On the Reconciliation of Logics of Agency and Logics of Event Types

Jan Broersen

Abstract This paper discusses Segerberg's view on agency, a view that is heavily influenced by his thinking about dynamic logic. The main work that puts forward Segerberg's ideas about agency is *Outline of a logic of action*. That article attempts to reconcile the *stit* view of agency with the dynamic logic view of event types. Here I discuss Segerberg's proposal. I will argue that the theory lacks some detail and explanatory power. I will suggest an alternative theory based on an extension of the logic XSTIT. Recently, the subject discussed here has attracted renewed attention of several researchers working in computer science and philosophy.

1 Introduction

Over the last 30 years, two different views on the logic of action have emerged in the computer science and philosophical literature. The first view comes from computer science, and I will call it the 'event type' approach. In this view the structures the logic talks about are labeled transitions systems, where the labels denote a type of event (think of a database update, a register update, a variable assignment, etc). Examples of formalisms of this kind are Hennessey-Milner logic [19], dynamic logic (DL) [21] and process logic [14], but I also take the situation calculus [18] to belong to this branch. The other kind of action formalism originates in philosophy, and focusses on the modeling of agency, that is, on the formal modeling of the connection between agents and the changes in the world they can be held responsible for. In this type of formalism, the structures are choice structures. Examples of formalisms of this kind are 'Bringing It About Logic' (BIAT) [17], *stit*-logic [3], Coalition Logic (CL) [20],

J. Broersen (✉)
Department of Information and Computing Sciences, Utrecht University,
3508 UTRECHT, Netherlands
e-mail: J.M.Broersen@uu.nl

R. Trypuz (ed.), *Krister Segerberg on Logic of Actions*,
Outstanding Contributions to Logic 1, DOI: 10.1007/978-94-007-7046-1_3,

and Alternating time Temporal Logic (ATL) [1] and Brown's logic of ability (which is a predecessor to CL and ATL) [9].

Many authors have sought to combine both views on action. Examples are the work of Herzig and Lorini [15], and the work of Xu [29]. Combining the computer science view on action (but from now on I will refer to this view as the event-type view) and the philosophical, agency-oriented view is of central importance to the understanding of the relation between computation and agency, and thus, it seems safe to claim, to the understanding of the possibilities of Artificial Intelligence.

Krister Segerberg, being one of the central researchers working on action formalisms at the time of their emergence, describes the problem as follows in *Outline of a logic of action* [26] (which extends [25] and is the culmination of ideas first put forward in *Bringing it about* [23] and *Getting started: Beginnings in the logic of action* [24].): "to combine action logic in the Scott/Chellas/Belnap tradition with Pratt's dynamic logic". In *Outline of a logic of action* Segerberg then puts forward a language, a class of structures and a semantics whose main aim is to reconcile the two different views on the logic of action.

Here I will explain and discuss Segerberg's theory of agency and action as put forward in *Outline of a logic of action*. In explaining and discussing this work, I will point to the places where I do not agree with the modeling choices made by Segerberg. Then, to explain my view on the matter in a coherent way, I will put forward my own outline of a theory of action.

That there is a problem to be solved here shines through clearly if we look at the practice of computer scientists to claim that agency in dynamic logic is modeled sufficiently by annotating event types with agents or groups of agents. However, this practice does not explain the logical differences between an action a performed by agent 1, an action a performed by agent 2 and an action a performed by agents 1 and 2 together. For instance, in a dynamic logic theory with event types annotated with agents and groups of agents, it is entirely unclear if there are logical relations between $[a_{ag_1}]\varphi$, $[a_{ag_2}]\psi$ and $[a_{ag_1,ag_2}]\chi$, and if there are such relations, it is unclear what they are (e.g.: since all formulas concern the same event type a, should there be a logical relation between the three formulas? What axioms describe this relation? If there is no such relation, then why introduce event type notations in the formulas at all?).

2 Segerberg's Action Theory

In *Outline of a logic of action*, Segerberg puts forward the following syntax for his unifying action formalism.

Definition 2.1 Given a countable set of atomic proposition letters P and $p \in P$, and given a countable set Ags of agent names, and $i \in Ags$, the formal language $\mathcal{L}_{\mathsf{SEG}}$ is:

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid [H\psi]\varphi \mid [F]\varphi \mid [P]\varphi \mid [NEXT:\psi]\varphi \mid [LAST:\psi]\varphi \mid$$
$$does_i(\alpha) \mid done_i(\alpha) \mid reals_i(\alpha) \mid realled_i(\alpha) \mid occs(\alpha) \mid occed(\alpha)$$
$$\alpha := \alpha;\beta \mid \delta_i\varphi \mid \epsilon\varphi$$

The reading of the event type terms is as follows:

$$\alpha;\beta = \text{the composite event type of beta after alpha}$$
$$\delta_i\varphi = \text{agent } i \text{ bringing it about that } \varphi \text{ (see [23] and [27])}$$
$$\epsilon\varphi = \text{the coming about of } \varphi$$

The reading of the modalities is as follows:

$$[H\psi]\varphi = \varphi \text{ holds for all histories for which } \psi$$
$$[F]\varphi = \text{henceforth } \varphi$$
$$[P]\varphi = \text{it has always been the case that } \varphi$$
$$[NEXT:\psi]\varphi = \text{next time that } \psi, \varphi \text{ holds}$$
$$[LAST:\psi]\varphi = \text{last time that } \psi, \varphi \text{ held}$$
$$does_i(\alpha) = \text{agent } i \text{ does an event of type } \alpha$$
$$done_i(\alpha) = \text{agent } i \text{ just did an event of type } \alpha$$
$$reals_i(\alpha) = \text{agent } i \text{ realizes an event of type } \alpha$$
$$realled_i(\alpha) = \text{agent } i \text{ just realized an event of type } \alpha$$
$$occs(\alpha) = \text{an event of type } \alpha \text{ occurs}$$
$$occed(\alpha) = \text{an event of type } \alpha \text{ just occured}$$

It has to be emphasized that all readings are relative to a state and a history (which is a 'timeline' extending infinitely into the past and the future). So 'always in the future' means always in the future *on the current history of evaluation*, and does *not* mean 'always in the future independent of whatever agents will do or whatever events will occur'. So, like in *stit* theory, Segerberg takes the Ockhamist approach to future contingencies [22], which means that truth of formulas is relative to a history. This means that histories or paths are viewed as possible worlds. That insight is essential. In his semantics Segerberg uses triples $\langle h, u, g\rangle$, where u is the current state, h a path from u into the past, and g a path from u into the future. He calls triples $\langle h, u, g\rangle$ 'articulated histories'. All truth conditions are relative to articulated histories.

