Modeling Attempt and Action Failure in Probabilistic stit Logic

Jan Broersen

Technical Report UU-CS-2011-002
January 2011

Department of Information and Computing Sciences
Utrecht University, Utrecht, The Netherlands
www.cs.uu.nl

ISSN: 0924-3275

Department of Information and Computing Sciences
Utrecht University
P.O. Box 80.089
3508 TB Utrecht
The Netherlands

Modeling Attempt and Action Failure in Probabilistic stit Logic

Abstract

We define an extension of stit logic that encompasses subjective probabilities representing beliefs about simultaneous choice exertion of other agents. The formalism enables us to express the notion of 'attempt' as a choice exertion that maximizes the chance of success with respect to an action effect. The notion of attempt (or effort) is central in philosophical and legal discussions on responsibility and liability.

1 Introduction

The predominant formal theory of agency in philosophy is *stit* theory [3]. *Stit* theory gives an elegant and thoroughly elaborated view on the question of how agents exercise control over the courses of events that constitute our dynamic world. Also *stit* theory provides a view on the fundamentals of cooperation and the limits and possibilities of acting together and / or in interaction. Recently, *stit* theory attracted the attention of computer scientist who are interested in deontic logic and logic for the specification of multi-agent systems [5, 6, 2].

One shortcoming of *stit* theory is that its central notion of choice exertion is one that assumes that a choice is always successful. But it is highly unrealistic for formalisms aimed at modeling (group) choice of intelligent agents to assume that action can never fail. This problem cannot be solved by making the connection with dynamic logic or the situation calculus, since these formalisms also lack a theory about how actions can be unsuccessful.

This paper assumes we measure success of action against an agent's beliefs about the outcome of its choice. So, the perspective is an internal, subjective one, and the criterion of success is formed by an agent's beliefs about its action. To represent these beliefs we choose here to use probabilities. In particular, we will represent beliefs about simultaneous choice exertion of other agents in a system as subjective probabilities. Several choices have to be made. We will pose that an agent can never be mistaken about its own choice, but that it can be mistaken about choices of others. The actual action performed results from a simultaneous choice exertion of all agents in the system. Then, if an agent can be mistaken about the choices of other agents (including possibly an agent with special properties called 'nature'), the action can be unsuccessful. As a very basic example, consider the opening of a door. An agent exercises its choice to open the door. It cannot be mistaken about that: it knows what it chooses to do. It does this under the belief that there is no other agent on the other side exercising its choice to keep the door closed. So it assigns a low probability to such a choice of any other agent. However, here the agent can be mistaken. And here comes in the notion of unsuccessful action modeled in this paper: as it turns out, in the situation described there actually is an agent at the other side of the door choosing to keep it closed and the agent's opening effort is unsuccessful.

To model this, we endow *stit* theory with probabilities in the object language, enabling us to say that an agent exercises a choice for which it believes to have a chance higher than c to see to it that φ results in the next state.

As far as we know, our proposal is the first combining *stit* logic and probability. Possibly unsuccessful actions have been considered in the context of Markov Decision Processes, temporal logic and ATL [10]. Two differences with the present work are that here we start from the richer *stit* theory and that we focus on fundamental properties of the resulting logic in stead of on issues related to planning, policy generation or model checking. An independent motivation for considering action with a chance of success comes from

the relation between *stit* theory and game theory. Kooi and Tamminga [11] investigate how to characterize pure strategy equilibria as *stit* formulas. An extension of *stit* logic with probabilistic effects would enable us to also characterize mixed strategy equilibria.

The distinction between successful and possibly unsuccessful action naturally leads to the concept of 'attempt'. Attempts are choices that can be unsuccessful. But an attempt is more than just a choice with a certain probability of success. Consider a choice with a chance p of bringing about φ. Then, necessarily, the chance of bringing about $\neg\varphi$ is $1 - p$. Then, if an attempt would be a choice with some probability different from 0 or 1 of bringing about a certain effect, in this case, the same choice is both an attempt for φ and an attempt for $\neg\varphi$. This is counter intuitive; we only call something an attempt if the agent exercising the choice took the *best* choice available relative to the effect it tries to obtain. This is even completely unrelated to the absolute chance of success for the choice exercised in the attempt. For instance, the buying of a lottery ticket can be an attempt to win the jackpot, even though the chance of success is very low. What this shows, is that attempt is a comparative notion. Here we will model it as the exertion of a choice that in comparison with other choices possible in a situation is maximal with respect to the chance of success of obtaining a condition φ.

