Introducing *attempt* in a modal logic of intentional action[*]

Emiliano Lorini[1], Andreas Herzig[2], and Cristiano Castelfranchi[1]

[1] Institute of Cognitive Sciences and Technologies-CNR, Rome, Italy
[2] Institut de Recherche en Informatique de Toulouse (IRIT), France

Abstract. The main objective of this work is to develop a multi-modal logic of Intention and Attempt. We call this logic *LIA*. All formal results are focused on the notion of *attempt*. We substitute the dynamic molecular notion *action* by his atomic constituent *attempt* and define the former from the latter. The relations between attempts, goals, beliefs and present-directed intentions are studied. A section of the paper is devoted to the analysis of the relations of our modal logic with a situation calculus-style approach.

1 Introduction

BDI (belief, desire, intention) logics [21, 18, 4, 12] are conceived as explicit formal models of the intentional pursuit. If this is true then they should be able to take into account notions such as the notion of *attempt* and *trying*. These two notions have been mainly discussed in the philosophical field and taken into account in some logics of agency[3] but few models exist that are able to integrate in the same formal framework a precise description of practical reasoning (motivational dynamics and functional properties of mental states) with a description of its external and physical counterpart: the executive phase of intentional action. The main objective of this work is to develop a multi-modal logic which enables us to deal with the notion of *attempt* inside the more general framework of Intentional Action theory. We call this logic *LIA*: Logic of Intention and Attempt. The main difference between *LIA* and standard dynamic logic is that the dynamic primitives are not atomic actions, but atomic attempts. In our view a model of intentional action should explicitly represent the process of action execution, from the agent "triggering" the action to the successful execution of the action (when the preconditions for action execution hold). The axioms of the logic will be presented and discussed in Section 3. In section 4 the notions of *attempt* and *action* will be compared. It will be shown how action theories can be specified starting from the primitive notion of *attempt*. In section 5 additional properties of *attempt* will be discussed and the formal definition of present-directed intention will be introduced.

2 Properties of attempts and basic actions

With "agent *i attempts to do an action* α", we mean that "agent *i triggers the execution of action* α", "agent *i exerts himself to do action* α". In our view the *attempt* is the core

---

[*] We thank the anonymous referees of this paper for their helpful comments.
[3] See for example [23] where attempts are defined as "not necessarily successful actions".

element of the causal process which leads from the present-directed intention [1] to the successful execution of the action in the external world. Several authors have emphasized the importance of this concept for a theory of Intentional action (see [3], [13], [19]). The Psycho-Psycho Law proposed by O'Shaughnessy [19] stresses the "bridging role" of the *attempt* between the present-directed intention and the execution of the action: "if an agent at an instant in time realizes that that instant is an instant at which he intends to perform action x, then logically necessarily he begins trying to do x at that very moment of realization". In our view the property relating *attempts* with *actions* is the following one: *an action is effectively performed if and only if the performing agent triggers the action under the appropriate preconditions for action execution. The fact that the preconditions for action execution hold, guarantees that the attempt will be successful: it will succeed in producing the associated action (i.e. in causing the intrinsic result of the associated action[4]).* In the present analysis we deal with *basic actions* of a given agent *i* and leave aside the issue of *complex actions* (therefore every time we use the term *action* we mean *basic action*). According to [10, 5] an agent can perform a basic action α without necessarily performing some other action β and without necessarily believing that β must be performed in order to perform α. On the other hand a complex action (for example *Jack killing Joe*) depends on some other action (*Jack shooting Joe*) which in turn depends on some other action (*Jack pulling the trigger*) and so on... Thus we assume that basic actions are precisely those actions which are always executed when the agent attempts to perform them and the preconditions for action execution hold. Complex actions do not have this property. There could be an agent *i*'s complex action α whose execution also depends on some external event or some agent *j*'s action β which must happen after the initiation of action α (for example Jack can not perform the complex action of *killing Joe* if Joe does not perform the action of *drinking the poisoned soda* after that Jack *has poured some poison into the glass of soda*). Moreover we focus in this paper on "intentional attempts" assuming that an attempt to do α is always produced by the goal to attempt to do action α (see also [28] with respect to this hypothesis). Finally let us observe that there are two ways to conceive the notion of *attempt* (or *trying*). Some authors [16, 19] conceive the attempt as a purely mental event. If we adopt this perspective we should postulate that *unsuccessful attempts*[5] (differently from actions) never change the physical (external) world and are not perceivable by other agents. Other authors [13] are more prone to accept that *attempt* already refers to the physical realization of the basic action. In this analysis we adhere to the latter view and assume that the category *unsuccessful attempt* includes all those cases of "partial" execution of a basic action not producing the intrinsic result of the action (for example an agent who attempts to *raise the hand above the head* and only moves the arm of few millimeters since the arm is blocked). In [10, 5, 13] it is assumed that basic actions include only bodily movements such as *raising the arm*, *moving the leg*, *turning the sensor* etc... Thus in the examples given for supporting our analysis we

---

[4] According to [26, 29] the intrinsic result of an action is "the result which logically must occur if the action is to have been done". For instance the agent cannot have *opened his eye* unless *his eye is open.*

[5] With *unsuccessful attempt* we mean that the action that the attempt should produce is not performed due to the fact that the preconditions for executing the action do not hold.

will often refer to basic actions by using names denoting human bodily movements. Our analysis can be extended to realistic applications where the agent would be a robot with an artificial body (artificial limbs, rotating wheels, moving sensors etc...).

