Action theories

Andreas Herzig, Emiliano Lorini, Nicolas Troquard

Abstract

We present the main logical theories of action. We distinguish theories identifying an action with its result from theories studying actions in terms of both their results and the means that result is obtained. The first family includes most prominently the logic of seeing-to-it-that and the logic of bringing-it-about-that. The second includes propositional dynamic logic and its variants. For all these logics we overview their extensions by other modalities such as modal operators of knowledge, belief, and obligation.

Keywords: action; agency; capability; seeing to it that; stit; bringing it about

Published as: Herzig A., Lorini E., Troquard N. (2018) Action Theories. In: Hansson S., Hendricks V. (eds) Introduction to Formal Philosophy. Springer Undergraduate Texts in Philosophy. Springer, Cham

1 Introduction

Actions such as raising one's arm, switching on a computer, jumping a traffic light, killing somebody, or waltzing are investigated in several areas of philosophy, among others in philosophy of action, philosophy of language and philosophy of law. Through the analogy between actions and programs the concept is also relevant in computer science, in particular in artificial intelligence, multi-agent systems and theoretical computer science. Several other concepts are intimately related to action. One that is directly related is that of the *ability* to act. Mental attitudes and norms also play an essential role in the study of action.

It has been attempted since Aristotle to systematise the analysis of action. Taking advantage of the mathematical advances in predicate logic, ontological perspectives on action were proposed in the form of first-order theories in the mid-20th century and have been very influential in philosophy. Concurrently, various research programs investigated the logic of action as such, trying to uncover the grand principles. These approaches are dominated by a modal view of action, and a first survey of this field is in a 1992 special issue of Studia Logica [45]. The present chapter overviews the resulting logics of action. We start by introducing the main issues at stake.

Actions as events brought about by agents. It is generally considered that an action can be identified with an *event* that is brought about by an *agent* [48, 14], as exemplified by Belnap talking about "an agent as a wart on the skin of an action" [5]. The dedicated term in the literature is that an agent is *agentive* for an event. Examples of events are that an arm goes up, that a computer starts, that somebody dies, etc. So my action of switching the computer on is identified with me bringing about the event that the computer starts.

Essentially, there exist two different semantical accounts of events: the first account identifies an event with a *set of possible worlds*, also called a *proposition*; the second account identifies an event with a *binary relation between possible worlds*, also called a transition relation. In the first view, events are facts of the world, identified with propositions: subsets of the set of possible worlds where the event occurs. To these propositions the usual set-theoretic operations can be applied. We thus obtain a way to interpret complex events and actions that are built with the logical connectives of propositional logic, such as negation, conjunction, and material implication. In the second view, the transition relations of atomic events are a given, and the transition relation of a complex event is built up from them.

Action as result vs. action as 'means+result'. The two views on the semantics of events yielded two traditions of logics of action. The difference is reflected by two different logical forms of action sentences they consider: the first family is about sentences such as "I bring it about that the computer is on" and focuses on *the result of an action*; the second family is about sentences such as "I bring it about that the computer is on by pushing the power button" and focuses on both *the result and the means by which it is obtained*.

The first family are the so-called logics of agency. The logic of seeing-to-it-that (STIT) [6, 4] and the logic of bringing-it-about-that (BIAT) [41, 19, 20] are two sub-families. These logics are studied in philosophy of action and more recently in multi-agent systems. The second family contains variants and extensions of propositional dynamic logic (PDL). These latter logics were introduced and studied in theoretical computer science, but were also investigated by philosophers.

Potential action. A notion that is often studied along with actual agency is the mere existence of a potential action. "He could have done otherwise"; "She can win this match"; "The Democrats have a strategy to undermine the influence of the Senate whatever the rest of the electorate does"; "I can switch on the light if you want"; "He can! But he would be lucky!" Loaded with many distinct but somewhat overlapping meanings, this notion has been called ability, capability, opportunity, power, etc. In this chapter we will simply use *ability* as an umbrella term for potential action.

Some meanings of the term *ability* have not yet been satisfyingly formalised in logic. One in particular is Kenny's sense of ability [29]: I am able to do an

	result	means+result
potential only	CL, ATL	PDL
actual only	BIAT	linear PDL
potential+actual	STIT, Elgesem's BIAT	PDL with actual actions, DLA

Table 1: Logical form and concepts of the logics of this chapter.

action if when I try to do that action under normal conditions then I usually succeed. Kenny's example is that of an expert dart player who is able to hit the bullseye while a layman is not. Although very close to our real world experience, one difficulty is to meaningfully capture in a formalism that ability is not a sufficient condition for actual agency and that actual agency is not evidence of ability *à la* Kenny.

Yet, possible action has been studied alongside actual action in some logical formalisms. All of the logics presented in this chapter that deal with both actual and potential agency subscribe the principle 'actual agency implies potential agency', for short: 'do implies can'.

- BIAT logic is about actual agency. Elgesem has added a notion of ability to bring about a proposition. In his logic an ability still can exist without actual agency: a lion in a zoo can catch a zebra. Both agency and ability are primitive concepts in his logic (although they are defined by means of the same semantic structure).

- STIT logic is primarily about actual agency and potential agency. It is equipped with quantification over possible unrolling of events. Potential agency for a proposition is then reduced to the existence of an unrolling of events where actual agency for that proposition is expressed.

- Coalition logic CL [38] and alternating-time temporal logic ATL [1] are about the ability of an agent to ensure something whatever the other agents do. There is no notion of actual agency, and the language does not explicitly refer to action terms.

- The standard version of PDL [22] enables to talk about the possibility of the occurrence of an event and about what is true afterwards. Linear versions of PDL also allow to capture actual agency. Furthermore, there are variants of PDL which allow to represent both actual action and potential action such as PDL with actual actions [32] and DLA [24].

Table 1 classifies the logics that we are going to overview in this chapter according to the distinctions 'potential and/or actual agency' and 'result vs. result+means' .