The notion of attempt we consider differs considerably from the one studied in [12], where the focus is on the idea that an attempt is a 'mental' action *not* having direct but only having indirect consequences for an agent's environment. One crucial way in which our logic is different from the one sketched in [14] is that we explicitly model the epistemic attitude in attempt using subjective probabilities.

2 The base logic: XSTIT^p

In this section we define the base logic, which is a variant of Broersen's XSTIT logic that we call XSTIT^p. The difference with XSTIT is embodied by an axiom schema concerning modality-free propositions p, which explains the name. Another difference with XSTIT is that we do not define the semantics in terms of *relations*, but in terms of functions. We introduce h-relative effectivity functions, which specialize the notion of effectivity function from Coalition Logic [13] by defining choices relative to histories. The function-based semantics explains the formalism better than Broersen's semantics in terms of relations.

Definition 2.1 *Given a countable set of propositions P and $p \in P$, and given a finite set Ags of agent names, and $A \subseteq Ags$, the formal language $\mathcal{L}_{\mathsf{XSTIT}^p}$ is:*

$$\varphi \quad := \quad p \mid \neg\varphi \mid \varphi \wedge \varphi \mid \Box\varphi \mid [A \ \mathsf{xstit}]\varphi \mid X\varphi$$

Besides the usual propositional connectives, the syntax of XSTIT^p comprises three 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* [8]. Another way of talking about this operator is to say that it expresses that φ is 'settled'. 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 state.

Definition 2.2 *A function-based XSTIT^p-frame is a tuple $\langle S, H, E\rangle$ such that*[1]*:*

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 of the form $\ldots s_{-2}, s_{-1}, s_0, s_1, s_2, \ldots$ with $s_x \in S$ for $x \in \mathbb{Z}$. 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 $s = prec(s', h)$. Furthermore:*
   a. *if $s \in h$ and $s \in h'$ then $prec(s, h) = prec(s, h')$*
3. *$E : S \times H \times 2^{Ags} \mapsto 2^S$ is an h-effectivity function yielding for a group of agents A the set of next static states allowed by the joint actions taken by the agents in the group A relative to a history*

[1] 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* $s \notin h$ *then* $E(s,h,A) = \emptyset$

b. *if* $s' \in E(s,h,A)$ *then* $\exists h' : s' = succ(s,h')$

c. $succ(s,h) \in E(s,h,A)$

d. $\exists h' : s' = succ(s,h')$ *if and only if* $\forall h :$ *if* $s \in h$ *then* $s' \in E(s,h,\emptyset)$

e. *if* $s \in h$ *then* $E(s,h,Ags) = \{succ(s,h)\}$

f. *if* $A \supset B$ *then* $E(s,h,A) \subseteq E(s,h,B)$

g. *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 2.2 above, we refer to the states s as 'static states'. This is to distinguish them from 'dynamic states', which are combinations $\langle s,h \rangle$ of static states and histories. Dynamic states function as the elementary units of evaluation of the logic. This means that the basic notion of 'truth' in the semantics of this logic is about dynamic conditions concerning choice exertions. This distinguishes *stit* from logics like Dynamic Logic and Coalition Logic whose central notion of truth concerns static conditions holding for static states.

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'. Condition **3.a** above states that h-effectivity is empty for history-state combinations that do not form a dynamic state. Condition **3.b** ensures that next state effectivity as seen from a current state s does not contain states s' that are not reachable from the current state through some history. Condition **3.c** states that the static state next of some other static state on a history is always in the effectivity set relative to that history state pair for any group of agents. Condition **3.d** above states that any next state is in the effectivity set of the empty set and vice versa. This underlines the special role of the empty set of agents. On the one hand, the empty set is powerless, since it does not have genuine alternatives for choices, like agents generally do. On the other hand, it is almighty, since whatever is determined by the effectivity of the empty set *must* occur in next states. We may refer to the empty set of agents as 'nature' and to its effectivity as 'causation'. Condition **3.e** 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 [1] and Coalition logic (CL for short). However, although **3.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'. In this formalism dynamic states are the units of evaluation. In ATL and CL, static states are the units of evaluation[2]. Conditions **3.f** expresses coalition (anti-)monotony. The second subset relation in this property is not strict, because we can always add a powerless agent with the same properties as the empty set of agents: it does not have real choices and always 'goes with the flow'. This increases the number of agents while leaving the choices of all agents as they are. Condition **3.g** above states that simultaneous choices of different agents never have an empty intersection. This is the central condition of 'independence of agency'. It reflects 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.