3 Formal Logic: syntax, semantics, axiom system

LIA is a multi-modal logic of time, attempts, actions, goals and beliefs.[6] The logic is based on a combination of an enhanced version of linear temporal logic where it is possible to talk about actions and Cohen and Levesque's logic of goal and intention [4]. The main difference with respect to standard dynamic logic [11] is that the notion of *atomic (basic) action* is substituted with the more primitive notion *basic attempt*. We will show that the former can be defined from the latter.

The syntactic primitives of the logic are the following: -a set of atomic (basic) actions $ACT = \{\alpha, \beta, ...\}$; -a set of agents $AGT = \{i, j, ...\}$; -a set of propositional atoms $\Pi = \{p, q, ...\}$. The set of propositional formulas of our language is denoted by $PROP$ (elements in $PROP$ are denoted by $\Phi, \Omega, \Psi, ...$). The set FOR of well formed formulas φ of our modal action language L is defined by the following BNF:

$\varphi := p|\top|\neg\varphi|\varphi \wedge \psi| [[i, \alpha]] \, \varphi|G\varphi|X\varphi|\varphi Until\psi|Bel_i\varphi|Goal_i\varphi$

where p ranges over Π, i ranges over *AGT* and α ranges over *ACT*.

$[[i, \alpha]] \, \varphi$ is read "φ holds after any agent *i*'s *attempt* to do α". Hence $[[i, \alpha]] \perp$ expresses "agent *i* does not attempt to do α". Three abbreviations are used. $\langle\langle i, \alpha\rangle\rangle \, \varphi$ abbreviates $\neg [[i, \alpha]] \neg\varphi$, $F\varphi$ abbreviates $\neg G\neg\varphi$ and $\varphi Before\psi$ abbreviates $\neg(\neg\varphi Until\psi)$. Hence $\langle\langle i, \alpha\rangle\rangle \, \varphi$ has to be read "agent *i* attempts to do α and φ holds after this attempt " and $\langle\langle i, \alpha\rangle\rangle \, \top$ has to be read "agent *i* attempts to do α". For example $\langle\langle Bill, raiseArm\rangle\rangle \, \top$ is read "Bill attempts to raise the arm". We briefly go into the basic semantics.

A model for *LIA* is defined by the tuple M = $(W, R_X, R^{att}, B, G, V)$.

- W is a set of worlds.

- R_X is a mapping $R_X : W \longrightarrow 2^W$ associating sets of possible worlds $R_X(w)$ to each possible world *w*. We suppose that R_X is a total function.

- R^{att} is a mapping $R^{att} : AGT \times ACT \longrightarrow (W \longrightarrow 2^W)$ associating sets of possible worlds $R^{att}_{i:\alpha}(w)$ to each possible world *w*. We assume that every $R^{att}_{i:\alpha}$ is a partial function.

- *B* is a mapping $B : AGT \longrightarrow (W \longrightarrow 2^W)$ associating sets of possible worlds $B_i(w)$ to each possible world *w*. We suppose that every B_i is serial, transitive and euclidean.[7]

- *G* is a mapping $G : AGT \longrightarrow (W \longrightarrow 2^W)$ associating sets of possible worlds $G_i(w)$ to each possible world *w*. We suppose that also every G_i is serial, transitive and euclidean.

- *V* is a mapping $V : \Pi \longrightarrow 2^W$ associating sets of possible worlds to propositional atoms.

---

[6] The logic is described more extensively in [15] where also formal proofs of the theorems are presented.

[7] We use a modal logic KD45 as the logic for Belief and Goals, i.e. an agent does not entertain inconsistent Beliefs (and inconsistent Goals) and is aware of his Beliefs and disbeliefs (and of his Goals and non-Goals).

After defining R_X^* as the reflexive and transitive closure of R_X, we look at truth conditions.

- $M, w \models p$ iff $w \in V(p)$.
- $M, w \models \neg\varphi$ iff not $M, w \models \varphi$
- $M, w \models \varphi \wedge \psi$ iff $M, w \models \varphi$ and $M, w \models \psi$
- $M, w \models X\varphi$ iff $\forall w'$ such that $w' \in R_X(w)$ it holds that $M, w' \models \varphi$
- $M, w \models G\varphi$ iff $\forall w'$ such that $w' \in R_X^*(w)$ it holds that $M, w' \models \varphi$.
- $M, w \models \varphi Until \psi$ iff $\exists\, w' \in R_X^*(w)$ such that $M, w' \models \psi$ and $\forall w''$ if $w'' \in R_X^*(w)$ and $w' \in R_X^*(w'')$ and $w'' \notin R_X^*(w')$ then $M, w'' \models \varphi$
- $M, w \models [[i, \alpha]]\, \varphi$ iff $\forall w'$ such that $w' \in R_{i:\alpha}^{att}(w)$ it holds that $M, v \models \varphi$
- $M, w \models Bel_i\varphi$ iff $\forall w' \in B_i(w)$ it holds that $M, w' \models \varphi$.
- $M, w \models Goal_i\varphi$ iff $\forall w' \in G_i(w)$ it holds that $M, w' \models \varphi$.