Figure 1 shows how in Segerberg's framework histories are build from sequences of actions, pictured as triangles. Histories are defined as maximal sets of subsequent actions.

I will not give the formal definitions of the models and the truth conditions, since the semantics is easy to describe in terms of pictures and natural language. In Fig. 1 we see the actions depicted as triangles. In the formal semantics, an action is a triple (i, a, p), where i is an agent, a is an event type, and p a finite sequence of states

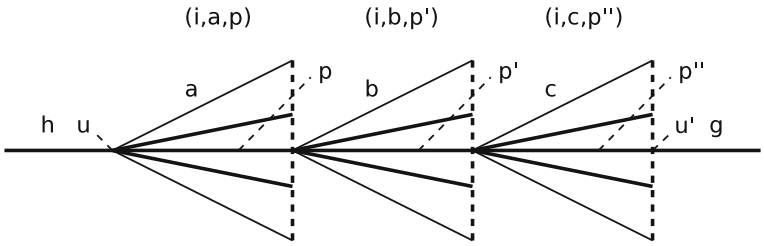

Fig. 1 Histories as sequences of actions in Segerberg's action semantics

representing the way the agent i performs the event of type a. In the picture, events are represented by triangles build from two fine lines and one interrupted line. We see three events, one of type a, one of type b and one of type c. The sub-triangles build from thick lines and parts of the interrupted lines represent the 'ways and means' in which agent i can perform the event types that are associated with the bigger triangles. In terms of this picture the semantics of the action operators is easy to explain. Events of type $\epsilon\varphi$ are those for which φ holds on all the points depicted by the interrupted line of a triangle. Events of type $\delta_i\varphi$ are those for which φ holds on all the points depicted by the interrupted line on a triangle for agent i. Now, $does_i(\alpha)$ holds at a point on the history just in case the event types that agent i 'does' are those interpreting α. For 'doing', the history of evaluation must be contained inside the inner triangles representing the agent's way and means to do the event of the given type. For instance, in point u in the picture it holds that $does_i(a; b; c)$. In point u' in the picture it holds that $done_i(a; b; c)$. But, also we have that in point u in the picture it holds that $reals_i(a; b; c)$ and in point u' that $realled_i(a; b; c)$. For 'realizing' the truth condition is only weaker, and the history of evaluation must run through the outer triangle. It is clear then that one validity of the logic is that doing an event of a given type implies realizing that same event. But the other way around does not hold, which is exemplified by cases where the history of evaluation runs through the part of the bigger triangle that is not included in the smaller triangle. The interpretations of $occs(\alpha)$ and $occed(\alpha)$ are similar to those of $reals_i(\alpha)$ and $realled_i(\alpha)$, the difference being that the agents are quantified out. The interpretation of the modalities $[F]\varphi$, $[P]\varphi$, $[NEXT : \psi]\varphi$ and $[LAST : \psi]\varphi$ is straightforward given their informal reading and the fact that their formal interpretation is relative to individual histories. The interpretation of $[H\psi]\varphi$ is clarified ones we realize that histories like the one depicted in Fig. 1 are elements of trees resulting from the fact that in each state agents in general have multiple event types to choose from and for each event type in general have multiple ways to perform them. Action 'trees' (maybe 'bundles' is a better word) are sets of histories closed under these alternatives for the agents. Now $[H\psi]\varphi$ is true in a state on a history if on all alternative continuations of the path from the past that satisfy ψ, also φ is true.

2.1 Realizing Versus Doing

As explained above, the theory distinguishes between doing and realizing. In particular, $\langle h, u, g\rangle \models does_i(\alpha)$ holds if in state u, along the future history g, along the first part agent i does an event of the type α. In terms of the triangle-based picture of Fig. 1: if the 'inner' triangles (that is, i's possible ways to perform and event of the given type) along future history g are those interpreting α. The semantics of an agent i *realizing* events of a given type is slightly different. $\langle h, u, g\rangle \models reals_i(\alpha)$ holds if in state u, along the future history g, in the first part the agent i realizes an event of the type α. In terms of the triangle-based picture of Fig. 1: if the 'outer' triangles (that is, the events of a given type) along future history g are part of 'outer' triangles that interpret α.

By introducing the difference between realizing and doing, Segerberg aims to accommodate the intuition that an agent can be part of an activity without really contributing to that activity. However, the theory lacks explanatory power here. In what sense can an agent be part of an activity without contributing to it? What is the exact sense in which the agent is still connected to an activity if it is not that it is in some sense responsible for that activity? In *stit* theory, these issue have been resolved quite satisfactory. Either an agent ensures that a condition occurs, or it allows for the negation of that condition to occur. So, by refraining to see to it that the negation of a condition occurs, an agent can play a role in an activity without being the 'author' (as Segerberg puts it) of that activity.

A related problem for the theory is that it does not allow for indeterminism. At least, if it does, it is unclear how. On the one hand it is able to define the *stit* operator (as I will explain later on), so it seems there should be a notion of non-determinism in the system. However, the theory talks about 'ways and means' for the performance of actions as if these are procedures to choose from for the agent. So, it sees the different possible ways of performing an event of a certain type as a choice that is fully under control of the agent, not leaving room for non-determinism.

2.2 More Actions at Once

I think a theory of action should allow for the possibility of single agents performing more than one action at the same time. And indeed, it seems to me that in Segerberg's theory the situation can be as in Fig. 2 (a picture that I will use later on to explain the simulation of *stit* semantics in the theory). The picture shows how along the current history an agent does both an event of type a and an event of type b. However, it is not completely clear if it is actually Segerberg's intention to allow for these situations. For instance can an agent at the same time *do* an event of type a and only *realize* an event of type b? And what would that mean? For instance, would the event of type b be an unintended side effect of the responsibility for event a?

