC) if $A \dashv\vdash A'$, then can do $A \dashv\vdash$ can do A'.

Then some conditions governing the behavior of 'str':

(S1a) str$(A \wedge B) \vdash$ str A,
(S1b) str$(A \wedge B) \vdash$ str B,
(S2) str A, str $B \vdash$ str $(A \wedge B)$.

Next several conditions for the causality operator '$\Rrightarrow$':

(PC1) $A \Rrightarrow B,\ B \Rrightarrow C \vdash A \Rrightarrow C$,
(PC2) $A \Rrightarrow B,\ A \vdash B$,
(PC3a) $(A \wedge B) \Rrightarrow C,\ A \vdash B \Rrightarrow C$,
(PC3b) $(A \wedge B) \Rrightarrow C,\ B \vdash A \Rrightarrow C$,
(PC4) $A \Rrightarrow B,\ A \Rrightarrow C \vdash A \Rrightarrow (B \wedge C)$,
(PC5a) $A \Rrightarrow (B \wedge C) \vdash A \Rrightarrow B$,
(PC5b) $A \Rrightarrow (B \wedge C) \vdash A \Rrightarrow C$.

Finally some conditions that relate 'does' and 'can do' to the other concepts :

(D11) str$(A \wedge B) \Rrightarrow A$, str$(A \wedge B) \vdash$ does A,

(D12) if : str$(A \wedge B) \Rrightarrow A$, str$(A \wedge B) \vdash C$, then: does $A \vdash C$,
provided that B and C have no propositional letter in common,

(D21) str$(A \wedge B) \Rrightarrow A \vdash$ can do A,

(D22) if : $\text{str}(A \wedge B) \Rrightarrow A \vdash C$, then: can do $A \vdash C$,
provided that B and C have no propositional letter in common.

In Fitch's own logic it holds that

$$\begin{array}{ll} \text{does } (A \wedge B) \vdash \text{does } A, & \text{does } (A \wedge B) \vdash \text{does } B, \\ \text{can do } (A \wedge B) \vdash \text{can do } A, & \text{can do } (A \wedge B) \vdash \text{can do } B, \end{array}$$

however, there is a suggestion that the converses do not hold. We note that the four deducibility statements are reproducible in our version of Fitch's logic and give a proof of the first as an example. Suppose that A and B are any formulae. Let C be any formula having no propositional letter in common with A. A deduction is readily constructed from the following outline :

1.	$\text{str}((A \wedge B) \wedge C)$	premise
2.	$\text{str}((A \wedge B) \wedge C) \Rrightarrow (A \wedge B)$	premise
3.	$\text{str}(A \wedge (B \wedge C))$	from 1 by (CS)
4.	$\text{str}(A \wedge (B \wedge C)) \Rrightarrow (A \wedge B)$	from 2 by (CS)
5.	$\text{str}(A \wedge (B \wedge C)) \Rrightarrow A$	from 4 by (PC5a)
6.	$\text{does} A$	from 3 and 5 by (D11)

This shows that

$$\text{str}(A \wedge B) \wedge C,\ \text{str}(A \wedge B) \wedge C \Rrightarrow (A \wedge B) \vdash \text{does } A$$

By assumption, A and C have no propositional letter in common. Therefore, by (D12),

$$\text{does}(A \wedge B) \vdash \text{does } A,$$

as we wanted to show. ■

One way to prove that the statement

$$(*) \qquad \text{does } A, \text{does} B \vdash \text{does}(A \wedge B)$$

is *not* forthcoming in our system would be first to develop a suitable semantics for our logic. Without doing this we can still see why an attempt to give a direct proof of (*) would be frustrated. For suppose that we embarked on a deduction as follows :

1.	$\text{str}(A \wedge C)$	premise
2.	$\text{str}(A \wedge C) \Rrightarrow A$	premise
3.	$\text{str}(B \wedge D)$	premise
4.	$\text{str}(B \wedge D) \Rrightarrow B$	premise

The strategy would now be to try to arrive at

5. $\mathrm{str}((A \wedge B) \wedge (C \wedge D))$
6. $\mathrm{str}((A \wedge B) \wedge (C \wedge D)) \Rrightarrow (A \wedge B)$.

If we could achieve this, then by (D11) we would be able to conclude does $(A \wedge B)$, and, on the assumption that C and D did not share any propositional letter with one another or with either A or B, the rest would be easy. And, indeed, line 5 is derivable. However, line 6 is not. It would be if we had adopted the condition

$$A \Rrightarrow C,\ B \Rrightarrow C \vdash (A \wedge B) \Rrightarrow C.$$

However this rule is conspicuously absent from Fitch's system (and from ours), and for good reason.