We use a complete axiomatization of linear temporal logic (axioms 0a-7a plus inference rules R1-R3) [8, 9] and the axioms and inference rules of the basic normal modal logic for *belief* modal operator, *goal* modal operator and *attempt* modal operator plus axioms 1b-12b (table 1).[8]

0a. All tautologies of propositional calculus	1b. $\neg(Bel_i\varphi \wedge Bel_i\neg\varphi)$
1a. $G(\varphi \to \psi) \to (G\varphi \to G\psi)$	2b. $Bel_i\varphi \to Bel_iBel_i\varphi$
2a. $X\neg\varphi \leftrightarrow \neg X\varphi$	3b. $\neg Bel_i\varphi \to Bel_i\neg Bel_i\varphi$
3a. $X(\varphi \to \psi) \to (X\varphi \to X\psi)$	4b. $\neg(Goal_i\varphi \wedge Goal_i\neg\varphi)$
4a. $G\varphi \to \varphi \wedge XG\varphi$	5b. $Goal_i\varphi \to Bel_iGoal_i\varphi$
5a. $G(\varphi \to X\varphi) \to (\varphi \to G\varphi)$	6b. $\neg Goal_i\varphi \to Bel_i\neg Goal_i\varphi$
6a. $\varphi Until\psi \to F\psi$	7b. $Bel_i\varphi \to Goal_i\varphi$
7a. $\varphi Until\psi \leftrightarrow \psi \vee (\varphi \wedge X(\varphi Until\psi))$	8b. $Bel_i\,[[j, \alpha]]\,\psi \wedge \neg Bel_i\,[[j, \alpha]]\,\bot \to [[j, \alpha]]\,Bel_i\psi$
Inference Rules:	9b. $[[j, \alpha]]\,Bel_i\psi \wedge \neg\,[[j, \alpha]]\,\bot \to Bel_i\,[[j, \alpha]]\,\psi$
R1. $\dfrac{\vdash\varphi \;\; \vdash\varphi\to\psi}{\vdash\psi}$ (*modus ponens*)	10b. $Bel_i(GBel_i\psi \leftrightarrow Bel_iG\psi)$
R2. $\dfrac{\vdash\varphi}{\vdash G\varphi}$ (*G-necessitation*)	11b. $Goal_i\,\langle\langle i, \alpha\rangle\rangle\,\top \leftrightarrow \langle\langle i, \alpha\rangle\rangle\,\top$
R3. $\dfrac{\vdash\varphi}{\vdash X\varphi}$ (*X-necessitation*)	12b. $X\varphi \to [[i, \alpha]]\,\varphi$

Table 1. Axiom system.

Semantic characterizations (model correspondence) of the previous axioms and inference rules are given in [15]. In [15] it is also proved that *LIA* is *sound* with respect to the set of *LIA models* satisfying all the semantic constraints imposed by the previous axioms and inference rules. With $\models_{LIA} \varphi$ we mean that φ is valid in all *LIA models* and with $\vdash_{LIA} \varphi$ we mean that φ is a theorem of *LIA*. Moreover we say that a formula φ is a consequence of the set of global assumptions $\{\Phi_1, ..., \Phi_n\}$ in the class of models LIA (noted $\{\Phi_1, ..., \Phi_n\} \models_{LIA} \varphi$) if and only if for all models $M \in LIA$ if $\models_M \Phi_i$ for every Φ_i, then $\models_M \varphi$.

---

[8] Notice that from our axiomatic system (axiom 2a and axiom 12b) it follows that: $\langle\langle i, \alpha\rangle\rangle\,\varphi \to [[i, \alpha]]\,\varphi$, i.e. action are deterministic.

Axiom 11b deserves some comments. This axiom has not been analyzed before in the literature on modal logic of intentional action. It establishes that an agent attempts to do some action α if and only if the agent has the goal to attempt to do action α. The axiom relates mental attitudes with the executive and behavioral element. [9] In our view it is relevant for a formal theory of action to account for the role of intention in producing a given performance. This fundamental issue is not enough stressed in formal models of intentional action. Axiom 11b has exactly this role. It accounts for the conditions for passing from a pure mental and motivational level to the executive and physical reality. We will show later (in section 5) that axiom 11b is central for the notion of *present-directed intention*. In this analysis we do not introduce at the formal level *sequential composition* of basic actions. We prefer to work with the simplest formal language for actions, leaving the problem of sequential composition to future work. Let us only stress that axiom 11b should also be applied to sequences of basic actions $\alpha;\beta;...$ performed by the same agent and which do not involve perception (epistemic actions). For example consider a football playing robot having the goal to perform the sequence of basic actions *turn-right;advance;shoot*. The goal to attempt to perform the sequence of actions triggers the performance and even if the robot is blocked by another player, he will attempt to execute the three basic actions in sequence, without realizing that the first action fails.[10]

4 Attempts and action theories

4.1 Definition of Action and execution preconditions[11]

Our definition of action is built on the special formula $Pre(\alpha)$ denoting the *physical preconditions for executing action α (execution preconditions)*. We assume that $Pre(\alpha)$ is a function returning some classical formula Φ[12] that is $Pre : ACT \rightarrow PROP$. For example we might have: $FreeLeg = Pre(kickBall)$.

---

[9] In [20] a similar axiom is proposed where actions are related with knowledge (the axiom says that if an agent *i* can do a certain action α from his repertoire then the agent knows that he can do it).

[10] We are supposing here some kind of *persistence* in the intentional execution of those sequences of actions which do not involve perception that is, when an agent has the goal to attempt to perform a sequence of actions which does not involve perception then the agent attempts to perform the complete (intended) sequence of actions (the agent cannot stop in the middle of the sequence and revise his pushing intentions) and when the agent attempts to perform a complete sequence of actions which does not involve perception then the agent has the goal to attempt to perform the complete sequence.

[11] Given that time is linear in our logic, we use the terms "execution precondition" and "execution law" instead of the terms "executability precondition" and "executability law".

[12] In realistic applications the function *Pre* should have an agent argument: the preconditions of kicking a ball may differ from agent to agent, for example for a lame agent $Pre(kickBall) = \perp$. Here we make the assumption that *execution preconditions* of an action are the same for all agents.