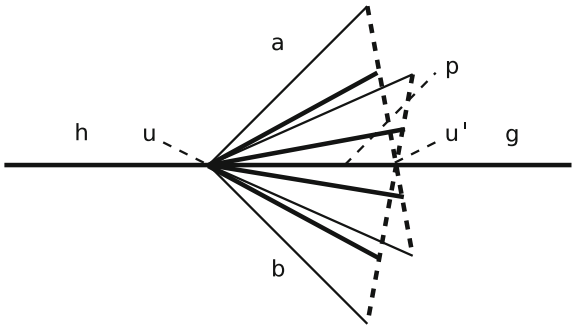

Fig. 2 Explanation of *stit* simulation in Segerberg's action semantics

2.3 Multi-agency and Collective Agency

Branching in action trees is defined in terms of closure of branching under the different possibilities individual agents have to perform an event of a certain type. But the basic theory does not give an answer to how simultaneous actions of different agents relate to each other. As Segerberg admits on page 381, in the base theory, only one agent can act at the time. However, for a theory of agency and action it seems important to ask whether or not one agent, by performing an action, can prevent another agent from performing his action simultaneously. That is, is there, or should there be, a notion of independence of agency like in *stit* theory? A related problem is that it is not clear how branching can occur as the result of collective action. However, Segerberg discusses an approach to this problem later on in the paper (page 375). The idea is to make the ways and means function relative to collectives of agents in stead of individual agents. I believe this is a correct idea, and I will use a closely related idea in my own theory in Sect. 3.

2.4 The Generalization to Complex Action

On page 371 Segerberg discusses a possible generalization of the action theory by allowing regular operations on event types, as in dynamic logic. The suggestion is that this generalization is easy and that, for instance, $\langle h, u, g\rangle \models does_i(\alpha; \gamma \cup \beta; \gamma)$ holds if the first part of the history g is either of the type $\alpha; \gamma$ or of the type $\beta; \gamma$. But, I disagree with that semantics. Assume that indeed in state u there is an alternative of the type β. But also assume that if the agent would have performed an action of type β, afterwards it would not have been possible to do an action of type γ. Then, is it still justified to say that the agent does an action of the type $(\alpha; \gamma \cup \beta; \gamma)$? So, in my opinion, checking if an agent does an action that corresponds with a complex event of the type $\alpha; \gamma \cup \beta; \gamma$ should involve checking that if it would have been β that was performed at the time of choice, afterwards γ is still a possible continuation. The underlying problem is, I believe, that complex actions cannot be interpreted in

an Ockhamist way relative to a single history only, because as soon as there is a non-deterministic choice involved (as the result of introducing a binary choice operation $\cup$ and/or the Kleene star *), also alternative histories will have to be considered to determine whether or not a complex action (i.e., a strategy) is actually performed. Also the idea of an agent performing an event of a type that is indeterministic needs much more clarification. What does this non-determinism represent? Uncertainty or practical ignorance on the part of the agent? Lack of agentive control? Intrinsic indeterminedness of the environment? Also, confinement of indeterminism to operators like $\cup$ and * suggests that we can explicitly point to the indeterminism in agency by specifying it in non-deterministic programs. But, in my opinion non-deterministic programs fall far short as an adequate model for agency.

2.5 Simulation of the Stit Operator

Segerberg argues that in his theory the Chellas stit operator is definable through the following definition.

$$[i \ \mathsf{cstit}]\varphi \equiv_{def} realled_i(\delta_i\varphi)$$

I will explain this definition using Fig. 2. The definition says (implicitly, using a function $\mathbf{D}(i, P)$ whose formal definition I do not give here) that an agent sees to it that φ if and only if agent i just 'realized' an action (i, a, p) for which it is true that it ensured the outcome φ independent of how the event a was 'done'. In terms of Fig. 2, the *stit* semantics is then as follows. In state u' (and *not* in u) along articulated history $\langle hp, u', g\rangle$, the agent sees to it that φ if either the event of type a or the event of type b (and we assume here that these are the only two events for which the agent i in state u' has a 'way or means' to perform it through p) have as a guaranteed result that φ holds (that is, on each point of at least one of the two dotted lines in the figure, φ must hold).

In my opinion there are three problems with this definition. The first is that since it is unclear what intuitions are behind the distinction between realizing and doing, it is also unclear why the *stit* operator is not defined in such a way that the condition φ is ensured independent of how the agent *does* the event (in stead of guaranteeing the outcome independent of how the agent *realizes* the event). That is, it is not clear why the definition could not be $[i \ \mathsf{cstit}]\varphi \equiv_{def} done_i(\delta_i\varphi)$. In terms of Fig. 2 this would amount to the condition that for at least one the two triangles, only the points on the interrupted line part belonging to the inner triangle satisfy φ.

The second problem is the existential quantification that is implicit in Segerberg's version of the *stit* operator. An agent can only see to a condition φ if *there is* an event of a certain type serving as a witness for this. This means that the truth condition for the modal *stit* operator defined in this way has an $\exists - \forall$ structure, which implies that the operator will be weak and will *not* satisfy the agglomeration schema $[i \ \mathsf{cstit}]\varphi \wedge [i \ \mathsf{cstit}]\psi \rightarrow [i \ \mathsf{cstit}](\varphi \wedge \psi)$ (in terms of the picture in Fig. 2: if φ is true on all points of the interrupted line for event of type a and ψ is true on all points of

the interrupted line for event of type b, the antecedent of the axiom is true while the consequent is not). However, all *stit* operators in the literature are normal and do satisfy this schema. Another problem with linking agency with the existence of events of a certain type is that events and their types are taken as the starting point for defining agency. I believe this should be the other way around: theories of agency can be used to understand and define the nature of events and their types. This is the approach I will take in Sect. 3.

The third problem for the encoding of the *stit* operator in the theory is that it is unclear how the central *stit* ideas about non-determinism take form in the models. The central idea of *stit* theory is that seeing to it that a condition holds is the same as ensuring that condition irrespective of the non-determinism of the environment (which includes the simultaneous choices of other agents). Now, saying that φ has to hold on all 'realization-alternative' outcomes (that is, the alternatives within the outer triangles) of the realization of an event of a certain type can hardly be seen as ensuring φ modulo *non-determinism* as in *stit* theory. But also if Segerberg would demand that φ would be true on all 'execution-alternative' outcomes (that is, the alternatives within the inner triangles), the semantics would not be one based on the *stit* idea of ensuring a condition modulo non-determism.