Actions and mental attitudes. Our actions are determined by our beliefs and desires: I switch my computer on because I want to know the weather forecast and believe I can find it on the Internet, or because I believe I got email and want to read it, or because I want to send an email and believe my Internet connection is not down.

According to an influential view due to Bratman, desires do not directly lead to actions, but it is rather the intermediate mental attitude of intention that triggers actions [7]. Cohen and Levesque designed a logic adding modal operators of belief and choice to PDL within which intention can be defined [17].

Actions and deontic concepts. What we do is not only influenced by our mental attitudes, but also by obligations and prohibitions. Indeed, there are cases where agents perform actions independently of their beliefs and desires merely because they are obliged to do so; think e.g. of soldiers blindly obeying their commander.

Meyer gave a logical account of obligation and action that is based on PDL [35], while Horty based his account on STIT [26, 25].

The rest of this chapter. We are now going to present the main logics of action and discuss their basic logical principles. In the next section we introduce the family of those logics allowing to talk about actions in terms of their results: BIAT and STIT. Thereafter we present the family of logics allowing to talk about actions in terms of results and means to achieve these results: PDL and its linear variants. For each family we discuss the interplay with ability, mental attitudes and norms.

Throughout this chapter $\varphi, \psi, \ldots$ denote formulas and $i, j, \ldots$ denote agents (individuals) that populate the world.

2 Action as result

According to Belnap and Perloff's 'stit-thesis' every agentive sentence can be transformed into a sentence of the form "i sees to it that φ", where i is an agent and φ is a proposition. In other words, an action is identified with the result it brings about. The sentence "agent i sees to it that φ" itself can then be viewed as a proposition. This allows a purely logical analysis of agentive sentences.

Let us start by formulating several principles that all of the logics in this section satisfy.

First, if we view agentive sentences as propositions then it is natural to require that the set of worlds where φ is true contains the set of worlds where i is agentive for φ. This is a *principle of success*: the proposition "i sees to it that φ" should imply the proposition φ. In other words, it should be valid that if i sees to it that φ then φ is true. Note that it follows from this principle that an agent can never see to it that $\varphi \land \neg\varphi$.

Second, the different approaches agree about the *principle of aggregation*: "if i sees to it that φ and i sees to it that ψ then i sees to it that $\varphi \wedge \psi$".

Third and as already discussed in the introduction, action implies ability. This is a *do implies can* principle: "if i sees to it that φ then i is able to achieve φ".

Fourth, a bringing about of a proposition is not sensitive to the syntactical formulation of that proposition. For example, if Zorro and Don Diego Vega are the same person and one considers that their being dead is the same proposition, then Sgt. Gonzales bringing about that Zorro is dead is equivalent to Sgt. Gonzales bringing about that Don Diego Vega is dead. This is the *principle of equivalents for actual agency.* A similar principle can be formulated for potential agency.

All variants of STIT and of BIAT satisfy the principles of success, of aggregation, 'do implies can', and equivalents for agency. Beyond these standard principles there are quite some differences that have been captured by quite different semantics. We therefore present the two families separately.

The main difference between BIAT logic and STIT logic is that the latter satisfies a principle of independence of agents while the former does not: in STIT it is assumed that each combination of the agents' individual actions can be *chosen jointly*, while this is not required in BIAT. It may be argued that while the principle of independence of agents is acceptable in the case of *choice* (or *trying*), it is less so in the case of *action.* Suppose two agents are standing in front of a room door and intend to enter the room. The door is too narrow to allow them to successfully enter at the same time, even though each agent can successfully enter if the other agent does nothing. While the two agents can simultaneously decide/try to enter the room, their attempts will fail to be performed successfully.

After the presentation of each family of logics we briefly mention extensions by concepts such as knowledge, belief, intention, and obligation.

2.1 The logic of bringing-it-about-that BIAT

BIAT logic, the logic of bringing-it-about-that, dates back to Kanger and Pörn [28, 41].[1] We here present Elgesem's semantics [20] whose validities were axiomatised by Governatori and Rotolo [21]. The semantics is in terms of selection function models $\langle W, \{f\}_i, V\rangle$ where W is some set of possible worlds, $V : \mathcal{P} \longrightarrow 2^W$ is a valuation function mapping propositional variables to subsets of W, and for every agent i, $f_i : W \times 2^W \to 2^W$ is a selection function associating a proposition to every possible world and proposition. The object $f_i(w, X)$ is the set of those worlds where i realises the ability he has in w to bring about his goal X. Therefore i is able to bring about X at w if $f_i(w, X)$ is nonempty; and i brings about X at w if w belongs to $f_i(w, X)$.

The functions f_i have to satisfy the following additional constraints:

[1] There is no well-established name in the literature, we therefore opted for the acronym BIAT, justs as the well-established STIT stands for 'seeing-to-it-that'.

- $f_i(w, X) \subseteq X$, for every $X \subseteq W$ and $w \in W$;
- $f_i(w, X_1) \cap f_i(w, X_2) \subseteq f_i(w, X_1 \cap X_2)$, for every $X_1, X_2 \subseteq W$ and $w \in W$;
- $f_i(w, W) = \emptyset$, for every $w \in W$.

The first two constraints correspond to the principle of success and to the principle of aggregation. The third constraint says that an agent cannot be agentive for a tautology.

The language of BIAT logic has modal operators of agency $\mathtt{Biat_i}$ and modal operators of ability $\mathtt{Can}_i$, one of each for every agent i. The formula $\mathtt{Biat_i}\varphi$ reads "i brings it about that φ", and the formula $\mathtt{Can}_i\varphi$ reads "i can achieve φ".[2]

The truth conditions are as follows:

$$
\begin{array}{lll}
M, w \models p & \text{iff} & w \in V(p); \\
M, w \models \mathtt{Biat_i}\varphi & \text{iff} & w \in f_i(w, ||\varphi||_M); \\
M, w \models \mathtt{Can}_i\varphi & \text{iff} & f_i(w, ||\varphi||_M) \neq \emptyset.
\end{array}
$$

In the last two conditions the set $||\varphi||_M$ is the extension of φ in M, i.e. the set of possible worlds where φ is true: $||\varphi||_M \stackrel{\text{def}}{=} \{w \in W : M, w \models \varphi\}$.