This is perhaps a good opportunity to mention Anthony Kenny's critique in his book *Will, freedom and power* [13] from 1975 of modal logic *qua* a logic of action. All normal modal logics satisfy the condition

$$\Diamond(A \vee B) \vdash \Diamond A \vee \Diamond B.$$

and some also satisfy the further condition

$$A \vdash \Diamond A.$$

Kenny notes that, while modal logic might offer a way to formalize the 'can' of *opportunity*, it is unable to formalize the 'can' of *ability*. Several interesting counterexamples are mentioned. The former condition is violated by anyone who has the ability to pick a card from a pack of cards (which of course is either red or black) without having the ability to pick a red or the ability to pick a black ([13], p. 137), the latter condition by any sufficiently hopeless darts player who, for once, manages to hit the bull but who lacks the ability to repeat this feet ([13], p. 136). If Kenny had considered Fitch's system he would have noticed that it goes some way towards meeting his requirements: if P and Q are distinct propositional letters, then

$$\text{can do}(P \vee Q) \nvdash \text{ can do}\, P \vee \text{ can do}\, Q,$$
$$P \nvdash \text{ can do}\, P.$$

However, according to Fitch,

$$\text{does}\, A \vdash \text{ can do}\, A,$$

which Kenny would probably have found unacceptable (unless he were to hold that, on the occasion when the hopeless darts player hits the bull, it would be inappropriate to represent this by a formula of the type 'does (the dart hits the bull)' or 'does (the dart is in the bull)').

4. Foundations for Rights : Kanger

von Wright's interest in the logic of action arose from his interest in deontic logic: as he was developing the latter he became convinced that an understanding of the deontic notions presupposes an understanding of action. In a similar way Stig Kanger was led to the logic of action in order to find a basis for another theory, in his case the logical analysis of the concept of a right along the lines first drawn by Wesley N. Hohfeld (see the latter's *Fundamental legal conceptions* (1919)).

Kanger sets himself the task of explicating the phrase

"i has versus j a right to the effect that A".

where i and j are agents and A is condition. Already in 1957, in *New foundations for ethical theory* [9], Kanger observed that an explication — or rather several explications — could be achieved in a logical language which included deontic and causal notions. This observation was developed in full detail in two later papers, *Rättighetsbegreppet* [10] from 1963 and *Rights and parliamentarism* [12] from 1966, the latter written jointly with Helle Kanger. Here we shall give a brief outline of Stig Kanger's theory from the point of view of the logic of action.

In [9] Kanger noted that each of the following four conditions may claim to be an explication of the phrase displayed above :

Ought (j causes that A),
Right $\neg$(i causes that $\neg A$),
Right (i causes that A),
Ought $\neg$(j causes that $\neg A$).

In [10] and the 1966 version of [12] we have 'it shall be that' and 'it may be that' instead of the classical notions 'Ought' and 'Right' — for our purposes not an important change — but the causal notion is the same. In the 1968 version of [12] 'causes that' is replaced by 'sees to it that'; a still later paper, *Law and logic* [11], suggests that Kanger regarded those operators as identical ([11], p. 111). It is clear that this operator is related to Fitch's operator 'does'. However, here we prefer to use the notation 'Do', which was Kanger's own choice in [11]. Thus for '$\mathrm{Do}_i A$', read "agent i sees to it that A". The four candidates then become,

Shall $\mathrm{Do}_j A$,
$\neg$ Shall $\mathrm{Do}_i \neg A$,
$\neg$ Shall $\neg\, \mathrm{Do}_i A$,

Shall $\neg\,\mathrm{Do}_j\neg A$.

The following readings are suggested :

"i has vs j a *claim* to the effect that A",
"i has vs j a *freedom* to the effect that A",
"i has vs j a *power* to the effect that A",
"i has vs j an *immunity* to the effect that A".

Thus we find that by introducing a deontic operator 'Shall' and two action operators 'Do_i' and 'Do_j' we are able, in a very simple way, to make several distinctions that seem fundamental and very natural. Let us now map out all possible relationships definable in this particular way; that is, those that are of the form

$$\pm\mathrm{Shall}\ \pm \mathrm{Do}_k\ \pm A,$$

where '$\pm$' stands for either the negation sign '$\neg$' or nothing, and 'k' stands for either 'i' or 'j'. Evidently there are sixteen possibilities. Kanger singled out eight of these, calling them the *simple types* of rights. Changing his list in an inessential way, we may regard the following as the eight simple types :

simple type	*abbreviation*	*explication*
1. Claim(i,j,A)	$C(i,j,A)$	Shall $\mathrm{Do}_j A$
2. Freedom(i,j,A)	$F(i,j,A)$	$\neg$ Shall $\mathrm{Do}_i\neg A$
3. Power(i,j,A)	$P(i,j,A)$	$\neg$ Shall $\neg\mathrm{Do}_i A$
4. Immunity(i,j,A)	$I(i,j,A)$	Shall $\neg\mathrm{Do}_j\neg A$
5. Inverse–claim(i,j,A)	$C^\circ(i,j,A)$	Shall $\mathrm{Do}_i A$
6. Inverse–freedom(i,j,A)	$F^\circ(i,j,A)$	$\neg$ Shall $\mathrm{Do}_j\neg A$
7. Inverse–power(i,j,A)	$P^\circ(i,j,A)$	$\neg$ Shall $\neg\mathrm{Do}_j A$
8. Inverse–immunity(i,j,A)	$I^\circ(i,j,A)$	Shall $\neg\mathrm{Do}_i\neg A$

Kanger remarks that these concepts correspond to concepts distinguished by Hohfeld as follows :

HOHFELD	KANGER
claim	claim
privilege	freedom
power	power

immunity	immunity
duty	inverse–claim
no–right	inverse–freedom
liability	inverse–power
disability	inverse–immunity