2.6 Simulation of the Dynamic Logic Operator

The simulation of the standard basic dynamic logic modality in the theory is as follows:

$$[\alpha]\varphi \equiv_{def} [H : occs(\alpha)][NEXT : occed(\alpha)]\varphi$$

I believe this simulation is intuitive and correct. It defines the modality $[a]\varphi$ as "directly after all possible continuations of the type α it holds that φ". However, it is important to bear in mind that evaluation is still relative to individual articulated histories. And in case $[\alpha]\varphi$ is true on an articulated history $\langle h, u, g\rangle$, it does *not* follow that along the future history g of that same articulated history, the agent i performs an event of type a resulting in φ. The right interpretation is: "for all possible future histories g' whose first part is an event of the type α, immediately after α is finished, φ is true". So it can be that α does not occur on the articulated history $\langle h, u, g\rangle$ relative to which is evaluated. For instance, in Fig. 3 it holds that $\langle h, u, g\rangle \models [a]\varphi$ and $\langle h, u, g'\rangle \models [a]\varphi$ and $\langle h, u, g''\rangle \models [a]\varphi$ in case φ holds on all points of the interrupted line belonging to the triangle of the event of type a. The reason for this interpretation is that $[\alpha]\varphi$ is what Prior [22] calls a Peircian temporal operator, while Segerberg's base semantics is, in Prior's terminology, Ockhamist.

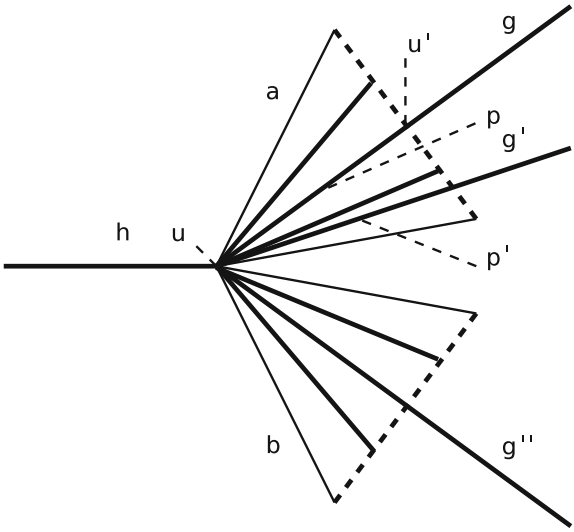

Fig. 3 Explanation of dynamic logic action type simulation in Segerberg's action semantics

3 Outline of an Alternative Theory of Action

I will now put forward an alternative outline for a theory of action. I will take the logic XSTIT as the base logic of agency and add a new operator to it that will enable me to simulate logics of event types (like dynamic logic) within the *stit* framework. This simulation requires a different view on the relation between event types and agency then the one put forward by Segerberg.

It is important, I think, to emphasize that in dynamic logic event types describe characteristics of *transitions*. In dynamic logic, if two transitions are of the same type, they have the same event type name, and using the logic we can specify that they have the same pre- and post-condition relation. For instance, if we want to specify that a-events are of the type whose instances have as a sufficient precondition ψ relative to the postcondition φ, we can write $\psi \rightarrow [a]\varphi$. The semantics of dynamic logic interprets these formulas in a transition system where a transition can only be of type a if in case it is a transition from a state where ψ it leads to a state where φ. As an example we might take the event type of "the closing of a door". Precondition is the door being open, postcondition is the door being closed. If as the result of agentive effort a door is moving from an open position to a closed position, the agent performs an action that is of the type "the closing of a door". But note that that same action or event might also have other types, such as "spending energy", or "producing a slamming noise", etc. But also note that it might take two or more agents to close a door (it might be very heavy). In that case the event of type "closing the door" cannot be linked to one agent exclusively.

The above described view on the relation between agency and event types will be the point of departure for the theory I will put forward here. It will be convenient to

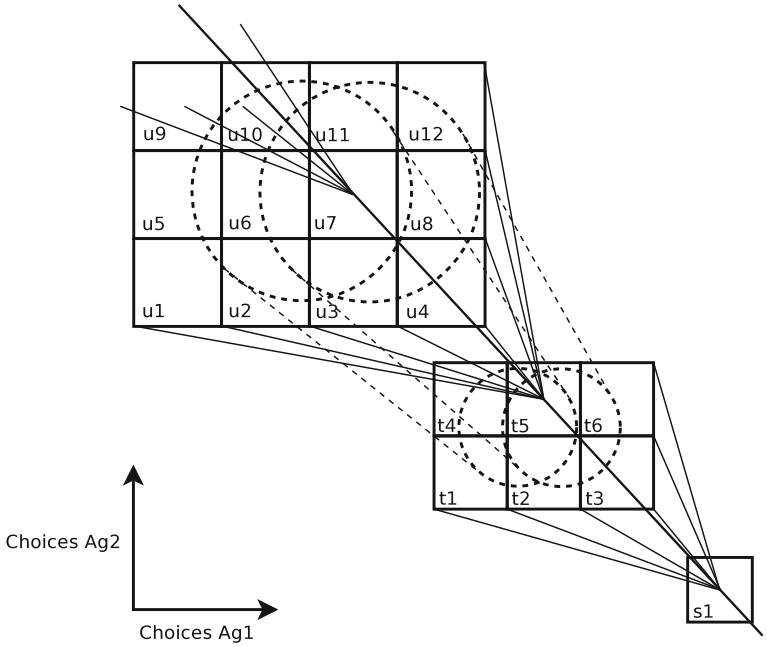

Fig. 4 Event types (*the dotted cylinders*) versus multi-agent choices (the game form structures)

see the central idea in a picture. In Fig. 4 we see the described view on event types pictured inside an XSTIT frame fragment. The XSTIT part gives the states, histories and choices for two agents. Three choice situations s, t and u along one central history are pictured, where in the first situation s, no genuine choices are possible. Of course, the central history (or more correct 'history bundle') pictured is only one of the many histories (bundles) that may result from the genuine choices the two agents have in t and u; the tree of possible histories is closed under the choices that are possible in the different situations. I will extend XSTIT frames to XSTIT.ET frames by adding event types. In the picture these appear as the cylinders build from interrupted line elements. For instance, the right cylinder might picture transitions of type b and the left cylinder transitions of type a. This set-up allows for: (1) different transitions for different situations throughout the frame being of the same event type, (2) single choices realizing more than one event type, and (3) events of a given type for which it takes strictly more than one agent to perform them.

With this conceptualization we arrive at the following ontology. Events are transitions at specific situations at specific moments in time. Actions are events that occur due to agentive involvement of agents.[1] Different agents at different times in

[1] Actually, in the present set-up the difference between events an actions is vacuous, since all transitions in the frames are due to agents. And seeing 'nature' as just another agent is problematic, since it seems natural to demand that nature does not have genuine choices.

different situations can execute an event of the same type. So, an event can be of a certain type α. But, it is never the case that α is the denotation of an action itself (as computer scientists are sometimes inclined to think).