Alternative semantic characterisations of the operators $\mathtt{Biat_i}$ exist in the literature: Pörn proposed to simulate it by combining two more elementary modal operators that are normal [41]; Carmo *et col.* have used neighborhood semantics [43]. However, there are no completeness results for these alternative semantics.

So, what are the axioms of BIAT, i.e., what are the formulas of the language that are true in every model? As announced above, the axioms of success, aggregation, and 'do implies can' are all valid in BIAT logic, and the rule of equivalents preserves BIAT validity:

$$\mathtt{Biat_i}\varphi \rightarrow \varphi \tag{1}$$

$$(\mathtt{Biat_i}\varphi \wedge \mathtt{Biat_i}\psi) \rightarrow \mathtt{Biat_i}(\varphi \wedge \psi) \tag{2}$$

$$\mathtt{Biat_i}\varphi \rightarrow \mathtt{Can}_i\varphi \tag{3}$$

$$\frac{\varphi \leftrightarrow \psi}{\mathtt{Biat_i}\varphi \leftrightarrow \mathtt{Biat_i}\psi} \tag{4}$$

$$\frac{\varphi \leftrightarrow \psi}{\mathtt{Can}_i\varphi \leftrightarrow \mathtt{Can}_i\psi} \tag{5}$$

A subject that has been a source of disagreement in the literature is whether an agent can bring about a logical tautology. Can John bring it about that $2 + 2 = 4$? BIAT rules it out:

$$\neg\mathtt{Can}_i\top \tag{6}$$

[2]Instead of $\mathtt{Biat_i}$ Jones and Pörn use $\mathtt{E}_i$ and Elgesem uses $\mathtt{Does}_i$. Instead of $\mathtt{Can}_i$ Elgesem uses $\mathtt{Ability}_i$.

is an axiom. Together with the 'do implies can' axiom of Equation 3, it entails that $\neg\mathtt{Biat_i}\top$ is valid. That is, no agent is agentive for a tautology.

The principle that is maybe most surprisingly absent is the axiom of monotony $\mathtt{Biat_i}(\varphi \wedge \psi) \to (\mathtt{Biat_i}\varphi \wedge \mathtt{Biat_i}\psi)$: i may bring it about that $\varphi \wedge \psi$ without necessarily bringing it about that φ. $\mathtt{Biat_i}$ is therefore not a normal modal 'box' operator. The same is the case for the logic of the ability operators $\mathtt{Can}_i$. Moreover, they do not satisfy the principle $\mathtt{Can}_i(\varphi \vee \psi) \to (\mathtt{Can}_i\varphi \vee \mathtt{Can}_i\psi)$; to see this take $\psi = \neg\varphi$. Therefore the latter cannot be modal 'diamond' operators either. Moreover they do not satisfy $\varphi \to \mathtt{Can}_i\varphi$. Due to these last two properties Elgesem's ability operators satisfy what Brown calls Kenny's constraint [11].

In presence of several agents, these operators can be combined to express interesting properties of interaction. One can say for instance that an agent i makes (resp. can make) another agent j bring it about that φ, in formula: $\mathtt{Biat_i}\mathtt{Biat_j}\varphi$ (resp. $\mathtt{Can}_i\mathtt{Biat_j}\varphi$). Following the common law maxim "quid facit per alium facit per se", some authors consider that when i makes another agent bring about something then i himself brings about that something [13]. Others disagree [20]. Troquard [47], in a group extension of BIAT suggests a principle $\mathtt{Biat_i}\mathtt{Biat_j}\varphi \to \mathtt{Biat_{\{i,j\}}}\varphi$, where $\mathtt{Biat_{\{i,j\}}}\varphi$ indicates that the group composed of i and j brings about φ together. Aiming at another kind of compromise, Santos *et al.* have proposed a logic with two kinds of agency operators: one of indirect agency (noted $\mathtt{G}_i$) satisfying the above principle and another one of direct agency (noted $\mathtt{E}_i$) which does not (and instead satisfies $\mathtt{E}_i\mathtt{E}_j\varphi \to \neg\mathtt{E}_i\varphi$) [42, 43].

Our next family of logics will validate this principle, and much more.

2.2 The logic of seeing-to-it-that STIT

While the temporal aspects were kept abstract in BIAT logics, the semantics of STIT logics inherits the Ockhamist conception of time [50] where the truth of statements is evaluated with respect to a moment that is situated on a particular history through time (that is identified with a sequence of moments). This is one of the reasons why the models of STIT logics that we are going to present now [4, 26, 25] are more intricate. A systematic comparison between Belnap et al.'s semantics for STIT and other semantics for STIT such as the Kripke-style semantics by [31] and the bundled-tree semantics by [16] has been recently proposed by [15].

A STIT model is based on a *tree of moments* which are the possible states of the world. Every moment occurs at an *instant*, a mere time-stamp. A *history* is a maximal path in the tree. When a moment belongs to a history we say that the history passes through the moment. Time is therefore indeterministic, and indeterminism is due mainly to agents making choices where they could have chosen otherwise: at every moment m, each of the agents has a repertoire of *choices*, and each of these choices consists in selecting a subset of the histories passing through m. The future is understood to be on one of the selected histories. Then the future lies among the histories at the intersection of the choices taken by all agents. Whatever each of the agents chooses, the intersection of all the agents' choices must be non-empty. This is the *independence constraint*.

Formulas are evaluated in a STIT model M with respect to moment-history pairs (m, h) such that m is on h. A significant variety of modalities of agency have been studied within STIT logic, with sometimes only little differences. We are going to mainly talk about two of them that are rather different: the *achievement stit* and the *Chellas stit*. Both have in common with the BIAT modality the principles of Equations 1, 2, 3, 4, and 5 of Section 2.1. The achievement stit moreover satisfies the principle of Equation 6, while the Chellas stit does not.