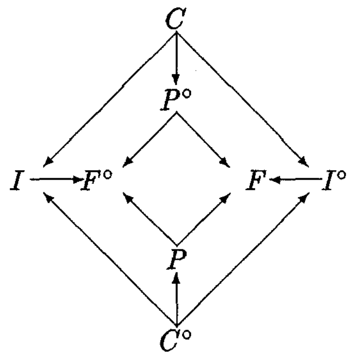

Fig. 4.1

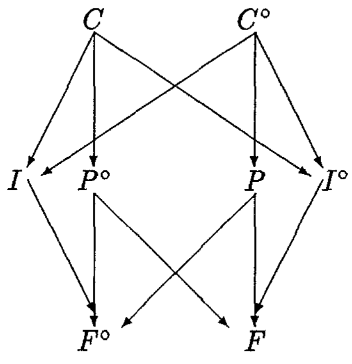

Fig. 4.2

Appealing to one's intuitions one might easily persuade oneself that there ought to be certain logical relationships between the eight simple types. The structure of these relationships is set out in Fig.4.1, where the types are abbreviated in an obvious manner and the arrows stand for logical implication. Fig. 4.1 gives the kind of representation preferred by Kanger. Another, more traditional, is given in Fig. 4.2.

The next step in Kanger's theory is to define the concept of an *atomic type* of right (between i and j to the effect that A). This we define as any condition logically equivalent to a condition of type

$$\pm C \wedge \pm F \wedge \pm P \wedge \pm I \wedge \pm C^{\circ} \wedge \pm F^{\circ} \wedge \pm P^{\circ} \wedge \pm I^{\circ}$$

where as before '$\pm$' stands for either a negation sign or nothing. A *molecular type* of right is any condition that is logically equivalent to a Boolean combination of atomic types.

In this way Kanger manages to develop with modest means a surprisingly rich theory; for, as it turns out, the number of molecular types of rights is enormous. This means that he has provided an instrument capable of high discrimination which should be useful for analytical purposes. Indeed, in the second half of [12], Helle Kanger goes on to apply Stig Kanger's theory to some actual problems in political science. Further applications are found in her doctoral dissertation *Human rights in the U. N. declaration* (1984). For a systematic study of these matters, including developments beyond Kanger's own theory but on the foundations provided by him, see the excellent treatise *Position and change* (1977) by his former student Lars Lindahl.

However, so far our presentation has no foundation as we have not accounted for the basic, nonclassical operators. In fact, Kanger himself does not tie his presentation to a unique underlying logic. Rather, he notes the conditions a logic must satisfy in order to fit his theory. Against the background provided by preceding exposition it is easy to see what those conditions are: the logic must support the logical relationships portrayed in the diagrams above. That is to say, the *minimum condition* is that the logic must be strong enough to allow the derivation of the logical implications in the diagrams; and the *maximum condition* is that the logic must no be so strong that any new implications are added in those diagrams.

The conditions on the new operators actually formulated by Kanger certainly satisfy the minimum condition. First, the logic is assumed to be classical in the sense of containing all tautologies as theses and being closed under *modus ponens*. Then the new operators are required to be congruential :

If $A \equiv B$ is a thesis, then so is Shall $A \equiv$ Shall B.
If $A \equiv B$ is a thesis, then so is $\mathrm{Do}_i A \equiv \mathrm{Do}_i B$, for every i.

Moreover, all instances of the following must be theses of the logic :

$\mathrm{Shall}(A \wedge B) \equiv (\mathrm{Shall}\ A \wedge \mathrm{Shall} B)$,
Shall $A \supset \neg$ Shall$\neg A$,
$\mathrm{Do}_i A \supset A$.

Kanger never bothers to say what the maximum condition comes to, understandably as the particular conditions are obvious as well as unwieldy. They might be summarized, schematically, as follows :

$$
\begin{array}{ll}
F^\circ \nvdash F & F \nvdash F^\circ, \\
F^\circ \nvdash I & F \nvdash I^\circ, \\
I \nvdash F & I^\circ \nvdash F^\circ, \\
P^\circ \nvdash I & P \nvdash I^\circ, \\
P^\circ \nvdash P & P \nvdash P^\circ, \\
P^\circ \nvdash I^\circ & P \nvdash I, \\
C \nvdash C^\circ & C^\circ \nvdash C.
\end{array}
$$

For reasons of symmetry it is enough to check the conditions in one of columns; that is, seven checks will suffice.

Checking for nonimplication is usually difficult if one does not have a semantics. It is noteworthy that Kanger, who was after all one of the inventors of possible worlds semantics for modal logic, developed his theory of rights entirely with syntactical means. In [9] he does provide semantics for the deontic operator but not for his causal operators. In [11], however, he does suggest a semantics for Do–operators. This is done in two steps. First, Kanger proposes a definition of the following type :

$$\mathrm{Do}_i A =_{df} \mathrm{D\acute{o}}_i A \wedge \mathrm{D\grave{o}}_i A,$$

where '$\mathrm{D\acute{o}}_i A$'and '$\mathrm{D\grave{o}}_i A$' are given the following readings, respectively :

"A is necessary for something that i does",
"A is sufficient for something that i does".

A complicated semantics is then given. A simpler but more accessible semantics in the same spirit is studied by another of Kanger's former students, Ingmar Pörn, in his doctoral dissertation *Action theory and social science* (1977) [17]. There he defines a Kripke type semantics for two action operators, associating with them alternativeness relations between possible worlds which we may denote by R_i' and R_i''. The intended meanings of these relations are roughly as follows :

$xR_i'y$ iff everything i does in x is the case in y;

$xR_i''y$ iff the opposite of everything i does in x is the case in y.