The conditions on the frames are not as tight as the conditions in the classical *stit* formalisms of Belnap, Perloff and Horty [3]. Appart from the crucial difference concerning the effect of actions (in XSTIT^p actions take effect in next states), the classical *stit* formalisms assume conditions that in our meta-language can be represented as:

h. $E(s,h,A) \neq E(s,h',A)$ implies
$E(s,h,A) \cap E(s,h',A) = \emptyset$

i. $E(s,h,A \cup B) = E(s,h,A) \cap E(s,h,B)$

Condition **h.** says that the choices of a group A are mutually disjoint. Condition **i.** says that the choices of a group are exactly the intersections of the choices of its sub-groups. Condition **i.** is strictly stronger than the coalition (anti-)monotony property **3.f**, which only says that the choices of a group are *contained* in the choices of its sub-groups. Since they result in much tidier pictures, in the example visualization of

[2]This is part of the reason why the coalition logic modality $[A]\varphi$ is not definable as $\Diamond[A \text{ xstit}]\varphi$ in XSTIT^p.

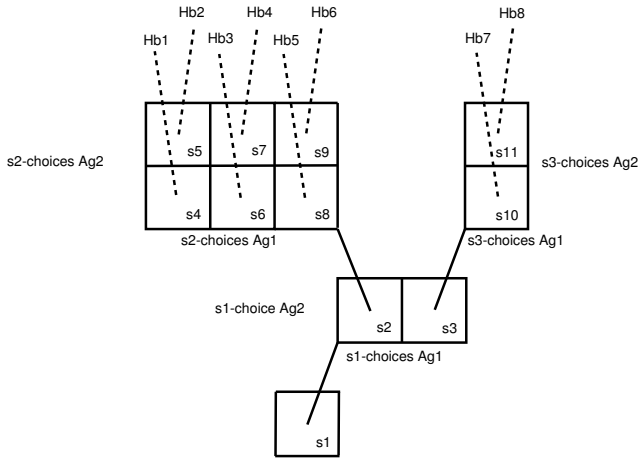

Figure 1: visualization of a partial two agent XSTIT^p frame

a frames we consider below, we assume both these conditions. However, we do not include them in the formal definition of the frames, because both conditions are not modally expressible (e.g., in modal logic we can give axioms characterizing that an intersection is non-empty, but we cannot characterize that an intersection is empty). This means that they will not have an effect on our modal logic of agency whose semantics we will define in terms of the above frames.

Figure 1 visualizes a frame of the type defined by definition 2.2. The small squares are static states in the effectivity sets of $E(s, h, Ags)$. Combinations of static states and histories running through 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. Choice exertion in this 'bundled' semantics is thought of as the separation of two bundles of histories: one bundle ensured by the choice exercised and one bundle excluded by that choice.

We now define models by adding a valuation of propositional atoms to the frames of definition 2.2. We impose that all dynamic state relative to a static state evaluate atomic propositions to the same value. This reflects the intuition that atoms, and modality-free formulas in general do not represent dynamic information. Their truth value should thus not depend on a history but only on the static state. This choice does however make the situation non-standard. It is a constraint on the models, and not on the frames.

Definition 2.3 *A frame $\mathcal{F} = \langle S, H, E\rangle$ 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 $\pi : P \longrightarrow 2^S$ assigning to each atomic proposition the set of static states relative to which they are true.*

We evaluate truth with respect to dynamic states built from a dimension of histories and a dimension of static states.