The theories are also equipped with an operator of historical possibility $\Diamond$. The formula $\Diamond\varphi$ reads "there is a possible history passing through the current moment such that φ". Formally speaking, given a history h and a moment m passing through h (i.e., such that m is on h), the formula $\Diamond\varphi$ is interpreted as follows:

$$M, h, m \models \Diamond\varphi \quad \text{iff} \quad M, h', m \models \varphi \text{ for some history } h' \text{ such that } m \text{ is on } h'.$$

We can define the dual modal operator $\Box$ by stipulating $\Box\varphi \overset{\text{def}}{=} \neg\Diamond\neg\varphi$ and thereby express the fact that "φ is settled true at the current moment".

The original stit modality proposed by Belnap and Perloff [6] is the *achievement stit*. Let us write $\mathtt{AStit_i}$ for that modal operator. An agent i sees to it that φ if a previous choice of i made sure that φ is true at the current instant, and φ could have been false at this instant had i done otherwise.

$M, h, m \models \mathtt{AStit_i}\varphi$ iff there is a moment m_0 preceding m on h such that
(1) $M, h', m' \models \varphi$ for every h' and m' such that
  (i) h and h' are in the same choice of i at m_0,
  (ii) m' is on h' and at the same instant as m;
(2) there is a history h'' and a moment m'' at the same instant as m with $M, h'', m'' \not\models \varphi$.

Just as in BIAT logic, the idea of achievement is conveyed by validity of the principle of success ($\mathtt{AStit_i}\varphi \to \varphi$) and by the principle that no agent sees to a tautology ($\neg\mathtt{AStit_i}\top$).

Now comes a rather fascinating insight from such a complex modality. If $\mathtt{AStit_i}\varphi$ is i doing φ, one can capture that agent i refrains from doing φ by the formula $\mathtt{AStit_i}\neg\mathtt{AStit_i}\varphi$. What the logic tells us is that *doing* is equivalent to *refraining from refraining from doing*:

$$\mathtt{AStit_i}\varphi \leftrightarrow \mathtt{AStit_i}\neg(\mathtt{AStit_i}\neg\mathtt{AStit_i}\varphi).$$

(Precisely, this holds under the assumption that an agent does not perform an infinite number of non-vacuous choices during a finite interval of time.)

Horty and Belnap [26] simplified the achievement stit into the *deliberative stit* where the decisive choice of the action is at the current moment. The idea of deliberativeness resides in that an agent is currently seeing to something but could as well see to something else. The logic of the *Chellas stit* further simplifies the deliberative stit by removing the negative part from the truth condition. Let us write $\mathtt{CStit_i}$ for Chellas's stit operator. Its semantics is as follows:

$$M, h, m \models \mathtt{CStit_i}\varphi \quad \text{iff} \quad M, h', m \models \varphi \text{ for every } h' \text{ such that } h \text{ and } h' \text{ are in the same choice of } i \text{ at } m.$$

Hence the Chellas stit operator is a simple quantification over the histories that the current choice of the agent allows. A trained logician may observe that 'being in the same choice cell' is an equivalence relation and that every operator $\mathtt{CStit_i}$ therefore obeys the principles of modal logic $\mathsf{S5}$.

While the axiom of monotony is invalid in BIAT logic, the corresponding formula is valid for the Chellas stit:

$$\mathtt{CStit_i}(\varphi \wedge \psi) \rightarrow (\mathtt{CStit_i}\varphi \wedge \mathtt{CStit_i}\psi). \tag{7}$$

The striking principle of the Chellas stit that earned it its name (because Chellas has been a strong advocate, see [45]) is:

$$\Box\varphi \rightarrow \mathtt{CStit_i}\varphi. \tag{8}$$

In words, an agent cannot avoid what is settled; in particular he can and does bring about every tautology.

Just as the achievement stit, both the Chellas stit operator and the deliberative stit operator satisfy that refraining from refraining from doing is doing (even without the assumption that an agent does not perform an infinite number of non-vacuous choices during a finite interval of time).

A common feature of all STIT logics is that the agents' choices are constrained to be *independent*, while they are not necessarily so in BIAT logic. This can be nicely characterised in the logic of the Chellas stit by the principle

$$(\Diamond\mathtt{CStit_i}\varphi \wedge \Diamond\mathtt{CStit_j}\psi) \rightarrow \Diamond(\mathtt{CStit_i}\varphi \wedge \mathtt{CStit_j}\psi), \text{ for } i \neq j. \tag{9}$$

It follows that when i and j are different then $\Diamond\mathtt{CStit_i}\varphi \wedge \Diamond\mathtt{CStit_j}\neg\varphi$ is unsatisfiable (because $\mathtt{CStit_i}\varphi \rightarrow \varphi$ is valid and because $\Diamond$ is a normal modal operator). This principle can straightforwardly be extended from two agents i and j to any finite number of agents and is central in Xu's axiomatisation of the Chellas stit ([4, Chap. 17]). In contrast, there is no BIAT formula corresponding to Equation 9, simply because the right hand side of the implication cannot be expressed (due to the absence of an operator of historic possibility in the existing BIAT logics).

A somewhat surprising consequence of the independence of agents is the validity of the following 'make do implies settled' principle:

$$\mathtt{CStit_i}\mathtt{CStit_j}\varphi \rightarrow \Box\varphi, \text{ for } i \neq j. \tag{10}$$

In words, i can make j see to it that φ only if φ is settled. This highlights that unlike in BIAT, in STIT logics we cannot reason about the power of agents over others. While this principle may be felt to be unfortunate from the point of view of common sense, it accommodates well with social choice theory and game theory. In [2] it is shown that the schema of Equation 10 is actually equivalent to the schema of Equation 9 and that its generalisation to any finite number of agents can substitute Xu's axiom of independence in the axiomatisation of STIT.