A single operator with the intuitive meaning "i brings it about that (causes it to be the case that, effects that) A" is then defined in terms of those operators. The intuitive significance of this semantics is not altogether clear, but in one respect it is similar to von Wright's logic of action. In von Wright, we saw, there is the result of the action (associated with the end–point), and there is what would have been the case, had the agent not done what he did (associated with the null–point). Similarly, R_i' is concerned with what the agent achieves, R_i'' with what would be the case "but for the agent's action".

On this ground Pörn is able to work out a nontrivial logic of action. We omit further details and refer the reader to Pörn's book. For a comment on Pörn's logic, see the author's paper "On the question of semantics in the logic of action" (1985).

5. The First Semantics : Chellas

The first fully articulated semantics for a logic of action is found in Brian F. Chellas's doctoral dissertation *The logical form of imperatives* [2] from 1969, written under the direction of Dana Scott. As indicated by the title, Chellas's primary interest was to analyse imperatives, but he became convinced that action operators are needed for this purpose. For this reason he introduced what he called an "instigative" operator, Δ, always relativized to an agent. We shall write '$\Delta_i A$' for what Chellas would read, "i sees to it that A".

Chellas's account is given within the general frame–work for intensional logic devised by Dana Scott. The following is a version of it, simplified in some ways (for example, the problem of individuals is glossed over). Suppose there is substratum T representing *time* — identified in [2] with the set of integers — and a set S of possible *states* of the world. By a *history* let us mean any function $T \to S$, and let H be some given set of histories. In this semantics the truth–value of a formula depends on both a history h and a point t of time; let us write $\models_{h,t} A$ to denote that A is true at h and t. To give truth–conditions for propositional letters and the Boolean connectives is trivial. The intensional operators considered by Chellas are all handled in a uniform manner, namely, by *alternativeness relations.* In the case of the instigative operator, this is done as follows. For each agent i and time t there is a binary relation $R_t(i) \subseteq H \times H$, and the truth–condition for 'Δ_i' has this general form :

$$\models_{h,t} \Delta_i A \text{ iff } \forall h' \in H(\langle h, h' \rangle \in R_t(i) \Rightarrow \models_{h',t} A).$$

There are various requirements to impose on $R_t(i)$ and its relationships with other alternativeness relations. One important necessary condition is that h and h' agree at all times preceding t (that is, for all $x < t$, $h(x) = h'(x)$).

This is a highly abstract formalism. Perhaps a picture (Fig. 5.1) can help explain the point of it all.

Think of T as the axis of abscissas and S as the axis of ordinates. (By doing so we allow ourselves a certain license, for while T is assumed to be linearly ordered, normally S would not be.) Now imagine how, as time runs through its range from beginning to end, a path unfolds before our eyes — a history. By time t a certain initial history has evolved, represented by the line OP. Usually it can be completed in various ways — many continuations of OP may be possible at t. In our picture two such continuations have been indicated by the lines PH and PH'; thus the lines OH and OH' represent two complete histories. The cone α (or APA') represents the class of continuations at P which are *logically possible* at t, while the cone β (or BPB') represents the class of continuations at P which are *physically possible* at t. Notice that β is subcone of α — continuations that are physically possible are logically possible, but the converse need not be true. It is easy to think of many other, similar cones of continuations of interest to philosophers, for example, the cone of all continuations that are *legally acceptable* at t and the cone of all continuations that are *morally acceptable* at t. In each such case there are operators corresponding to the cone in question: it is logically (physically, legally, etc.) necessary/possible with respect to h and t that A if it holds that A with respect to all/some h and t, where h is a continuation within the appropriate cone at t.

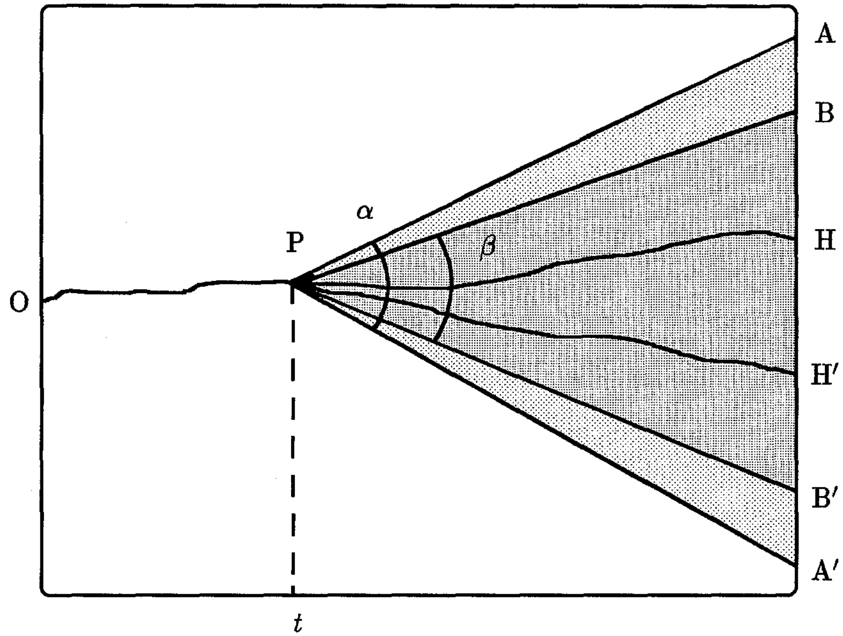

Fig. 5.1

It is in this spirit that Chellas introduces his action operator. What he needs is, at every point, a cone of what he terms 'actional alternatives', that is, cones of continuations that are possible in some "actional" sense yet to be explained. There is a complication, though: evaluating a formula $\Delta_i A$ at some history h and time t, the cone Chellas wishes to consider has its apex at the immediately preceding time, which we write $t-1$ (this is why he presupposes discrete time). Fig. 5.2 illustrates this idea.