DEFINITION 1: Action. $\langle i, \alpha \rangle \, \varphi =_{def} \langle \langle i, \alpha \rangle \rangle \, \varphi \wedge Pre(\alpha)$

Definition 1 relates the notion of *action* with the primitive notion of *attempt*. It says that action executions are attempts whose execution preconditions hold. An instance of definition 1 is: $\langle i, \alpha \rangle \top =_{def} \langle \langle i, \alpha \rangle \rangle \top \wedge Pre(\alpha)$. It says that a given action α is executed by agent *i* if and only if agent *i* attempts to do α and the preconditions for executing action α are holding. This is an explicit way to relate actions with attempts and to express execution laws. On the basis of definition 1 execution laws are defined by referring to the *attempt* notion. This kind of solution is quite different from standard solutions (see for example [2] and [22]) where execution laws are expressed by taking actions as primitive elements, without deconstructing them into more elementary constituents (viz. attempts). Moreover, from definition 1 it follows that a consequence of an attempt to perform α is also a consequence of the successful performance of basic action α and if the *execution preconditions* hold then the consequences of the attempt are equivalent to the consequences of the associated basic action. Indeed: a) $[[i, \alpha]] \, \varphi \rightarrow [i, \alpha] \, \varphi$ and b) $Pre(\alpha) \rightarrow ([[i, \alpha]] \, \varphi \leftrightarrow [i, \alpha] \, \varphi)$ are two valid formulas in our logic. In the next sections we analyze effect laws by substituting the notion of action with the more primitive notion of attempt and show that by distinguishing attempts from actions we get some important conceptual refinements.

4.2 Effect preconditions

Similarly to Situation Calculus [22] we take for each propositional atom $p \in \Pi$ and basic action $\alpha \in ACT$ a propositional formula $\gamma^+(\alpha, p)$ describing the *positive effect preconditions* of the attempt to do α with respect to *p* and $\gamma^-(\alpha, p)$ describing the *negative effect preconditions* of the attempt to do α with respect to *p*. For example the following *positive effect preconditions* can be associated to the attempt to perform the actions *loading*, *pulling* and *picking up*.[13]

$\gamma^+(load, loadedGun) = freeHand \wedge holdsGun$
$\gamma^+(pull, wounded) = holdsGun \wedge loadedGun \wedge pointedGun \wedge freeHand$
$\gamma^+(pull, pulledTrigger) = holdsGun \wedge freeHand$
$\gamma^+(pull, scared) = holdsGun \wedge pointedGun$
$\gamma^+(pickUp, holdsGun) = gunOnTable \wedge freeArm \wedge freeHand$

Effect laws are specified accordingly in terms of global assumptions in Fitting's sense [7] of the form: $\gamma^+(\alpha, p) \rightarrow [[\alpha]] \, p$ and $\gamma^-(\alpha, p) \rightarrow [[\alpha]] \, \neg p$.

For instance positive effect axioms for the previous actions are specified by the following global assumptions:[14]

$holdsGun \wedge pointedGun \rightarrow [[pull]] \, scared$
$holdsGun \wedge loadedGun \wedge pointedGun \wedge freeHand \rightarrow [[pull]] \, wounded$
$freeHand \wedge holdsGun \rightarrow [[load]] \, loadedGun$
$holdsGun \wedge freeHand \rightarrow [[pull]] \, pulledTrigger$
$gunOnTable \wedge freeArm \wedge freeHand \rightarrow [[pickUp]] \, holdsGun$

---

[13] To simplify our exposition we suppose in our examples that for each action α and possible effect *p* we have $\gamma^-(\alpha, p) = \bot$. Thus we do not need to specify *negative effect preconditions*.

[14] Notice that in effect laws actions are not indexed by agents. Indeed we assume that effects laws do not depend on the performing agent.

We could assume as in [22] that (*positive* and *negative*) effects *preconditions* are complete. *Completeness assumption* can be formulated by means of global assumptions of the form: $\neg\gamma^+(\alpha, p) \wedge \neg p \rightarrow [[\alpha]]\,\neg p$ and $\neg\gamma^-(\alpha, p) \wedge p \rightarrow [[\alpha]]\,p$.

For instance, given the effect law $holdsGun \wedge pointedGun \rightarrow [[pull]]\,scared$ for the action *pulling*, we can establish that: $\neg(holdsGun \wedge pointedGun) \wedge \neg scared \rightarrow [[pull]]\,\neg scared$.

We make a *Consistency assumption* saying that negative effect preconditions and positive effect preconditions must be consistent that is: $\gamma^+(\alpha, p) \rightarrow \neg\gamma^-(\alpha, p)$.[15]

Finally we need to specify *execution preconditions* for our three actions *loading*, *pulling* and *picking-up*: $Pre(pull) = freeHand$, $Pre(load) = freeHand$, $Pre(pickUp) = freeArm \wedge freeHand$.

Given effect preconditions and appropriate assumptions successor state axioms can be specified as standard Situation Calculus requires.

Indeed suppose that $\gamma^-(\alpha, p), \gamma^+(\alpha, p)$ are given and that the *completeness assumption* and *consistency assumption* are made then the following equivalences holds:

$$[[i, \alpha]]\,p \leftrightarrow \neg Goal_i\,\langle\langle i, \alpha\rangle\rangle \top \vee \gamma^+(\alpha, p) \vee (p \wedge \neg\gamma^-(\alpha, p))$$

In this paper we do not investigate any modal regression technique for our logic.[16] Let us only notice that according to the previous logical equivalence the effects of an attempt to do some action α are completely specified by positive effect preconditions ($\gamma^+ - preconditions$), negative effect preconditions ($\gamma^- - preconditions$) and the goal to attempt to do α. Execution preconditions are not mentioned in the successor state axiom. Thus we can argue that under our logical framework every planning task can in principle be reduced to the task of finding the correct sequence of attempts for reaching a given result. Given successor state axioms built on the primitive notion of attempt, for every planning problem there is no need to verify whether execution preconditions hold. This implies that in *LIA* the notion of *execution precondition* is not necessary for formulating action theories.