We just mention that when combined with the operator of historical possibility, the Chellas stit operator can express the *deliberative stit operator* $\mathtt{DStit_i}$ as follows:

$$\mathtt{DStit_i}\varphi \stackrel{\text{def}}{=} \mathtt{CStit_i}\varphi \wedge \Diamond\neg\varphi.$$

The other way round, the Chellas stit operator can be expressed by $\mathtt{DStit_i}$ as:

$$\mathtt{CStit_i}\varphi \stackrel{\text{def}}{=} \mathtt{DStit_i}\varphi \vee \Box\neg\varphi.$$

The Chellas stit operator together with historical possibility also allows to express by $\Diamond\mathtt{CStit_i}\varphi$ that an agent has the *ability* to see to it that φ. The schema $\mathtt{CStit_i}\varphi \to \Diamond\mathtt{CStit_i}\varphi$ is valid and provides a 'do implies can' principle. While the aggregation principle is clearly invalid for that ability operator, it satisfies monotony and the principle $\Diamond\mathtt{CStit_i}\top$. Hence every $\mathtt{CStit_i}$ is a normal modal diamond operator (violating therefore Kenny's constraint for ability operators).

2.3 Extensions

Temporal operators. Broersen *et al.* [10] have added the temporal operators of linear-time temporal logic LTL to the stit language. In that language they introduce another modality of ability different from the above as $\Diamond\mathtt{CStit_i}\mathtt{X}\varphi$, where $\mathtt{X}$ is the temporal 'next' operator. They show that this definition of ability matches the ability operator of Pauly's coalition logic CL [38]. They also show that the further addition of the 'eventually' modality of LTL allows to reduce alternating-time temporal logic ATL [1] to that temporal extension of STIT.

Lorini recently extended the stit language by future tense and past tense operators and provided a complete axiomatization for this temporal extension of stit [31]. The semantics for temporal stit used by Lorini is based on the concept of temporal Kripke stit model which extends Zanardo's concept of Ockhamist model [50] with a choice component.

Ciuni and Zanardo extended the stit language by (restricted) branching-time operators of computational tree logic CTL and proved a completeness result [16].

Mental attitudes and deontic concepts. Starting with Kanger and Lindahl [30], many researchers working on logics of agency were interested in deontic concepts such as the obligation or the permission to act. Starting from the neighbourhood semantics for BIAT logic, Santos *et al.* added a modal operator of obligation $\mathtt{Obl}$ to the language [42, 43, 12]. Then the formula $\mathtt{OblBiat_i}\varphi$ expresses that agent i is obliged to bring it about that φ.

Horty proposed to integrate obligation into branching-time structures by means of a function *idl* which for every moment m selects the ideal histories among all the histories running through m: those where all the obligations are fulfilled [25].

$$M, h, m \models \mathtt{Obl}\varphi \quad \text{iff} \quad M, h', m \models \varphi \text{ for every } h' \text{ such that } h' \in idl(m).$$

Much less work was done on the integration of mental attitudes into logics of agency. For some first attempts see [49, 9]. More recently, some authors have worked on the combination of epistemic logic and STIT logic by enriching the STIT semantics with names for choices and action tokens [34, 27].

Resource-sensitive agency. In [39, 40], Porello and Troquard have proposed a variant of BIAT logic, where the modality of agency is used to formalise agents using, transforming, and producing consumable resources. Using Linear Logic in place of classical logic, one can write sentences like

$$(egg \otimes egg \otimes \mathtt{Biat_i}(egg \otimes egg \multimap omelet)) \multimap omelet,$$

saying that if agent i transforms two eggs into one omelet, and two eggs are available, then one omelet can be produced. On the other hand, *omelet* does not follow from $egg \otimes \mathtt{Biat_i}(egg \otimes egg \multimap omelet)$ as the resources are too few.

3 Action as 'means+result'

The preceding analysis of actions was merely in terms of their results. Another tradition studies not only the result, but also the means the agent employs to attain that result. The logical form of such sentences is "i brings it about that φ by doing α".

If we identify "i does α" as "i brings it about that ψ", for some appropriate proposition ψ, then we end up with an analysis of a dyadic agency operator, as studied by Segerberg [46].

We will not present that view in more detail here and just note that Segerberg's logic turns out to be an instance of the action theory that we are going to present now. Instead of identifying events and actions with propositions, that theory views them as 'things that happen', coming with some change in the world. It is then natural to interpret events and actions as *transitions* between possible worlds, just as computer programs running from an initial state to an end state. This view is taken by propositional dynamic logic PDL, which has *names* to identify these transitions. It is a view of action whose development has benefited from the synergies between philosophy and the formal science of computer programming.

The availability of names for actions allows to build complex actions from atomic actions. The latter may then be identified with *basic actions*: actions that make up an agent's repertoire. In practice, the choice of granularity for the set of these actions depends on the application at hand. While raising an arm could be taken as a basic action when modeling a voting procedure, a choreographer might want to decompose the raising of an arm into more basic performances of bodily movements.

In the interpretation of actions, an edge between two possible worlds may stand for two different things, depending on how the events of the world will unroll: first, it might be an *actual transition* corresponding to the event actually taking place; second, it might be a *possible transition* that does not actually occur. The logic that we are going to present now mainly adopts the latter perspective.

3.1 Propositional dynamic logic PDL

Standard PDL has names for events. In this section we describe an *agentive version* of PDL as used in several places in the artificial intelligence literature (e.g., [36, 23]. In that version, atomic actions take the form $i{:}\alpha$ where i is an agent and α is an atomic event. Complex actions —alias programs— are then built recursively from these atomic actions by means of the PDL connectives ";" (sequential composition), "$\cup$" (nondeterministic composition), "*" (iteration), and "?" (test). For instance, the complex event

$$\pi_1 = (\neg \mathit{treeDown}?; i{:}\mathbf{chop})^*; \mathit{treeDown}?$$

describes i's felling a tree by performing the atomic 'chop' action until the tree is down.

The language of PDL has modal operators $\mathtt{Poss}_\pi$ where i is an agent and π is an action. The formula $\mathtt{Poss}_\pi\varphi$ reads "there is a possible execution of π after which φ is true".[3] Due to indeterminism, there might be several possible executions of π. While $\mathtt{Poss}_\pi$ quantifies existentially over these executions, the dual modal operator $\mathtt{After}_\pi$ quantifies universally. It is definable from the former by $\mathtt{After}_\pi\varphi \overset{\text{def}}{=} \neg\mathtt{Poss}_\pi\neg\varphi$.