So as not to clutter up the picture we have omitted the cones of continuations of the initial history OQ (with apex at Q) that are logically possible or physically possible at time $t-1$. Similarly, we have omitted the corresponding cones of continuations of the initial history OP (with apex at P). Only one complete history, h (or OH), is indicated in the picture. The cone γ (or CQC') contains all and only the continuations of OQ at $t-1$ that are relevant for evaluating $\Delta_i A$ with respect to h and t. In fact $\Delta_i A$ is regarded by Chellas as true with respect to h and t if and only if A is true with respect to h' and t for every complete history h' that shares OQ with h and continues within the cone γ. (The points $h'(t)$ for all those h' make up the line XY.)

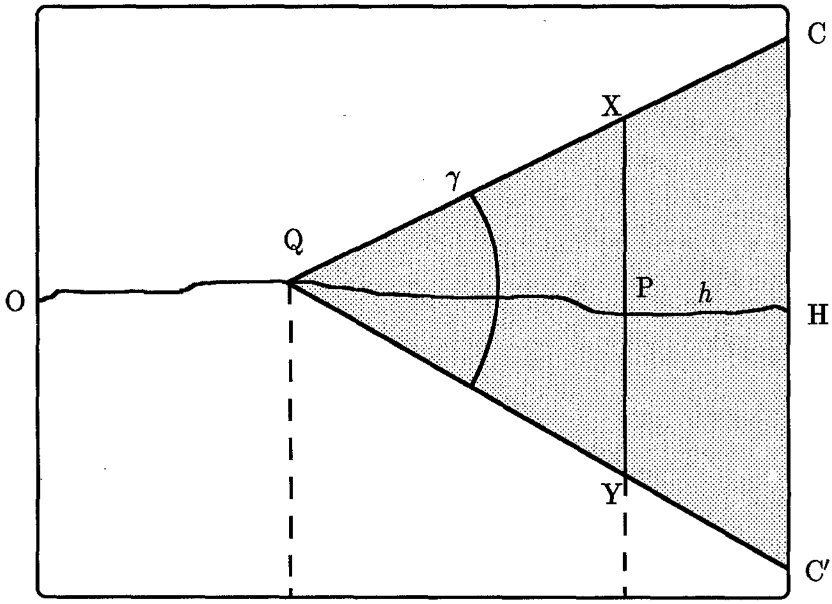

Fig. 5.2

Chellas, preoccupied with technical problems, never makes it entirely clear just how he thinks of the continuations in the cone γ. About all we are offered in the way of informal explanation is the comment that they are "under the control of" or "responsive to the actions of" the agent ([2], p. 63). One way to think of the cone γ would be to take Chellas to mean that it is the cone of continuations to which the agent, by his action, is able to confine the future development of the initial history OQ. This would make sense: we may not be able completely to determine the way history develops, but unless we are utterly powerless we can always rule out some otherwise possible continuations. For example, if I smash a glass at t, then I have effectively ruled out all continuations at t in which that particular glass is intact at any later time. However, this way of reading Chellas cannot be right, for then Δ_i would be an operator not of *agency* but of *ability*: it would tell us not what the agent *does* but what he *can do*.