Definition 2.4 *Relative to a model $\mathcal{M} = \langle S, H, E, \pi\rangle$, truth $\langle s, h\rangle \models \varphi$ of a formula φ in a dynamic state $\langle s, h\rangle$, with $s \in h$, is defined as:*

$$
\begin{array}{lcl}
\langle s,h\rangle \models p & \Leftrightarrow & s \in \pi(p) \\
\langle s,h\rangle \models \neg\varphi & \Leftrightarrow & \textit{not } \langle s,h\rangle \models \varphi \\
\langle s,h\rangle \models \varphi \wedge \psi & \Leftrightarrow & \langle s,h\rangle \models \varphi \textit{ and } \langle s,h\rangle \models \psi \\
\langle s,h\rangle \models \Box\varphi & \Leftrightarrow & \forall h' : \textit{ if } s \in h' \textit{ then } \langle s,h'\rangle \models \varphi \\
\langle s,h\rangle \models X\varphi & \Leftrightarrow & \textit{if } s' = succ(s,h) \textit{ then } \langle s',h\rangle \models \varphi \\
\langle s,h\rangle \models [A \textit{ xstit}]\varphi & \Leftrightarrow & \forall s', h' : \textit{if } s' \in E(s,h,A) \textit{ and} \\
 & & s' \in h' \textit{ then } \langle s',h'\rangle \models \varphi
\end{array}
$$

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. Now we proceed with the axiomatization of the base logic.

Definition 2.5 *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^p*:*

(p)	$p \rightarrow \Box p$ *for* p *modality free*
	S5 for $\Box$
	KD for each $[A$ *xstit*$]$
(Det)	$\neg X \neg \varphi \rightarrow X \varphi$
($\emptyset$ = SettX)	$[\emptyset$ *xstit*$]\varphi \leftrightarrow \Box X \varphi$
(Ags = XSett)	$[Ags$ *xstit*$]\varphi \leftrightarrow X \Box \varphi$
(C-Mon)	$[A$ *xstit*$]\varphi \rightarrow [A \cup B$ *xstit*$]\varphi$
(Indep-G)	$\Diamond[A$ *xstit*$]\varphi \wedge \Diamond[B$ *xstit*$]\psi \rightarrow$
	$\Diamond([A$ *xstit*$]\varphi \wedge [B$ *xstit*$]\psi)$ *for*
	$A \cap B = \emptyset$

Theorem 2.1 *The Hilbert system of definition 2.5 is complete with respect to the semantics of definition 2.4.*

The proof strategy is as follows. First we establish completeness of the system *without* the axiom $p \rightarrow \Box p$, relative to the frames of definition 2.2. All remaining axioms are in the Sahlqvist class. This means that all the axioms are expressible as first-order conditions on frames and that together they are complete with respect to the frame classes thus defined, cf. [4]. It is easy to find the first-order conditions corresponding to the axioms, for instance, by using the on-line SQEMA system [7]. So, now we know that every formula consistent in the slightly reduced Hilbert system has a model based on an abstract frame. Left to show is that we can associate such an abstract model to a concrete model based on an XSTIT^p frame as given in definition 2.2. This takes some effort, since we have to associate worlds in the abstract model to dynamic states in the frames of definition 2.2 and check all the conditions of definition 2.2 against the conditions in the abstract model (**3.c** corresponds with the D axiom, **3.d** corresponds to ($\emptyset$ = SettX), **3.e** to (Ags = XSett), **3.f** to (C-Mon), **3.g** to (Indep-G)). Once we have done this, we have established completeness of the axioms relative to the conditions on the frames. Now the second step is to add the axiom $p \rightarrow \Box p$. This axiom does not have a corresponding frame condition. Indeed, the axiom expresses a condition on the models. But then, to show completeness, we only have to show that we can always find a model obtained by the construction just described that satisfies the axiom $p \rightarrow \Box p$. But this is straightforward. From all the possible models resulting from the first step, we select the ones where propositional atoms in dynamic states based on the same static state have identical valuations. Since consistent formulas also have to be consistent with the axiom $p \rightarrow \Box p$ for any non-modal formula p, we can always do that. This means that a satisfying model for a consistent formula is always obtainable in this way and that completeness is preserved.

3 Choice with a bounded chance of success

We introduce operators $[\{ag\}\ \mathsf{xstit}^{\geq c}]\varphi$ with the intended meaning that agent ag exercises a choice for which it believes to have a chance of at least c of bringing about φ. Roughly, the semantics for this new operator is as follows. We start with the multi-agent *stit*-setting of the previous section. Now to the semantic structures we add belief functions such that in the little game-forms, as visualized by figure 1, for each choice of an agent ag we have available the subjective probabilities applying to the choices of the other agents in the system. For agent ag the sum of these probabilities over the choices of each particular other agent in the system add up to one. So, the probabilities represent agent ag's beliefs concerning what choices are exerted simultaneously by other agents. In terms of the subjective probability function we define for each choice the sum of the probabilities for each of the choices of all other agents in the system leading to a situation obeying φ.