While in the 'action-as-result' view of BIAT and STIT logics actions are interpreted as propositions, in PDL an atomic action $i{:}\alpha$ is interpreted as a set of edges of the transition relation: there is an edge from world w_1 to world w_2 that is labeled $i{:}\alpha$ if it is possible to execute $i{:}\alpha$ in w_1 and w_2 is a possible outcome world. The set of all these edges makes up the accessibility relation $R_{i:\alpha}$ associated to $i{:}\alpha$. Complex actions are then interpreted by operations such as relation composition in the case of sequential composition ";" or set union in the case of nondeterministic composition "$\cup$". For instance, our example action π_1 is interpreted by the set of couples (w, w') such that one can go from w through finite **chop**-paths running through possible worlds satisfying $\neg \mathit{treeDown}$ and whose last possible world w' satisfies $\mathit{treeDown}$.

The formula $\mathtt{Poss}_\pi\varphi$ is true at a world w if there is a couple (w, w') in R_π such that φ is true at world w':

$$M, w \models \mathtt{Poss}_\pi\varphi \quad \text{iff} \quad M, w' \models \varphi \text{ for some } w' \text{ such that } wR_\pi w'.$$

[3] The standard notation is $\langle\pi\rangle\varphi$; we here deviate in order to be able to distinguish actual action from potential action.

The formula $\texttt{Poss}_\pi\varphi$ therefore expresses a weak notion of ability: the action π might occur and φ could be true afterwards. The modal operators $\texttt{Poss}_\pi$ are normal modal diamond operators. Hence the axiom $\texttt{Poss}_\pi(\varphi \vee \psi) \to \texttt{Poss}_\pi\varphi \vee \texttt{Poss}_\pi\psi$ is valid (violating therefore Kenny's principle for ability operators).

As we have announced above, Segerberg's dyadic agency operator can be viewed as an instantiation of PDL. His atomic events α take the form $\delta_i\psi$ where ψ is a proposition. In that framework he argues for principles such as transitivity: when i brings about φ_2 by bringing about φ_1 and i brings about φ_3 by bringing about φ_2, does i bring about φ_3 by bringing about φ_1? This can formally be written as $(\texttt{After}_{\delta_i\varphi_1}\varphi_1 \wedge \texttt{After}_{\delta_i\varphi_2}\varphi_3) \to \texttt{After}_{\delta_i\varphi_1}\varphi_3$.

3.2 Linear-time propositional dynamic logic PDL

Probably Cohen and Levesque were the first to adapt PDL in order to model actual agency [17]. The modalities are interpreted in *linear-time* PDL models: every world w has a unique history running through it. We distinguish modal operators of actual action by writing them as $\texttt{Happ}_\pi\varphi$, read "π is performed, and φ is true afterwards". Then the following principle for basic actions characterises linear PDL models:

$$(\texttt{Happ}_{i:\alpha}\top \wedge \texttt{Happ}_{j:\alpha'}\varphi) \to \texttt{Happ}_{i:\alpha}\varphi \tag{11}$$

Cohen and Levesque's linear PDL being only about actual action, Lorini and Demolombe [32] proposed a logic combining PDL operators of potential action $\texttt{Poss}_{i:\alpha}$ with linear PDL operators of actual action $\texttt{Happ}_{i:\alpha}$. In this logic, that we call here PDL with actual actions, the 'do implies can' principle takes the form of the valid schema for atomic actions:

$$\texttt{Happ}_{i:\alpha}\varphi \to \texttt{Poss}_{i:\alpha}\varphi. \tag{12}$$

Another logic which allows to represent both actual action and potential action is the *Dynamic Logic of Agency* (DLA) [24]. That logic combines linear PDL operators of actual action $\texttt{Happ}_{i:\alpha}$ with the historical possibility operator of STIT logic: potential action is expressed by the formula $\Diamond\texttt{Happ}_{i:\alpha}\varphi$ which has to be read "there is a possible history passing through the current moment such that agent i performs α, and φ is true afterwards".

An extension of DLA with program constructions of PDL, called Ockhamist PDL (OPDL), has been recently proposed in [3]. It is shown that both PDL and Full Computation Tree Logic CTL^* can be polynomially embedded into OPDL.

3.3 Extensions

PDL plus knowledge and belief. The first to add a modal operator of knowledge to a PDL-like logic was Moore [37]. This allowed him to formulate and study a principle of perfect recall (aka 'no forgetting') $\texttt{Know}_i\texttt{After}_\alpha\varphi \to \texttt{After}_\alpha\texttt{Know}_i\varphi$, as well as the converse principle of 'no miracles' (aka 'no learning'). Similar axioms for belief have also been studied in the literature, in particular under the 'denomination successor state axiom for knowledge' in artificial intelligence [44]. Principles of perfect recall and 'no miracles' play a central role in public announcement logic and more generally dynamic epistemic logics. These logics consider particular atomic events: announcements of (the truth of) formulas. Such events do not change the world, but only the agents' epistemic states. An overview of dynamic epistemic logics can be found in [18].

PDL plus obligations. Meyer's account extends PDL by a *violation constant* $\mathtt{V}$ that was first proposed by Anderson. Agent i's being forbidden to do basic action α is then reduced to all possible executions of α by i resulting in possible worlds where $\mathtt{V}$ is true; and i's permission to do α is reduced to some execution of α resulting in a possible world where $\mathtt{V}$ is false. In formulas:

$$\begin{array}{lclclcl} \mathtt{Perm}(i{:}\alpha) & \stackrel{\text{def}}{=} & \mathtt{Poss}_{i:\alpha}\neg\mathtt{V} & & & & \\ \mathtt{Forb}(i{:}\alpha) & \stackrel{\text{def}}{=} & \neg\mathtt{Perm}(i{:}\alpha) & \stackrel{\text{def}}{=} & \neg\mathtt{Poss}_{i:\alpha}\neg\mathtt{V} & \stackrel{\text{def}}{=} & \mathtt{After}_{i:\alpha}\mathtt{V} \end{array}$$

One may account for the obligation to perform an action by stipulating that every non-performance of α by i results in a violation state. It is however subject to debate how the complement of an action should be defined (see e.g. the discussion in [8]).