4.3 Discussion

Being able to characterize attempts and actions, we can provide a further relevant distinction: the distinction between *stable effects* and *successful effects* of an attempt.

Let us go back to our previous example. We have identified the execution preconditions for *pulling* with $Pre(pull) = freeHand$. Moreover we have specified the following effect laws: $holdsGun \wedge loadedGun \wedge pointedGun \wedge freeHand \rightarrow [[pull]]\,wounded$ and $holdsGun \wedge pointedGun \rightarrow [[pull]]\,scared$.

Given definition 1 the first effect law can be rewritten as:

$holdsGun \wedge loadedGun \wedge pointedGun \rightarrow [pull]\,wounded$.

On one side a *stable positive effect* of an attempt to do some action α is a result that an attempt to perform α can produce even if the execution preconditions of action α do not

---

[15] Due to our hypothesis that $\gamma^-(\alpha, p) = \bot$ for each action α and possible effect p such consistency is always the case.

[16] On the problem of how handling regression in dynamic logic see [6].

hold. For instance *scared* is a stable positive effect of the attempt to *pull*. Indeed I can scare you simply by pointing a gun toward you and attempting to pull the trigger.[17] On the other side a *successful positive effect* of an attempt to do some action α is a result that an attempt to perform α causes only if the execution preconditions of action α hold. For instance *wounded* is a *successful positive effect* of the attempt to *pull*. Indeed I can wound you if after pointing the gun toward you and attempting to pull the trigger, I correctly execute the pulling movement (the execution preconditions of *pulling* hold) and the gun is loaded.

Formally:

– *p* is a *successful positive effect* of the attempt to perform the basic action α if and only if $\models_{LIA} \gamma^+(\alpha, p) \rightarrow Pre(\alpha)$.[18]
– *p* is a *stable positive effect* of the attempt to perform the basic action α if and only if there is a model $M \in LIA$ such that $\gamma^+(\alpha, p) \wedge \neg Pre(\alpha)$ is *satisfiable* in M.[19]

In our view there is always some stable effects associated with attempts. Even assuming that the *attempt* to do a basic action is a mere mental process, we can still identify stable effects of the attempt. Indeed under some appropriate preconditions attempting to do something can cause some modification of the mental states of the performing agent (and these modifications do not depend on the fact that the attempt is successful). For example if I believe that after raising my arm my arm goes up and I believe that the preconditions for raising my arm are holding (for instance I believe that my arm is not blocked) then after attempting to raise my arm I believe that my arm is up. This is made explicit by the next theorem of our logic: $Bel_i Pre(\alpha) \wedge Bel_i [i, \alpha] \varphi \rightarrow [[i, \alpha]] Bel_i \varphi$. We can also write plausible effect laws which mention stable effects of attempts at the level of mental attitudes and dispositions of the performing agent. For example if in the morning I am still half-awake and I attempt to stand up then I am awake after this attempt: $asleep \rightarrow [[stand - up]] awake$.

---

[17] We are assuming that you become aware of the risk of being killed only if you can perceive that I am attempting to pull the trigger (fear is not simply triggered by your seeing that I am pointing the gun toward you).

[18] Notice that the class of *successful positive effects* of an attempt to do some basic action α also includes the *intrinsic effect* of (basic) action α [29, 26]. Indeed the *intrinsic effect* of some (basic) action α is the state of affairs that it is guaranteed to hold when α is attempted and the execution preconditions of action α hold. For instance the intrinsic effect of the (basic) action of *raising the arm* is *raised arm,* the intrinsic effect of the (basic) action of *opening the mouth* is *open mouth* and so on... Formally: *p* is a *intrinsic effect* of some basic action α if and only if $\models_{LIA} \gamma^+(\alpha, p) \leftrightarrow Pre(\alpha)$. According to Stoutland also complex actions have intrinsic results. For instance the (complex) action of *opening the door* (opening the door is performed *by* moving the arm in a certain way) has the *door is open* as intrinsic effect.

[19] From the two definitions it follows that the category *stable positive effects* and the category *successful positive effects* are disjoint. Moreover the same kind of definitions apply to *successful negative effects* and *stable negative effects* of an attempt, that is: 1) $\neg p$ is a *successful negative effect* of some basic action α if and only if $\models_{LIA} \gamma^-(\alpha, p) \rightarrow Pre(\alpha)$; 2) $\neg p$ is a *stable negative effect* of some basic action α if and only if it exists a model $M \in LIA$ such that $\gamma^-(\alpha, p) \wedge \neg Pre(\alpha)$ is *satisfiable* in M.

Application to "count as" scenarios In the context of institutions, actions may "count as" implementations of others. Many actions in the social world acquire a different meaning when some institutional fact holds in that world.[20]

For instance take the action of *signing a document* (or the action of *voting*). This action has the same physical realization of the action *writing*, but it differentiates from a simple writing since it is performed under some particular institutional preconditions. It is not the aim of this paper to investigate the exact institutional preconditions which are needed in order to make some physical action an institutional action.[21] Indeed several kinds of conditions concerning social roles, norms etc... must be satisfied: for example the performing agent needs to be entitled to perform the institutional action (he must play some institutional role)[22] and there should be some other agent with institutional power who verifies the correct execution of the action[23] etc... Just consider the following simple example.