For the definition of the probabilistic frames, we first define an augmentation function returning the choices a group of agent has in a given state.

Definition 3.1 *The range function* $Range : S \times 2^{Ags} \mapsto 2^{2^S}$ *yielding for a state* s *and a group of agents* A*, the combined choices these agents have in* s *is defined as:*
$Range(s, A) = \{Ch \mid \exists h : Ch = E(s, h, A)\}$

A range function corresponds to what in Coalition Logic is called an 'effectivity function'. Now we are ready to define the probabilistic *stit* frames.

Definition 3.2 *A probabilistic* XSTIT^p*-frame is a tuple* $\langle S, H, E, B\rangle$ *such that:*

1. $\langle S, H, E\rangle$ *is a function based* XSTIT^p*-frame*
2. $B : S \times Ags \times Ags \times 2^S \mapsto [0, 1]$ *is a subjective probability function such that* $B(s, ag_1, ag_2, Ch)$ *expresses agent 1's believe that in static state* s *agent 2 performs a choice resulting in one of the static states in* Ch*. We apply the following constraints.*
   - **a.** $B(s, ag, ag', Ch) = 0$ *if* $Ch \notin Range(s, \{ag'\})$
   - **b.** $B(s, ag, ag', Ch) > 0$ *if* $Ch \in Range(s, \{ag'\})$
   - **c.** $B(s, ag, ag, Ch) = 1$
   - **d.** $\sum_{Ch \in Range(s,\{ag'\})} B(s, ag, ag', Ch) = 1$

Condition **2.a** says that agents only assign non-zero subjective probabilities to choices other agents objectively have. Condition **2.b** says these probabilities are strictly larger than zero. Condition **2.c** says that agents always know what choice they exercise themselves. Note that this is not the same as claiming that agents always know what action they perform (which is not the case in our conceptualization). We already explained this difference between choice and action in section 2. Condition **2.d** says that the sum of the subjective probabilities over the possible choices of other agents add up to 1.

In the sequel we will need an augmentation function yielding for an agent and an arbitrary next static state the chance an agent ascribes to the occurrence of this state (given its belief, i.e., subjective probabilities about simultaneous choice exertion of other agents). For this, we first need the following proposition.

Proposition 3.1 *For any static state* $s' = E(s, h, Ags)$ *in the static state* s *there is a unique 'choice profile' determining for any agent* ag *in the system a unique choice* $Ch = E(s, h, \{ag\})$ *relative to* s *and* h*.*

The proposition follows from the conditions posed on the frames in definition 2.2. Now we can define the subjective probabilities agents assign to possible system outcomes. Because of the idea of independence of agency, we can multiply the chances for the choices of the individual agents relative to the system outcome (the resulting static state). Note that this gives a new and extra dimension to the notion of independence that is not available in standard *stit* theories.

Definition 3.3 $BX : S \times Ags \times S \mapsto [0, 1]$ *is a subjective probability function concerning possible next static states, defined by*
$BX(s, ag, s') = \prod_{ag' \in Ags} B(s, ag, ag', E(s, h, \{ag'\}))$ *if* $\exists h : s' = E(s, h, Ags)$ *or* 0 *otherwise.*

It is basic arithmetic to establish that also the subjective probabilities an agent ag assigns to the choices of the complete set of agents Ags add up to 1.

Proposition 3.2 $\sum_{s' \in Range(s,Ags)} BX(s, ag, s') = 1$

Now before we can define the notion of 'seeing to it under a minimal probability of success' formally as a truth condition on the frames of definition 3.2 we need to do more preparations. First we observe that the intersection of the h-effectivity functions of complementary groups of agents yields a unique static state. This justifies the following definition, that establishes a function characterizing the static states next of a given state that satisfy a formula φ relative to the current choice of an agent.

Definition 3.4 *The 'possible next static φ-states' function $PosX : S \times H \times Ags \times \mathcal{L} \mapsto 2^S$ which for a state s, a history h, an agent ag and a formula φ gives the possible next static states obeying φ given the agent's current choice determined by h, is defined by: $PosX(s, h, ag, \varphi) = \{s' \mid E(s, h, \{ag\}) \cap Ch = \{s'\}$ for $Ch \in Range(s, Ags \setminus \{ag\})$ and $\langle s', h' \rangle \models \varphi$ for all h' with $s' \in h'\}$.*

Now we can formulate the central 'chance of success' (CoS) function that will be used in the truth condition for the new operator. The chance of success relative to a formula φ is the sum of the chances the agent assigns to possible next static states validating φ.