Linear PDL plus belief and intentions. Cohen and Levesque have analysed intention in linear PDL [17]. In their account intentions are defined in several steps from the concept of *strongly realistic preference*: among the worlds that are possible for an agent there is a subset the agent prefers. There is a modal operator $\mathtt{Pref}_i$ for each agent i, and $\mathtt{Pref}_i\varphi$ reads "i chooses φ to be true".[4] Such a notion of preference is strongly realistic in the sense that belief logically implies preference. Furthermore, there are the temporal operators "eventually" (noted $\mathtt{F}$), "henceforth" (noted $\mathtt{G}$), and "until" (noted $\mathtt{U}$) that are interpreted on histories of linear PDL models just as in linear-time temporal logic LTL.

The incremental construction is then as follows. (1) Agent i has the *goal* that φ if i prefers that φ is eventually true, formally $\mathtt{Goal}_i\varphi \stackrel{\text{def}}{=} \mathtt{Pref}_i\mathtt{F}\varphi$. (2) i has the *achievement goal* that φ if i has the goal that φ and believes that φ is currently false, formally $\mathtt{AGoal}_i\varphi \stackrel{\text{def}}{=} \mathtt{Goal}_i\varphi \wedge \mathtt{Bel}_i\neg\varphi$. (3) i has the *persistent goal* that φ if i has the achievement goal that φ and will keep that goal until it is either fulfilled or believed to be out of reach, formally $\mathtt{PGoal}_i\varphi \stackrel{\text{def}}{=} \mathtt{AGoal}_i\varphi \wedge (\mathtt{AGoal}_i\varphi)\,\mathtt{U}\,(\mathtt{Bel}_i\varphi \vee \mathtt{Bel}_i\mathtt{G}\neg\varphi)$. (4) i has the *intention* that φ if i has the persistent goal that φ and believes he can achieve that goal by an action of his. The formal definition requires quantification over i's actions; we do not go in the details here.

Lorini and Herzig [33] complemented Cohen and Levesque's approach by integrating the concept of an *attempt* to perform an action. The central principle there is "can and attempts implies does": if i intends to (attempt to) perform

[4] The original notation is $\mathtt{Choice}_i$ instead of $\mathtt{Pref}_i$, but we preferred to avoid any confusion with the concept of choice in stit theory.

α and α is feasible then α will indeed take place. This principle is a sort of converse to the 'do implies can' principle.

References and recommended readings

[1] Rajeev Alur, Thomas A. Henzinger, and Orna Kupferman. Alternating-time temporal logic. In *Proceedings of the 38th IEEE Symposium on Foundations of Computer Science*, Florida, October 1997.

[2] Philippe Balbiani, Andreas Herzig, and Nicolas Troquard. Alternative axiomatics and complexity of deliberative STIT theories. *Journal of Philosophical Logic*, 37(4):387–406, 2008.

[3] Philippe Balbiani and Emiliano Lorini. Ockhamist propositional dynamic logic: A natural link between PDL and CTL*. In *Proceedings of the 20th International Workshop on Logic, Language, Information, and Computation (WOLLIC 2013)*, volume 8071 of *Lecture Notes in Computer Science*, pages 251–265. Springer, 2013.

[4] $***$ Nuel Belnap, Michael Perloff, and Ming Xu. *Facing the Future: Agents and Choices in Our Indeterminist World.* Oxford University Press, Oxford, 2001. [A compilation of a over a decade of work of the authors on agency in branching-time.].

[5] Nuel Belnap. Backwards and forwards in the modal logic of agency. *Philosophy and Phenomenological Research*, 51(4):777–807, 1991.

[6] Nuel Belnap and Michael Perloff. Seeing to it that: A canonical form for agentives. *Theoria*, 54(3):175–199, 1988.

[7] Michael E. Bratman. *Intentions, plans, and practical reason.* Harvard University Press, MA, 1987.

[8] Jan Broersen. *Modal Action Logics for Reasoning about Reactive Systems.* PhD thesis, Vrije Universiteit Amsterdam, Amsterdam, January 2003.

[9] Jan Broersen. Making a start with the stit logic analysis of intentional action. *Journal of Philosophical Logic*, 40:399–420, 2011.

[10] Jan Broersen, Andreas Herzig, and Nicolas Troquard. Embedding Alternating-time Temporal Logic in strategic STIT logic of agency. *Journal of Logic and Computation*, 16(5):559–578, 2006.

[11] Mark A. Brown. Normal bimodal logics of ability and action. *Studia Logica*, 52:519–532, 1992.

[12] José Carmo and Olga Pacheco. Deontic and action logics for organized collective agency, modeled through institutionalized agents and roles. *Fundamenta Informaticae*, 48, 2001.

[13] Brian F. Chellas. *The Logical Form of Imperatives.* Perry Lane Press, Stanford, CA, 1969.

[14] Roderick M. Chisholm. The descriptive element in the concept of action. *Journal of Philosophy*, 61:613–624, 1964.

[15] Roberto Ciuni and Emiliano Lorini. Comparing semantics for temporal STIT logic. *Logique et Analyse*, 2017.

[16] Roberto Ciuni and Alberto Zanardo. Completeness of a branching-time logic with possible choices. *Studia Logica*, 96(3):393–420, 2010.

[17] Philip R. Cohen and Hector J. Levesque. Intention is choice with commitment. *Artificial Intelligence*, 42(2–3):213–261, 1990.

[18] H. P. van Ditmarsch, W. van der Hoek, and B. Kooi. *Dynamic Epistemic Logic.* Kluwer Academic Publishers, 2007.