$\gamma^+(write, closedHand) = freeHand$

$\gamma^+(write, written) = hasDoc \land holdsPen \land freeHand$

$\gamma^+(write, signed) = hasDoc \land holdsPen \land lastPage \land freeHand \land director$

$\gamma^+(write, voted) = election \land citizen \land holdsPen \land VotingPaper \land freeHand$

According to the previous formulations of positive effect preconditions and negative effect preconditions, if an agent has the hand free and attempts to write then the hand gets closed; if the agent has a document in front of him and a pen is in his hand and his hand is free and he attempts to write then the document gets written on;[24] if the agent is the director of the organization, has the last page of the document in front of him and a pen in the hand, his hand is free, and attempts to write then the document gets signed. Finally if it is election day, the agent is a citizen of the country, a pen is in his hand, his hand is free, has a voting paper in front of him and attempts to write then the agent gives his vote. We do not specify here the completeness laws (their specification is straightforward). Finally we formulate the execution preconditions of the *writing* action: $Pre(write) = freeHand$. Let us only use two abbreviations for indicating the institutional version of the attempt to do action α and the institutional version of action α:

$\langle\langle Ist - \alpha \rangle\rangle \varphi =_{def} \langle\langle \alpha \rangle\rangle \varphi \land Ist(\alpha)$;

$\langle Ist - \alpha \rangle \varphi =_{def} \langle \alpha \rangle \varphi \land Ist(\alpha)$ which can be rewritten as

$\langle Ist - \alpha \rangle \varphi =_{def} \langle\langle \alpha \rangle\rangle \varphi \land Pre(\alpha) \land Ist(\alpha))$ where $Ist(\alpha)$ denotes all conditions which make α become an *institutional action* or to "count as" an *institutional action* (we call them *institutional preconditions*).[25]

---

[20] See also [14] for a formal approach to institutional actions.

[21] On this point see [24, 27].

[22] In order to marry a couple the agent must be a priest.

[23] In signing a contract is not enough to sign the document at the correct place. An institutional witness (the notary) is needed who verifies the correct execution of the procedure.

[24] Notice that in common sense language *writing* is not a proper basic action. Indeed agent generally *writes by* performing a *certain movement with the hand.* Thus our label "write" denotes rather the basic action (bodily movement) on which the complex action of *writing* is based.

[25] We assume that $Ist(\alpha)$ is a function returning some classical formula Φ that is: $Ist : ACT \to PROP$.

Since the institutional version of a basic action α is physically identical to α we can safely assume that the execution preconditions of α and the execution preconditions of the institutional version of α are identical.

Notice that basic actions might have more than one institutional version. For instance the basic action of *writing* "counts as" the institutional action of *signing* under some institutional preconditions whereas it "counts as" the institutional action of *voting* under some different institutional preconditions. In order to account for different kinds of institutional actions based on the same basic action, $Ist(\alpha)$ must denote several alternative groups of institutional preconditions. For instance in order to distinguish the institutional action of *signing* from the institutional action of *voting* we must operate at the level of institutional preconditions and identify different subsets of institutional preconditions corresponding to each institutional version of the basic action. Let us consider the simple scenario where the action of *writing* has only two institutional versions (*signing* and *voting*). $Ist(\alpha)$ denotes only two subsets of institutional preconditions:

$Ist(write) = (lastPage \wedge director) \vee (election \wedge citizen \wedge VotingPaper)$.

Having introduced $Ist(write)$, the institutional version of the action *writing* and the institutional version of the attempt to *write* can be specified:

$\langle\langle Ist - write \rangle\rangle \varphi =_{def} \langle\langle write \rangle\rangle \varphi \wedge (lastPage \wedge director) \vee (election \wedge citizen \wedge VotingPaper)$;

$\langle Ist - write \rangle \varphi =_{def} \langle write \rangle \varphi \wedge (lastPage \wedge director) \vee (election \wedge citizen \wedge VotingPaper)$.

The action *signing* (the attempt to *sign*) and the action *voting* (the attempt to *vote*) are specific institutional versions of the action *writing* (the attempt to *write*), that is they are defined as instances of *writing* under some specific subsets of the set of institutional preconditions of *writing*. Indeed:

$\langle sign \rangle \varphi =_{def} \langle write \rangle \varphi \wedge lastPage \wedge director$;

$\langle\langle sign \rangle\rangle \varphi =_{def} \langle\langle write \rangle\rangle \varphi \wedge lastPage \wedge director$;

$\langle vote \rangle \varphi =_{def} \langle write \rangle \varphi \wedge election \wedge citizen \wedge VotingPaper$

$\langle\langle vote \rangle\rangle \varphi =_{def} \langle\langle write \rangle\rangle \varphi \wedge election \wedge citizen \wedge VotingPaper$.

This means that: 1) getting φ after my attempt to perform the action of *signing* (or after performing the action of *signing*) means being the director and getting φ after the attempt to write my name (or after the action of *writing* my name) on the last page of the document; 2) getting φ after my attempt to perform the action of *voting* (or after performing the action of *voting*) means being a citizen of the country on the election day and getting φ after the attempt to write (or after the action of *writing*) on the voting paper.

The fact that *signing* and *voting* are specific institutional versions of the basic action *writing* is made explicit by the following four formal consequences of our definitions: a) $\langle sign \rangle \varphi \rightarrow \langle Ist - write \rangle \varphi$; b) $\langle vote \rangle \varphi \rightarrow \langle Ist - write \rangle \varphi$; c) $\langle\langle sign \rangle\rangle \varphi \rightarrow \langle\langle Ist - write \rangle\rangle \varphi$ and d) $\langle\langle vote \rangle\rangle \varphi \rightarrow \langle\langle Ist - write \rangle\rangle \varphi$.