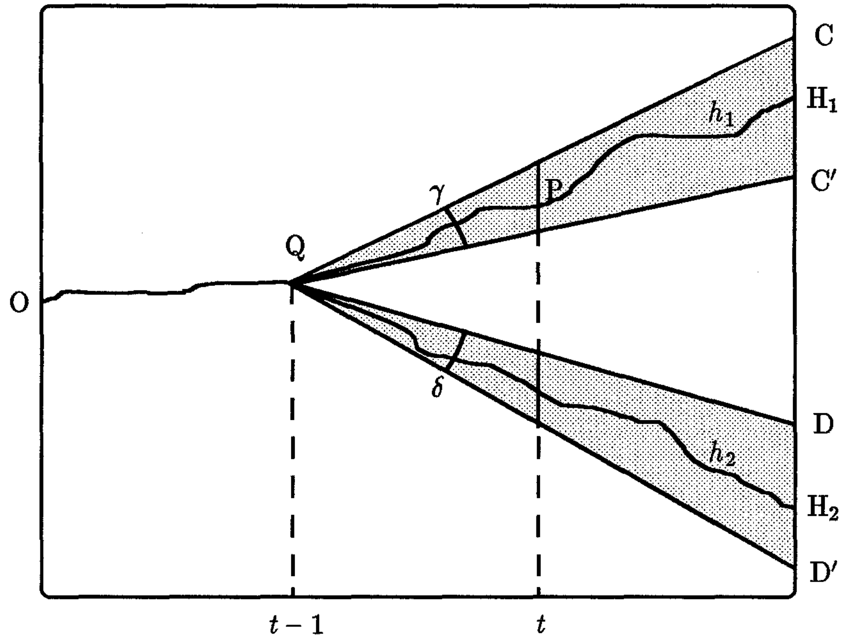

Fig. 5.3

Another way to think of γ is as the class of continuations which are the relevant ones at t, *given that* the agent does or has done whatever it is he does or has done. That makes Δ_i an operator of agency, as intended. But if this is what Chellas has in mind, then our picture in Fig. 5.2 is misleading: there is not just one cone of "actional alternatives", but one cone for every possible continuation. Thus the picture is rather as in Fig. 5.3. Because of the difficulty of drawing legible pictures involving many possible continuations we have indicated only two, OH_1 and OH_2: both h_1 and h_2 are complete histories which agree on OQ but diverge at Q. The cone γ (or CQC') is the cone corresponding to h_1, while δ (or DQD') is the cone corresponding to h_2. (The diagram represents γ and δ as disjoint, but that is not a necessary feature.)

Thus, from the vantage point of P, $\Delta_i A$ is true with respect to h_1 and t if and only if, for all h' such that h' and h_1 share OQ and h' is contained in the cone γ, A is true with respect to h' and t. Because of the agent's action, the "world-line", which at time $t-1$ had developed into OQ, has at time t developed into OP and might eventually, in the fullness of time, develop into h_1. But if the agent had acted differently, then h_2 might have become the resulting complete history. Notice that, whether or not $\Delta_i A$ is true with respect to h_1 and t, $\Delta_i A$ may or may not be true with respect to h_2 and t.

That this really is the way Chellas should be read is supported by the following consideration. If the picture suggested by Fig. 5.2 were what he had intended, then he should have adopted the assumption that, for all histories g, g' and h, if g and g' agree at all times preceding t, then $(g,h) \in R_t(i)$ if and only if $(g',h) \in R_t(i)$. He actually imposes the corresponding condition on one of his other alternativeness relations ([2], p. 83). The fact that he does not even contemplate imposing it on $R_t(i)$ is therefore highly suggestive.

Thus it seems that the cones needed in (this informal presentation of) Chellas's theory is determined by what the agent does. While this is not implausible, it would have been interesting to have been told something about the connexion between the agent and those cones. What is it that makes an initial history continue in one fashion rather than another? Does the agent "do" anything at $t-1$ to define a certain cone — does action consist in choosing or somehow committing oneself to a cone? Otherwise, where does action come from? And when does it take place — at $t-1$, at t, at the interval $[t-1,t]$, or what?

That Chellas's action semantics provides no picture of action itself is perhaps a feature it shares with any semantics that validates (or can easily be modified to validate) all formulae of type $\Delta_i A \supset A$. In Chellas's case this validity is ensured by the requirement that the alternativeness relations $R_t(i)$ be reflexive. But there seems to be no intuitive support for this requirement other then the wish to bestow validity on those formulae. Chellas's own comment is brief :

> This is perhaps the most minimal substantive axiom for Δ. One can see to it that such–and–such is, or be responsible for such–and–such's being, the case only if such and such is the case. ([2], p. 66)

This view is in agreement with the logics of Fitch, Kanger and Pörn, but it contrasts with the position of von Wright who, as we saw, does not accept $\mathcal{B}_i A \supset A$ as valid (although, unobjectionably, he would have to accept the validity of $\mathcal{S}_i A \supset A$). It is of some interest to note that of these authors it is only von Wright who has made a genuine effort to provide an analysis of action.

There is much power and beauty in Scott's and Chellas's conception, and it is surely fruitful in the sense that it can be taken much further.

In fact, recent work by Nuel Belnap and Michael Perloff may be seen as lending support to this assessment. In an already large and still growing sequence of papers, beginning with their 1989 paper "Seeing to it that: a canonical form for agentives"[0], they are jointly and severally developing a distinctive theory which, although of independent origin, is nevertheless within the general Scott/Chellas framework. Their work is probably one of the two most promising avenues of research in current logic of action. The other is the topic of the following section.

6. Dynamic Logic : Pratt

Let us modify the Scott/Chellas frame–work slightly. First, while we retain the set S of states, we eliminate the time set T. The elements (points) of S represent possible total states of the world; the assumption is that the world is always in some state or other, and that change in the world can be represented as a sequence of states. The way the world changes depends to a great extent on what agents do, but nature is also important: even if all agents remain totally passive, there will or might be a change in the world. (For some purposes it might be convenient to cast nature as an agent in her own right. In such a case all change in the world is determined by what the agents do.)

We may now tackle a question never faced by any of the previous authors, namely, how we might semantically represent the contribution agents make to the changing world. In the previous section we complained that the rôle of the agent — what he "really does" — seemed unclear in Chellas's semantics. In the section on von Wright, we felt the same problem; there we actually went beyond von Wright's text and introduced the concept of "options". But we still did not say what an option is or indicate how change in the world comes about.