3.1 The Logic XSTIT.ET

I will define a logic with the acronym XSTIT.ET by extending my earlier definitions for the logic XSTIT (first put forward in [7] and corrected, adapted and extended in various ways in [5], [6] and [8]). The characters ET stand for 'Event Types'. The modal language of XSTIT.ET is given by the following definition:

Definition 3.1 Given a countable set of propositions P and $p \in P$, a finite set Ags of agent names with $A \subseteq Ags$, and a countable set of event type names Et with $a \in Et$, the formal language $\mathcal{L}_{\mathsf{XSTIT.ET}}$ is:

$$\varphi := p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid [A\ xstit]\varphi \mid X\varphi \mid [A\ perf\ a]$$

Besides the usual propositional connectives, the syntax of XSTIT.ET comprises four modal operators. The operator $\Box\varphi$ expresses 'historical necessity', and plays the same role as the well-known path quantifiers in logics such as CTL and CTL* [12]. Another way of talking about this operator is to say that it expresses that φ is 'settled'. However, settledness does *not* necessarily mean that a property is *always* true in the future (as often thought). Settledness may, in general, apply to the condition that φ occurs 'some' time in the future, or to some other temporal property. This is reflected by the fact that settledness is interpreted as a universal quantification over the *branching* dimension of time, and *not* over the dimension of duration. The operator $[A\ \mathsf{xstit}]\varphi$ stands for 'agents A jointly see to it that φ in the next state'. The third modality is the next operator $X\varphi$. It has a standard interpretation as the transition to a next system state. The new operator introduced in this context is $[A\ \mathsf{perf}\ a]$. It expresses that the group of agents A *performs* an event of the type a.

To give a formal interpretation to the new operator $[A\ \mathsf{perf}\ a]$ we extend XSTIT frames (in their version using functions in stead of relations) with a function A returning the event types of a transition between two subsequent states.

Definition 3.2 An XSTIT.ET-frame is a tuple $\langle S, H, A, E\rangle$ such that[2]:

1. S is a non-empty set of static states. Elements of S are denoted s, s', etc.
2. H is a non-empty set of possible system histories isomorphic to infinite sequences $\ldots s_{-2}, s_{-1}, s_0, s_1, s_2, \ldots$ with $s_i \in S$ for $i \in \mathbb{Z}$. Elements of H are denoted h, h', etc. We denote that s' succeeds s on the history h by $s' = succ(s, h)$ and by $s = pred(s', h)$. We have the following bundeling constraint on the set H:

[2] In the meta-language we use the same symbols both as constant names and as variable names, and we assume universal quantification of unbound meta-variables.

a. if $s \in h$ and $s' \in h'$ and $s = s'$ then $pred(s, h) = pred(s, h')$

3. $A : S \times S \mapsto 2^{Et}$ is a function mapping subsequent states to a set of basic event types characterizing the transition between the two states. We have the following constraints on the function A:

   a. $A(s, t) = \emptyset$ if there is no $h \in H$ with $s \in h$ and $t \in h$ and $t = succ(s, h)$
   b. for any $h \in H$ and $h' \in H$: if $s \in h$ and $s' \in h'$ and $s = s'$ then $A(pred(s, h), s) = A(pred(s', h'), s')$

4. $E : S \times H \times 2^{Ags} \mapsto 2^S$ is an h-effectivity function yielding for a group of agents A the set of next static states allowed by the simultaneous choices exercised by the agents relative to a history. On the function E we have the following constraints:

   a. if $s \notin h$ then $E(s, h, A) = \emptyset$
   b. $succ(s, h) \in E(s, h, A)$
   c. $\exists h : s' = succ(s, h)$ if and only if $\forall h$: if $s \in h$ then $s' \in E(s, h, \emptyset)$
   d. if $s \in h$ then $E(s, h, Ags) = \{succ(s, h)\}$
   e. if $A \supset B$ then $E(s, h, A) \subseteq E(s, h, B)$
   f. if $A \cap B = \emptyset$ and $s \in h$ and $s \in h'$ then $E(s, h, A) \cap E(s, h', B) \neq \emptyset$

In definition 3.2 above, we refer to the states s as 'static states'. This is to distinguish them from what we call 'dynamic states', which are combinations $\langle s, h \rangle$ of static states and histories. Dynamic states will function as the elementary units of evaluation of the logic. This is very much like in Segerberg's semantics, the only difference being that we do no articulate the past of a history. We do not need to refer to the past in our models, since we do not have backwards looking operators in the logical language.

The name 'h-effectivity functions' for the functions defined in item 3. above is short for 'h-relative effectivity functions'. This name is inspired by similar terminology in Coalition Logic whose semantics is in terms of 'effectivity functions'. An effectivity function in Coalition Logic is a function $E : S \times 2^{Ags} \mapsto 2^{2^S}$ mapping static states to sets of sets of static states. Each set in 2^{2^S} then represents a choice. In our h-effectivity functions, choices are always relative to a history (the history that is part of the dynamic state we evaluate against), which is why h-effectivity functions map to sets instead of to sets of sets.

Condition 2.a above ensures that the structure of histories is isomorphic to that of a tree.

Condition 3.a ensures that event types are only assigned to state pairs where one state successes the other.

Condition 3.b ensues that if histories are still undivided, transitions between their subsequent states are uniform, that is, they are characterized by the same set of event type labels.

Condition 4.b says that the next state on the current history is always in the current effectivity set of any group of agents. This gives a notion of success (in instantaneous *stit* semantics [4] the success property is modeled by the truth axiom).

Condition 4.c above states that any next state is in the effectivity set of the empty set and vice versa. This implies the empty set of agents is powerless: it cannot choose between different options and has to 'go with the flow'.

Condition 4.d above implies that a simultaneous choice exertion of all agents in the system uniquely determines a next static state. A similar condition holds for related formalisms like ATL [2] and Coalition logic (CL for short). However, we want to point here to an important difference with these formalisms. Although 4.d uniquely determines the next state relative to a simultaneous choice for all agents in the system, it does not determine the unique next 'dynamic state'. This is important, because dynamic states are the units of evaluation. In ATL and CL, static states are the units of evaluation. As a consequence, CL is not definable in this logic.

Condition 4.e expresses coalition monotony, saying that whatever is ensured by the choice of a group of agents is also ensured by the simultaneous choice of any supergroup of agents.