Indeed if I attempt to sign (or to vote) and φ holds after this attempt then I also attempt to perform the institutional version of *writing* and φ holds after the attempt, and if I sign (or vote) and φ holds after this action then I also attempt to perform the institutional version of *writing* and φ holds afterward.

Moreover on the basis of the previous definitions of institutional action and institutional attempt we get (besides the validity $[[i, \alpha]] \varphi \rightarrow [i, \alpha] \varphi$) the following four validities: a) $[[\alpha]] \varphi \rightarrow [[Ist - \alpha]] \varphi$; b) $[\alpha] \varphi \rightarrow [Ist - \alpha] \varphi$; c) $[[\alpha]] \varphi \rightarrow [Ist - \alpha] \varphi$ and d) $[[Ist - \alpha]] \varphi \rightarrow [Ist - \alpha] \varphi$.

It is evident that the relation among attempt and physical (basic) action is symmetrical to the relation among physical (basic) action and institutional action. Indeed if φ is a consequence of the attempt to do action α then φ is also a consequence of the successful execution of the basic action α and if φ is a consequence of performing the basic action α then φ is also a consequence of performing the institutional version of action α (moreover if φ is a consequence of the attempt to do action α then φ is also a consequence of the attempt to do the institutional version of action α; if φ is a consequence of the attempt to do the institutional version of action α then φ is also a consequence of doing the institutional version of action α).

Besides the distinction between *stable effects* and *successful effects* of an attempt we can provide the distinction among *institutional effects* and *natural effects* of an attempt. We define *institutional positive effects* of a given attempt to perform action α all those positive effects that the attempt to perform α causes only if the institutional preconditions of α hold. We distinguish these effects from *natural positive effects* of a given attempt to perform action α which are those positive effects that the attempt to perform α causes even if the institutional preconditions of α do not hold.

Formally:

- p is a *institutional positive effect* of the attempt to perform the basic action α if and only if $\models_{LIA} \gamma^+(\alpha, p) \rightarrow Ist(\alpha)$.
- p is a *natural positive effect* of the attempt to perform the basic action α if and only if there is a model $M \in LIA$ such that $\gamma^+(\alpha, p) \wedge \neg Ist(\alpha)$ is *satisfiable* in M.[26]

To sum up, we can distinguish four different sub-categories of effects of an attempt:

1. Institutional and stable effects of an attempt.
2. Natural and stable effects of an attempt.
3. Institutional and successful effects of an attempt.
4. Natural and successful effects of an attempt.

Going back to our initial example of *pulling* action, *scared* is a natural and stable effect of the attempt to *pull*, *wounded* (or *dead*) is a natural and successful effect of the attempt to *pull*. Finally *attempted homicide* (or *attempted capital punishment*) is an institutional and stable effect of the attempt to *pull* and *homicide* (or *capital punishment)* is an institutional and successful effect of the attempt to *pull*.[27]

---

[26] From the definition it follows that the category *institutional positive effects* and the category *natural positive effects* are also disjoint. Again the same kind of definitions apply to *natural negative effects* and *institutional negative effects* of an attempt, that is : 1) $\neg p$ is a *institutional negative effect* of some basic action α if and only if $\models_{LIA} \gamma^-(\alpha, p) \rightarrow Ist(\alpha)$; 2) $\neg p$ is a *natural negative effect* of some basic action α if and only if it exists a model $M \in LIA$ such that $\gamma^-(\alpha, p) \wedge \neg Ist(\alpha)$ is *satisfiable* in M.

[27] *Attempted homicide* and *homicide* are recognized as violations of the law in every civil society when a private person kills someone and is not entitled to do so (therefore they are institutional

5 Concluding remarks: attempt and present-directed intention

In this final section we discuss additional properties of attempts, introduce the notion of present-directed intention and present some formal relations between the two concepts. The formula $\langle\langle i, \alpha \rangle\rangle \top \leftrightarrow Bel_i \langle\langle i, \alpha \rangle\rangle \top$ is valid and establishes that our notion of *attempt* is related with agent's awareness (when agent i attempts to do some action α, he believes to be attempting and viceversa).

The valid formula $Goal_i [[i, \alpha]] \perp \rightarrow [[i, \alpha]] \perp$ establishes that if agent i has the goal to *avoid* to attempt to perform action α then action α is not attempted by agent i. We introduce next the notion of *present-directed intention* [1, 25].

DEFINITION 2: Present-directed Intention. $PDI_i(\alpha) =_{def} Goal_i \langle i, \alpha \rangle \top$

The definition of present-directed intention is intimately related with axiom 11b stating the logical equivalence of the goal to attempt to do action α and attempt itself. Indeed the present-directed intention stage coincides with the stage at which the agent triggers the motor intentional behaviour. According to the present model when agent has the present-directed intention to do some action α: 1) he can attempt to do α (he can send the action to execution); 2) he is aware of this possibility; 3) he has the goal to attempt to do α, 4) he has the goal that the preconditions for executing action α hold, 5) he cannot believe that the preconditions for executing action α do not hold. [28] The previous statements are formally expressed by the following valid formulas of our logic:

a) $PDI_i(\alpha) \rightarrow \langle\langle i, \alpha \rangle\rangle \top$;
b) $PDI_i(\alpha) \rightarrow Bel_i \langle\langle i, \alpha \rangle\rangle \top$;
c) $PDI_i(\alpha) \rightarrow Goal_i \langle\langle i, \alpha \rangle\rangle \top$;
d) $PDI_i(\alpha) \rightarrow Goal_i Pre(i, \alpha)$;
e) $PDI_i(\alpha) \rightarrow \neg Bel_i \neg Pre(i, \alpha)$.