This is where dynamic logic comes in: why not think of the option of an agent as something like a computer program? To simplify the situation and avoid the problems to do with joint or concurrent action, let us consider the case where there is just one agent. If he exercises an option he has, the world will leave its current state, traverse a number of states and then perhaps end up in some end–state; call this a *path* according to the option (the program in question). Of course, as nature never tires, there is no final state, so what is to be regarded as the end–state of an action is something that must be considered; here von Wright's extensive discussion of *results* of actions as distinguished from *consequences* of actions should be useful.

In this way it becomes natural to associate with each program α of the agent a set $P(\alpha)$, the set of paths according to α. For any point x in S, if $P_x(\alpha)$ is defined as the subset of $P(\alpha)$ whose paths begin with x, then $P_x(\alpha)$ would represent the possible computations according to α if α where started at x. Note that such a path may or may not terminate, may or may not "fail". A determinist might wish to require $P_x(\alpha)$ to be a singleton, but there is no logical compulsion to do so.

This modelling might be said to be "action driven". Tense–logic, by contrast, is "time–driven": there time is primary, somehow pulling the ever–changing "now" through the course of history. But for somebody interested in action it makes more sense to see time created from the actions of agents, a by-product of sorts.

The picture just drawn is inspired by dynamic logic. If you think of S as the set of possible total states of some automaton, then the notion of a path according to a program becomes not just suggestive but rigorous. This was the idea that occurred to Vaughan Pratt (see his papers in the references). As he was interested in program verification he introduced a new operator "after α", where α is a program. That is to say, as a first step he wanted to be able to formalize the concept

"after α, A",

where α is a program and A is a proposition. For that purpose the modelling just described is unnecessarily rich: only the end–state (if there is one) of a path matters, not the intermediate states. Thus for Pratt it was enough to postulate, for each program α, in lieu of the set $P(\alpha)$ of paths, a binary relation $R(\alpha)$ consisting of all ordered pairs $\langle x, y\rangle$ such that y is the end–state of some path in $P_x(\alpha)$. This means that there are two possible ways of formalizing "after α, A":

"after every terminating computation according to α, A", and
"after some terminating computation according to α, A",

for which Pratt introduced the notations '$[\alpha]A$' and '$\langle\alpha\rangle A$', respectively. The choice of shapes was not fortuitous: the square brackets suggest a box operator, the corners a diamond operator. And, indeed, dynamic logic may be regarded as a generalized modal logic, as seen from the following truth–conditions :

$$\models_x [\alpha]A \text{ iff } \forall y(xR(\alpha)y \Rightarrow \models_y A),$$
$$\models_x \langle\alpha\rangle A \text{ iff } \exists y(xR(\alpha)y \;\&\; \models_y A).$$

Programs can be combined to form more complex programs. In particular, dynamic logicians have studied the regular operations used to combine programs: $\alpha+\beta$, $\alpha;\beta$, α^*. It is simplest to explain their meaning by saying what it would mean to tell somebody to run one of those programs: you run $\alpha + \beta$ if and only if you run either α or β (you decide which); you run $\alpha;\beta$ if and only if you run α and then immediately after run β; and you run α^* if and only if you run α some finite number of times (you decide the number).

To preserve this intuition the following formal requirements must be made :

$$\begin{aligned} R(\alpha+\beta) &= R(\alpha) \cup R(\beta), \\ R(\alpha;\beta) &= R(\alpha) \mid R(\beta), \\ R(\alpha^*) &= (R(\alpha))^*. \end{aligned}$$

As it turns out, this logic (PDL = Propositional Dynamic Logic) is axiomatized by the following system (see Goldblatt's 1986 review article [6] for the history of this result). As rules adopt *modus ponens* and necessitation for each box operator $[\alpha]$. As axioms adopt all instances of the minimal normal modal logic K (for each box operator $[\alpha]$) plus all instances of the following schemata :

1. $[\alpha + \beta]A \equiv ([\alpha] \wedge [\beta]A)$,
2. $[\alpha;\beta]A \equiv [\alpha][\beta]A$,
3. $[\alpha^*]A \supset A$,
4. $[\alpha^*]A \supset [\alpha]A$,
5. $[\alpha^*]A \supset [\alpha^*][\alpha^*]A$,
6. $(A \wedge ([\alpha^*](A \supset [\alpha]A))) \supset [\alpha^*]A$.

The decidability problem for this logic was settled — in the affirmative — before the completeness problem was (see [3]). This is one of the first nontrivial *results* in the logic of action.

Is dynamic logic a logic of action? It seems to this author that one might well say so: certainly there is a strong kinship between the formal semantics of Scott and Chellas and the informal semantics of von Wright, on the one hand, and Pratt's semantics on the other. Even so, it has to be admitted that dynamic logic, as described here, lacks resources in the object language directly to express agency and ability. For example, it is not possible in PDL to define Kanger's '$\mathrm{Do}_i A$' or Chellas's '$\Delta_i A$'. But it is of course not difficult to add to the machinery of dynamic logic to express the idea that a certain program is run. Similarly, concepts of ability are close to hand: for example, the agent is able at x to see to it that A

in a strong sense if he has some program α such that α can be started at x, and A is true at all y such that $xR(\alpha)y$, and there is some y such that $xR(\alpha)y$ (and α always terminates if started at x);

in a weak sense if he has a program α such that A is true at some y such that $xR(\alpha)y$.

These observations are closely related to the ideas in an interesting recent paper by Mark Brown, "On the logic of ability" (1988) [1].