Condition 4.f above states that simultaneous choices of different agents never have an empty intersection. In *stit* this is referred to as the condition of 'independence of agency'. It says that a choice exertion of one agent can never have as a consequence that some other agent is limited in the choices it can exercise simultaneously.

I briefly explain the formal definition of the frames in definition 3.2 using Fig. 4. The small squares are static states in the effectivity sets of $E(s, h, Ags)$. Combinations of static states and histories running thought them form dynamic states. The big, outmost squares forming the boundaries of the game forms, collect the static (and implicitly also the dynamic) states in the effectivity sets of $E(s, h, \emptyset)$. Independence of choices is reflected by the fact that the game forms contain no 'holes' in them. The semantics is a so called 'bundled' semantics. In a bundled semantics choice exertion is always thought of as the separation of two bundles of histories: one bundle ensured by the choice exercised and one bundle excluded by that choice. In the figure the bundles are depicted as bundles.

We now define models by adding a valuation of propositional atoms to the frames of definition 3.2.

Definition 3.3 A frame $\mathcal{F} = \langle S, H, A, E\rangle$ is extended to a model $\mathcal{M} = \langle S, H, E, \pi\rangle$ by adding a valuation π of atomic propositions:

- π is a valuation function $\pi : P \longrightarrow 2^{S\times H}$ assigning to each atomic proposition the set of dynamic states relative to which they are true.

The truth conditions for the semantics of the operators are fairly standard. The non-standard aspect is the two-dimensionality of the semantics, meaning that we evaluate truth with respect to dynamic states built from a dimension of histories and a dimension of static states.

Definition 3.4 Relative to a model $\mathcal{M} = \langle S, H, A, E, \pi\rangle$, truth $\mathcal{M}, \langle s, h\rangle \models \varphi$ of a formula φ in a dynamic state $\langle s, h\rangle$, with $s \in h$, is defined as:

$$
\begin{aligned}
\mathcal{M}, \langle s, h\rangle \models p &\Leftrightarrow s \in \pi(p)\\
\mathcal{M}, \langle s, h\rangle \models \neg\varphi &\Leftrightarrow \text{not } \mathcal{M}, \langle s, h\rangle \models \varphi\\
\mathcal{M}, \langle s, h\rangle \models \varphi \wedge \psi &\Leftrightarrow \mathcal{M}, \langle s, h\rangle \models \varphi \text{ and}\\
&\qquad \mathcal{M}, \langle s, h\rangle \models \psi\\
\mathcal{M}, \langle s, h\rangle \models \Box\varphi &\Leftrightarrow \forall h' : \text{ if } s \in h' \text{ then}\\
&\qquad \mathcal{M}, \langle s, h'\rangle \models \varphi\\
\mathcal{M}, \langle s, h\rangle \models X\varphi &\Leftrightarrow \forall s' : \text{ if } s' = succ(s, h) \text{ then}\\
&\qquad \mathcal{M}, \langle s', h\rangle \models \varphi\\
\mathcal{M}, \langle s, h\rangle \models [A\ xstit]\varphi &\Leftrightarrow \forall s', h' : \text{ if } s' \in E(s, h, A) \text{ and}\\
&\qquad s' \in h' \text{ then } \mathcal{M}, \langle s', h'\rangle \models \varphi\\
\mathcal{M}, \langle s, h\rangle \models [A\ perf\ a] &\Leftrightarrow \forall s', h' : \text{ if } s' \in E(s, h, A) \text{ and}\\
&\qquad s' \in h' \text{ then } a \in A(s, s')
\end{aligned}
$$

Satisfiability, validity on a frame and general validity are defined as usual.

Note that the historical necessity operator quantifies over one dimension, and the next operator over the other. The *stit* modality combines both dimensions.

Definition 3.5 The following axiom schemas, in combination with a standard axiomatization for propositional logic, and the standard rules (like necessitation) for the normal modal operators, define a Hilbert system for XSTIT.ET:

	S5 for $\Box$
	KD for each $[A\ xstit]$
(Det)	$\neg X\neg\varphi \rightarrow X\varphi$
($\emptyset$ = Sett)	$[\emptyset\ xstit]\varphi \leftrightarrow \Box X\varphi$
(Ags = XSett)	$[Ags\ xstit]\varphi \leftrightarrow X\Box\varphi$
(CMon)	$[A\ xstit]\varphi \rightarrow [B\ xstit]\varphi$ for $A \subseteq B$
(Indep-G)	$\Diamond[A\ xstit]\varphi \wedge \Diamond[B\ xstit]\psi \rightarrow \Diamond([A\ xstit]\varphi \wedge [B\ xstit]\psi)$ *for* $A \cap B = \emptyset$
(a-CMon)	$[A\ perf\ a] \rightarrow [B\ perf\ a]$ *for* $A \subseteq B$
(a-Indep-G)	$\Diamond[A\ perf\ a] \wedge \Diamond[B\ perf\ b] \rightarrow \Diamond([A\ perf\ a] \wedge [B\ perf\ b])$ *for* $A \cap B = \emptyset$
(Aa-Lnk)	$[A\ perf\ a] \wedge \Box([Ags\ perf\ a] \rightarrow X\varphi) \rightarrow [A\ xstit]\varphi$

Conjecture 3.1 *The Hilbert system of definition 3.5 is complete with respect to the semantics of definition 3.4.*

The logic of the new operator $[A$ perf $a]$ is very simple. It is not a traditional modal operator, since it works on event type terms and not on arbitrary formulas. Since the event type terms are atomic here, it is close to obvious that the above system is complete. Of course we can get more interesting logics by generalizing to boolean event types or to a regular language like in full propositional dynamic logic. But this

is left for future work. Here the central aim is to put forward the central idea about the relation between agency and event types.

Now it is time to explain how in the logic XSTIT.ET we can simulate the basic dynamic logic operator $[a]\varphi$. This is accomplished by the following definition.

Definition 3.6 $[a]\varphi \equiv_{def} \Box([Ags\ perf\ a] \rightarrow X\varphi)$

Proposition 3.1 *Any event type operator* $[a]\varphi$ *as given by definition 3.6 is a normal modal K operator (like in Hennessey-Milner logic)*

The proposition claims that the simulation is indeed a correct simulation of a dynamic logic like operator, and is easily verified by inspection of the semantics. Now I briefly mention three simple properties that follow in the logic.