[19] Dag Elgesem. *Action theory and modal logic.* PhD thesis, Institut for filosofi, Det historiskfilosofiske fakultetet, Universitetet i Oslo, 1993.

[20] Dag Elgesem. The modal logic of agency. *Nordic J. Philos. Logic*, 2(2):1–46, 1997.

[21] $***$ Guido Governatori and Antonino Rotolo. On the axiomatization of elgesem's logic of agency and ability. *J. of Philosophical Logic*, 34:403–431, 2005. [A semantics for the logic of bringing-it-about-that in terms of neighbourhood frames.].

[22] $***$ David Harel, Dexter Kozen, and Jerzy Tiuryn. *Dynamic Logic.* MIT Press, 2000. [A standard textbook for dynamic logics.].

[23] Andreas Herzig and Dominique Longin. C&L intention revisited. In Didier Dubois, Chris Welty, and Mary-Anne Williams, editors, *Proc. 9th Int. Conf. on Principles on Principles of Knowledge Representation and Reasoning(KR2004)*, pages 527–535. AAAI Press, 2004.

[24] Andreas Herzig and Emiliano Lorini. A dynamic logic of agency I: STIT, abilities and powers. *Journal of Logic, Language and Information*, 19:89–121, 2010.

[25] $***$ John F. Horty. *Agency and Deontic Logic.* Oxford University Press, 2001. [A thorough analysis of obligations to do in the models of branching-time and choice of agents.].

[26] John Horty and Nuel Belnap. The deliberative stit: a study of action, omission, ability and obligation. *Journal of Philosophical Logic*, 24(6):583–644, 1995.

[27] John Horty and Eric Pacuit. Action Types in Stit Semantics. *Review of Symbolic Logic*, 2017.

[28] Stig Kanger and Helle Kanger. Rights and Parliamentarism. *Theoria*, 32:85–115, 1966.

[29] Anthony Kenny. *Will, Freedom, and Power*. Oxford: Blackwell, 1975.

[30] Lars Lindahl. *Position and change: A study in law and logic*. D. Reidel publishing Company: Dordrecht, 1977.

[31] Emiliano Lorini. Temporal STIT logic and its application to normative reasoning. *Journal of Applied Non-Classical Logics*, 23(4):372–399, 2013.

[32] Emiliano Lorini and Robert Demolombe. Trust and norms in the context of computer security: toward a logical formalization. In R. Van der Meyden and L. Van der Torre, editors, *Proceedings of the International Workshop on Deontic Logic in Computer Science (DEON 2008)*, volume 5076 of *LNCS*, pages 50–64. Springer-Verlag, 2008.

[33] Emiliano Lorini and Andreas Herzig. A logic of intention and attempt. *Synthese KRA*, 163(1):45–77, 2008.

[34] Emiliano Lorini, Dominique Longin, and Eunate Mayor. A logical analysis of responsibility attribution : emotions, individuals and collectives. *Journal of Logic and Computation*, 24(6):1313–1339, 2014.

[35] John-Jules Ch. Meyer. A different approach to deontic logic: Deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic*, 29:109–136, 1988.

[36] John-Jules Ch. Meyer, Wiebe van der Hoek, and Bernardus van der Linder. A logical approach to the dynamics of commitments. *Artificial Intelligence*, 113(1-2):1–40, 1999.

[37] Robert C. Moore. A formal theory of knowledge and action. In J.R. Hobbs and R.C. Moore, editors, *Formal Theories of the Commonsense World*, pages 319–358. Ablex, Norwood, NJ, 1985.

[38] $***$ Marc Pauly. A modal logic for coalitional power in games. *Journal of Logic and Computation*, 12(1):149–166, 2002. [A now classic article on group abilities in game and social choice theory.].

[39] Daniele Porello and Nicolas Troquard. A resource-sensitive logic of agency. In *ECAI 2014 - 21st European Conference on Artificial Intelligence*, pages 723–728, 2014.

[40] Daniele Porello and Nicolas Troquard. Non-normal modalities in variants of linear logic. *Journal of Applied Non-Classical Logics*, 25(3):229–255, 2015.

[41] Ingmar Pörn. *Action Theory and Social Science: Some Formal Models*. Synthese Library 120. D. Reidel, Dordrecht, 1977.

[42] Felipe Santos, Andrew J.I. Jones, and José Carmo. Action concepts for describing organised interaction. In R.H. Sprague, editor, *Proc. Thirtieth Annual Hawaii International Conference on System Sciences (HICSS-30)*, volume 5, pages 373–382. IEEE Computer Society Press, 1997.

[43] Felipe Santos, Andrew J.I. Jones, and José Carmo. Responsibility for action in organisations: a formal model. In G. Holmström-Hintikka and R. Tuomela, editors, *Contemporary Action Theory*, volume 1, pages 333–348. Kluwer, 1997.

[44] Richard Scherl and Hector J. Levesque. The frame problem and knowledge producing actions. *Artificial Intelligence*, 144(1-2), 2003.

[45] Krister Segerberg, editor. *"Logic of Action": Special issue of Studia Logica*, volume 51:3/4. Springer Netherlands, 1992.

[46] Krister Segerberg. Two traditions in the logic of belief: bringing them together. In Hans Jürgen Ohlbach and Uwe Reyle, editors, *Logic, Language and Reasoning: essays in honour of Dov Gabbay*, volume 5 of *Trends in Logic*, pages 135–147. Kluwer Academic Publishers, Dordrecht, 1999.

[47] Nicolas Troquard. Reasoning about coalitional agency and ability in the logics of "bringing-it-about". *Autonomous Agents and Multi-Agent Systems*, 28(3):381–407, 2014.

[48] Georg H. Von Wright. *Norm and Action. A Logical Inquiry.* Routledge and Kegan Paul, London, 1963.

[49] Heinrich Wansing and Caroline Semmling. From BDI and stit to bdi-stit logic. *Logic and Logical Philosophy*, 17:185–207, 2008.

[50] Alberto Zanardo. Branching-time logic with quantification over branches: the point of view of modal logic. *Journal of Symbolic Logic*, 61:1–39, 1996.