Finally just pay attention to the distinction among *the goal to attempt to do a certain action α ($Goal_i \langle\langle i, \alpha \rangle\rangle \top$)* and the notion of *present-directed intention to do a certain action α*. Notice that the inverse direction of previous formula c) is not a valid statement: in our logic an agent can have the goal to attempt to do α without having the present-directed intention to do α.[29]

References

1. Bratman, M. E. (1987). *Intentions, plans and practical reason*. Cambridge, MA: Harvard University Press.

---

effects of a given action). On the other hand a firing squad executing a death penalty is entitled to kill and the effect of its action is recognized by the institution either as a *capital execution* or as an *attempted capital execution*.

[28] Notice that an agent can have the goal to attempt to do α believing that the preconditions for executing α do not hold ($Goal_i \langle\langle i, \alpha \rangle\rangle \top \wedge Bel_i \neg Pre(i, \alpha)$ is satisfiable).

[29] Suppose that "Brett promises to pay Belton fifty dollars if Belton *attempts* to solve a certain chess problem within five minutes". Imagine that Brett assures Belton that he need not actually solve the problem for getting the fifty dollars. According to [17] it is plausible to say that Belton is motivated to attempt to solve problem even if he does not intend to solve the problem.

2. Castilho, M. A., Gasquet, O., Herzig, A. (1999). Formalizing action and change in modal logic I: the frame problem. *Journal of Logic and Computation*, 9(5), pp. 701-735.

3. Chisholm, R. M. (1966). Freedom and Action. In Keith Lehrer (Ed.), *Freedom and Determinism*, Random House; New York, NY, pp. 105-39.

4. Cohen, P. R. , Levesque, H. J. (1990). Intention is choice with commitment. *Artificial Intelligence*, 42, pp. 213-261.

5. Danto, A (1965). What we can do. *The Journal of Philosophy,* 60, pp. 435-445.

6. Demolombe, R., Herzig, A., Varzinczak, I. (2003). Regression in modal logic. *Journal of Applied Non-Classical Logics,* 13, pp. 165-185.

7. Fitting, M. (1983). *Proof Methods for Modal and Intuitionistic Logics*. D. Reidel, Dordrecht.

8. Gabbay, D., Pnueli, A., Shelah, S., Stavi, J. (1980). On the temporal analysis of fairness. In *Proceedings 7th ACM Symposium on Principles of Programming Languages*, pp. 163-173.

9. Goldblatt, R. (1992). *Logics of Time and Computation, 2nd edition*. CSLI Lecture Notes, Stanford, California.

10. Goldman, A. (1970). *A Theory of Human Action*. Englewood Cliffs: Prentice-Hall.

11. Harel, D., Kozen D., Tiuryn, J.(2000). *Dynamic Logic*. Cambridge, MA: MIT Press.

12. Herzig, A., Longin, D. (2004). C&L Intention Revisited. In *Proceedings of KR2004*, pp. 527-535.

13. Hornsby, J. (1980). *Actions*. Routledge & Kegan Paul, London.

14. Jones, A., Sergot, M. J. (1996). A formal characterisation of institutionalised power. *Journal of the IGPL*, 4(3), pp. 429–445.

15. Lorini, E. (2006). A logic of Intention and Attempt. *Technical Report*, Institute of Cognitive Science and Technologies-CNR, Rome.

16. McCann, H. J. (1974). Volition and Basic Action. *The Philosophical Review*, 83, pp. 451 473.

17. Mele, A. R. (1992). *Springs of action.* Oxford University Press, New York.

18. Meyer, J.J. Ch., van der Hoek, W., van Linder, B. (1999). A Logical Approach to the Dynamics of Commitments. *Artificial Intelligence*, 113(1-2), pp. 1-40.

19. O'Shaughnessy, B. (1973). Trying (as the Mental Pineal Gland.) *The Journal of Philosophy*, 70, pp. 365-86.

20. Pacuit, E., Parikh, R., Cogan, E. (2005). The Logic of Knowledge Based Obligation. To appear in *Knowledge, Rationality and Action*.

21. Rao, A. S., Georgeff M. P. (1991). Modelling rational agents within a BDI-architecture. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*, Morgan Kaufmann Publishers, San Mateo, CA.

22. Reiter, R. (2001). *Knowledge in action: logical foundations for specifying and implementing dynamical systems*. Cambridge, MA: MIT Press.

23. Santos, F., Carmo, J., Jones, A. (1997). Action concepts for describing organised interaction. In *Proceedings Thirtieth Annual Hawai International Conference on System Sciences*, pp. 373-382.

24. Searle, J. R. (1995). *The construction of social reality*. Free Press, New York.

25. Searle, J. R. (1983). *Intentionality*. Cambridge University Press.

26. Stoutland, F. (1968). Basic Actions and Causality. *Journal of Philosophy*, 65, pp. 467-475.

27. Tummolini, L., Castelfranchi, C. (in press). The cognitive and behavioral mediation of institutions: Towards an account of institutional actions. *Cognitive Systems Research*, 7(2-3).

28. Vanderveken, D. (2003). Attempt and action generation: towards the foundations of the logic of action. *Cahiers d'pistmologie*, 293.

29. Von Wright, G. H. (1963). *Norm and Action*. London: Routledge and Kegan Paul.