(a) $\langle a\rangle\varphi \leftrightarrow \Diamond([Ags\ \mathsf{perf}\ a] \wedge X\varphi)$
(b) $[A\ \mathsf{perf}\ a] \wedge [a]\varphi \rightarrow [A\ \mathsf{xstit}]\varphi$
(c) $[a]\varphi \rightarrow \Box[a]\varphi$

Property (a) follows as the dual of definition 3.6. One thing it says is that an event of some type can only occur if the complete group of agents can perform it. Property (b) says that if a group performs an act of a certain type, and if acts of that type, when they occur guarantee that φ holds, then the group sees to it that φ. This property embodies the central relationship between agency and event type reasoning in this theory. Finally, property (c) emphasizes the Peircian character of the dynamic logic operator.

The axiom (a-Indep-G) expresses independence of event types in the sense that if one agent can perform an event of type a and another agent can perform an event of type b, it is always possible for them to perform these events jointly. It might seem then that here the theory goes wrong. For instance, if a is the type 'the closing of a door' and b is the type 'the opening of a door', then we cannot have that events of these types can occur at the same time, which means that the axiom does not apply. However, when we say that an agent has the ability to perform an event of the type 'the opening of a door', we never mean that this agent has the ability *under all possible circumstances*. Indeed if another agent obstructs, or if moisture has caused the door to expand the agent cannot open the door even though we would still say that the agent has the ability to open a door. So, an ability is always a conditional: the capacity to perform an action 'under normal circumstances'. Often we are not even aware of what these circumstance are (which relates directly to what in formal theories of action is called the 'qualification problem' [13]). But we know that in most cases we will be able to perform the event associated with the ability.[3] So examples as the one given here are not a counter example to the axiom.

At the end of the introduction I said that a good theory of agency and event types should explain how one agent performing an event of type a differs from another

[3] If we model knowledge using probabilities, as in [6], we might also say that an ability is the capacity to significantly higher the chance that an event occurs.

agent performing an event of type a and from both of them performing the event of type a at the same time. Here I will show how the logic makes a difference between these situations. In the logic these three positions can be represented by $[A\ \mathsf{perf}\ a]$, $[B\ \mathsf{perf}\ a]$ and $[A \cup B\ \mathsf{perf}\ a]$ with $A \cap B = \emptyset$ (for the sake of generality we generalize to groups). Because of axiom $(a - CMon)$ we have that the third condition follows from each of the first two conditions. Furthermore we have that this is the only logical dependency there is between the formulas. So, $[A\ \mathsf{perf}\ a]$ can be satisfied while $[B\ \mathsf{perf}\ a]$ is not, and $[A \cup B\ \mathsf{perf}\ a]$ can be satisfied, while neither of $[A\ \mathsf{perf}\ a]$ or $[B\ \mathsf{perf}\ a]$ is.

3.2 Collective Responsibility for an Action

Figure 5 pictures two situations of collective responsibility for an event of type a. In the left picture we have that "$[ag_1\ \mathsf{perf}\ a] \wedge [ag_2\ \mathsf{perf}\ a]$" (the grey row is the choice exerted by agent 2 and the grey column is the choice exerted by agent 1). Here both agent ag_1 and agent ag_2 perform an event of type a, and if one of them had chosen differently, the event of type a would still have occurred due to the agentive effort of the other agent. In the right picture we have that "$[ag_1 \cup ag_2\ \mathsf{perf}\ a]$" and the combined agentive effort of both agents is required for the event of type a to occur.

It is a very interesting question to ask in what sense the collective responsibilities differ in these two situations. Assume that the event of type a is one that is wrong (relative to some normative system) and that we have to decide in which situation the agents are more to blame. Interestingly enough we can argue in two opposite directions. We might say that the individual agents in the left 'full cross' case are more to blame, because each on their own their effort would have been enough to ensure that the bad event occurs. This can be interpreted as pointing to a strong determination on the side of both agents involved. On the other hand we might argue

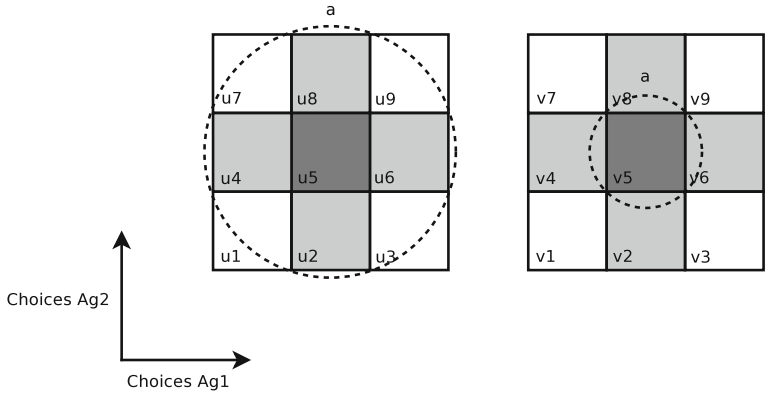

Fig. 5 "$[ag_1\ \mathsf{perf}\ a] \wedge [ag_2\ \mathsf{perf}\ a]$" versus "$[ag_1 \cup ag_2\ \mathsf{perf}\ a]$"

that in the full cross case, both agent's actions are not 'sine qua non', meaning that their effort was not strictly necessary for the bad event to occur. Here the argument would be that each of the agents can claim that the bad event would occur anyway, because the other agent already ensured it. We can make similar arguments starting from the right picture; the one where a is associated with the center of the cross. On the one hand we can argue that relative to the full cross scenario, the agents each individually are more to blame, because each of them had the power to prevent a from happening. On the other hand, we can say that each of them is less to blame in comparison to the full cross case, because their action alone was not enough to ensure a; they both needed the other, and in that sense they could not have been 100 % sure about the outcome.

I believe to analyze this twin example further, and understand how collective responsibility and individual responsibility relate to each other, we need to bring the epistemic dimension into the picture. I will leave this to future work.