But if dynamic logic is a logic of action, it is primarily a logic of computer action. Human action is considerably more complicated. For example, a viable semantics for a comprehensive logic of action must account for concepts such as 'intention' and 'goal'. Some efforts have already been made to tackle these problems. But almost all the work remains to be done.

References

[0] NUEL BELNAP and MICHAEL PERLOFF, *Seeing to it that: a canonical form for agentives*, **Theoria** 54 (1988), pp. 175 – 199. Reprinted with some corrections in **Knowledge representation and defeasible reasoning**, H. E. Kyburg et al. (eds), Kluwer Academic Publishers, Dordrecht 1990, pp. 167 – 190.

[1] MARK A. BROWN, *On the logic of ability*, **Journal of philosophical logic** 17 (1988), pp. 1 – 26.

[2] BRIAN F. CHELLAS, **The logical form of imperatives**, Perry Lane Press, Stanford 1969.

[3] MICHAEL J. FISCHER and RICHARD E. LADNER, *Propositional dynamic logic of regular programs*, **Journal of computer and system sciences** 18 (1979), pp. 194 – 211.

[4] FREDERIC B. FITCH, *A logical analysis of some value concepts*, **Journal of symbolic logic** 28 (1963), pp. 135 – 142.

[5] ROBERT GOLDBLATT, **Axiomatising the logic of computer programming**, Springer–Verlag, Berlin, Heidelberg, New York, 1982.

[6] ROBERT GOLDBLATT, Review article, **Journal of symbolic logic** 51 (1986), pp. 225 – 227.

[7] PAUL DESMOND HENRY, **The logic of Saint Anselm**, Clarendon Press, Oxford 1967.

[8] HELLE KANGER, **Human rights in the U. N. Declaration**, Acta universitatis upsaliensis, Uppsala 1984.

[9] STIG KANGER, **New foundations for ethical theory**, Part I, Mimeographed, privately distributed, Stockholm 1957. Reprinted in **Deontic logic: introductory and systematic readings**, Risto Hilpinen (ed.), D. Reidel Publ. Co., Dordrecht 1971.

[10] STIG KANGER, *Rättighetsbegreppet*, **Sju filosofiska studier tillägnade Anders Wedberg, Stockholm 30 mars 1963**, Philosophical studies published by the Department of Philosophy, University of Stockholm, vol. 3, Stockholm 1963, pp. 79 –102.

[11] STIG KANGER, *Law and logic*, **Theoria** 38 (1972), pp. 105 – 132.

[12] STIG KANGER and HELLE KANGER, *Rights and parliamentarism*, **Theoria** 32 (1966), pp. 85 – 115. Reprinted with some changes in **Contemporary philosophy in Scandinavi**, Raymond E. Olson and Anthony M. Paul (eds), The Johns Hopkins Press, Baltimore and London 1972, pp. 213 – 236.

[13] ANTHONY KENNY, **Will, freedom and power**, Blackwell, Oxford 1975.

[14] LARS LINDAHL, **Position and change: a study in law and logic**, D. Reidel Publ. Co., Dordrecht 1977.

[15] VAUGHAN R. PRATT, *Semantical considerations on Floyd–Hoare logic*, **17th Annual Symposium on Foundations of Computer Science**, I. E. E. E., New York 1976, pp. 109 – 121.

[16] VAUGHAN R. PRATT, *Application of modal logic to programming*, **Studia Logica** 39 (1980), pp. 257 – 274.

[17] INGMAR PÖRN, **Action theory and social science: some formal models**, D. Reidel Publ. Co., Dordrecht 1977.

[18] KRISTER SEGERBERG, *A topological logic of action*, **Studia Logica** 43 (1985), pp. 415 – 419.

[19] KRISTER SEGERBERG, *On the question of semantics in the logic of action: some remarks on Pörn's logic of action*, **Action, logic and social theory**, Ghita Holmström and Andrew J. I. Jones (eds.), **Acta philosophica fennica** 38 (1985), pp. 282 – 298.

[20] GEORG HENRIK VON WRIGHT, **Norm and action: a logical inquiry**, Routledge & Kegan Paul, London 1963.

[21] GEORG HENRIK VON WRIGHT, **Varieties of goodness**, Routledge & Kegan Paul, London 1963.

[22] GEORG HENRIK VON WRIGHT, **Explanation and understanding**, Cornell University Press, Ithaca, N. Y. 1971.

[23] GEORG HENRIK VON WRIGHT, **An essay in deontic logic and the general theory of action**, **Acta philosophica fennica**, fasc. 21 (1972)

[24] GEORG HENRIK VON WRIGHT, **Causality and determinism**, Columbia University Press, New York 1974.

[25] GEORG HENRIK VON WRIGHT, *Handlungslogik*, in **Normenlogik**, Hans Lenk (ed.), Pullach bei München: Verlag Dokumentation, 1974. Reprinted as *Elemente der Handlungslogik* in **Handlungstheorien interdisziplinär I**, Hans Lenk (ed.), Wilhem Fink Verlag, München 1980, pp. 21 – 34.

[26] GEORG HENRIK VON WRIGHT, **Practical reason**, Blackwell, Oxford 1983.

DEPARTMENT OF PHILOSOPHY
VILLAVÄGEN 5
S–752 36 UPPSALA, SWEDEN

Received July 17, 1991