4 Related Work

Fairly recently several authors addressed the problem of combining logics of agency and dynamic logic. Here I will mention these works only briefly. Recently Marek Sergot proposed the logic of unwitting collective agency [28]. The adjective 'unwitting' refers to the absence of any epistemic or motivational aspects, which, as is explained in [4], is also a starting point of *stit* theory. However, Sergot is very critical of the *stit* notion of 'independence of agency'. Sergot's semantics takes transitions between system states as the central semantic entities the formulas of his language are evaluated against. There are some similarities with the work of Segerberg discussed in this paper: it departs from dynamic logic intuitions and switches to an Ockhamist view (as Prior calls it) on the evaluation of truth, which enables him to simulate *stit*-like operators. Another work in the same spirit is that of Herzig and Lorini [15]. In this work the central operator is of the form $\langle Ag{:}a\rangle\varphi$ and the reading is 'agent Ag does an action of type a resulting in a state where φ'. However, any agent can only do one action of one particular type a at the time, which is a conceptual limitation that inhibits on the explanatory power of the theory (the proposals of Sergot, Segerberg and myself do not have this limitation). Also in this approach, dynamic logic intuitions are the point of departure and *stit* operators are simulated. A third work is that by Ming Xu [29]. Xu studies a slightly different problem though. He does not talk about event types, but aims to reconcile logics of agency with a language that talks directly about events. Xu takes operators $[Ag, e]\varphi$ as the central objects of study, where Ag is an agent and e is an event or action, and not an event *type*. Finally, there are several papers discussing the problem addressed here, without committing to a possible solution, like the papers of Brian Chellas [11], Risto Hilpinen [16] and Mark Brown [10].

5 Conclusion

I have discussed Segerberg's approach to combining two views on action—the dynamic logic view and the *stit* view—within one framework, as put forward in *Outline of a logic of action* [26]. I have placed some critical remarks on the theory. These remarks do not so much concern the fact that the framework lacks certain concepts or makes some oversimplifications (which is, as for any theory, also true, as Segerberg discusses in the final words of the paper), but directly question the idea that a description in terms of dynamic logic event types is appropriate for understanding agency. In stead I suggest to turn this view 180°; I have put forward an alternative action theory outline where it is the dynamic logic event type reasoning that is simulated in a *stit* framework. Further explorations and comparisons in future research will have to shed light on which of the two approaches best explains the relation between computation and agency.

References

1. Alur, R., Henzinger, T.A. & Kupferman. O. (1997). Alternating-time temporal logic. In *FOCS '97: Proceedings of the 38th Annual Symposium on Foundations of Computer Science (FOCS '97)* (pp. 100–109). IEEE Computer Society.
2. Alur, R., Henzinger, T. A., & Kupferman, O. (2002). Alternating-time temporal logic. *Journal of the ACM*, *49*(5), 672–713.
3. Belnap, N., & Perloff, M. (1988). Seeing to it that: a canonical form for agentives. *Theoria*, *54*(3), 175–199.
4. Belnap, N., Perloff, M. & Xu, M. (2001). *Facing the future: agents and choices in our indeterminist world*. Oxford University Press
5. Broersen, J. (2011). Making a start with the stit logic analysis of intentional action. *Journal of Philosophical Logic*, *40*, 399–420.
6. Broersen, J. (2011). Modeling attempt and action failure in probabilistic stit logic. In T. Walsh, (Ed.), *Proceedings of Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI 2011)* (pp. 792–797). IJCAI.
7. Broersen, J.M. (2009). A complete stit logic for knowledge and action, and some of its applications. In M. Baldoni, T. Cao Son, M.B. van Riemsdijk & M. Winikoff, (Eds.), *Declarative Agent Languages and Technologies VI (DALT 2008), volume 5397 of, Lecture Notes in Computer Science* (pp. 47–59). Springer
8. Broersen, J. M. (2011). Deontic epistemic stit logic distinguishing modes of mens rea. *Journal of Applied Logic*, *9*(2), 127–152.
9. Brown, M. A. (1988). On the logic of ability. *Journal of philosophical logic*, *17*(1), 1–26.
10. Brown, M. A. (2008). Acting, events and actions. In R. van der Meyden & L. van der Torre (Eds.), *Deontic Logic in Computer Science, 9th International Conference, DEON 2008, Luxembourg, Luxembourg, July 15–18, 2008. Proceedings, volume 5076 of Lecture Notes in Computer Science* (pp. 19–33). Springer.
11. Chellas, B. F. (1995). On bringing it about. *Journal of Philosophical Logic*, 24, 563–571 (1995). doi:10.1007/BF01306966.
12. Emerson, E.A. (1990). Temporal and modal logic. In J. van Leeuwen (Ed), Handbook of theoretical computer science, volume B: Formal models and semantics, chapter 14, pp. 996–1072. Elsevier Science.

13. Ginsberg, M. L., & Smith, D. E. (1988). Reasoning about action II: The qualification problem. *Artificial Intelligence*, *35*, 311–342.
14. Harel, D. & Peleg, D. (1985). Process logic with regular formulas. *Theoretical Computer Science*, *38*, 307–322.
15. Herzig, Andreas, & Lorini, Emiliano. (2010). A dynamic logic of agency I: Stit, capabilities and powers. *Journal of Logic, Language and Information*, *19*(1), 89–121.
16. Hilpinen, R. (1997). On action and agency. In E. Ejerhed & S. Lindström (Eds.), *Logic, Action and Cognition: Essays in Philosophical Logic* (pp. 3–27). Kluwer Academic Publishers.
17. Kanger, S. (1972). *Law and logic. Theoria*, *38*(3), 105–132.
18. McCarthy, J. & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine Intelligence*, 4, pp. 463–502. Edinburgh University Press.
19. Milner, R. (1989). *Communication and concurrency*. Prentice-Hall.
20. Pauly, Marc. (2002). A modal logic for coalitional power in games. *Journal of Logic and Computation*, *12*(1), 149–166.
21. Pratt, V. R. (1976). Semantical considerations on Floyd-Hoare logic. In *Proceedings 17th IEEE Symposium on the Foundations of Computer Science* (pp. 109–121). IEEE Computer Society Press.
22. Prior, A. N. (1967). *Past, present, and future*. Clarendon Press.
23. Segerberg, K. (1989). Bringing it about. *Journal of Philosophical Logic*, *18*(4), 327–347.
24. Segerberg, K. (1992). Getting started: Beginnings in the logic of action. *Studia Logica*, *51*, 347–378.
25. Segerberg, K. (2000). Outline of a logic of action. Technical Report 5–2000, Department of Philosophy University of Uppsala.
26. Segerberg, K. (2002). Outline of a logic of action. In *Advances in Modal Logic, 3*, pp. 365–387. World Scientific.
27. Segerberg, K. (1996). The delta operator at three levels of analysis. In *Logic, action, and, information* (pp. 63–78). De Gruyter
28. Sergot, M. (2008). The logic of unwitting collective agency. Technical Report 2008/6, Department of Computing, Imperial College London.
29. Ming, Xu. (2010). Combinations of stit and actions. *Journal of Logic, Language and Information*, *19*(4), 485–503.