Definition 3.5 *The chance of success function $CoS : S \times H \times Ags \times \mathcal{L} \mapsto [0, 1]$ which for a state s and a history h an agent ag and a formula φ gives the chance the agent's choice relative to h is an action resulting in φ is defined by: $CoS(s, h, ag, \varphi) = 0$ if $PosX(s, h, ag, \varphi) = \emptyset$ or else $CoS(s, h, ag, \varphi) = \sum_{s' \in PosX(s,h,ag,\varphi)} BX(s, ag, s')$.*

Extending the probabilistic frames of definition 3.2 to models in the usual way, the truth condition of the new operator is defined as follows.

Definition 3.6 *Relative to a model $\mathcal{M} = \langle S, H, E, B, \pi \rangle$, truth $\langle s, h \rangle \models [\{ag\}\ \mathsf{xstit}^{\geq c}]\varphi$ of a formula $[\{ag\}\ \mathsf{xstit}^{\geq c}]\varphi$ in a dynamic state $\langle s, h \rangle$, with $s \in h$, is defined as:*

$$\langle s, h \rangle \models [\{ag\}\ \mathsf{xstit}^{\geq c}]\varphi \Leftrightarrow CoS(s, h, ag, \varphi) \geq c$$

Three validities reflecting interesting properties of this semantics are the following.

Proposition 3.3 *Each instance of any of the following formula schemas is valid in the semantics following from definitions 2.4 and 3.6.*

a. $[\{ag\}\ \mathsf{xstit}]\varphi \leftrightarrow [\{ag\}\ \mathsf{xstit}^{\geq 1}]\varphi$
b. $[\{ag\}\ \mathsf{xstit}^{\geq 0}]\varphi$
c. $[\{ag\}\ \mathsf{xstit}^{\geq c}]\varphi \rightarrow [\{ag\}\ \mathsf{xstit}^{\geq k}]\varphi$ *for* $c \leq k$

Validity schema **a.** shows that the probabilistic *stit* operator we gave in definition 3.6 faithfully generalizes the *stit* operator of our base XSTIT^p system: the objective *stit* operator $[\{ag\}\ \mathsf{xstit}]\varphi$ discussed in section 2 comes out as the probabilistic *stit* operator assigning a probability 1 to establishing the effect φ. This is very natural. Where in the standard *stit* setting we can talk about 'ensuring' a condition, in the probabilistic setting we can only talk about establishing an effect with a certain lower bound on the probability of succeeding. Schema **b.** expresses that any effect φ is always brought about with a probability of zero or higher, which should clearly hold. Finally, validity schema **c.** expresses that seeing to it with a chance of success of at least c implies seeing to it with a chance of success of at least k provided $c \leq k$.

4 Choice with an optimal chance of success

As explained in the introduction, we see an attempt for φ as the exertion of a choice that is maximal in the sense that it has the highest chance of achieving φ. So we aim to model attempt as a comparative notion. This means, that in our formal definition for the attempt operator $[\{ag\}\ \mathsf{xatt}]\varphi$ that we introduce here, we drop the absolute probabilities. The truth condition for the new operator $[\{ag\}\ \mathsf{xatt}]\varphi$ is as follows.

Definition 4.1 *Relative to a model $\mathcal{M} = \langle S, H, E, B, \pi \rangle$, truth $\langle s, h \rangle \models [\{ag\}\ \mathsf{xatt}]\varphi$ of a formula $[\{ag\}\ \mathsf{xatt}]\varphi$ in a dynamic state $\langle s, h \rangle$, with $s \in h$, is defined as:*

$\langle s, h \rangle \models [\{ag\}\ \mathsf{xatt}]\varphi \Leftrightarrow$
$\forall h'$: *if* $s \in h'$ *then* $CoS(s, h', ag, \varphi) \leq CoS(s, h, ag, \varphi)$
and
$\exists h''$: $s \in h''$ *and* $CoS(s, h'', ag, \varphi) < CoS(s, h, ag, \varphi)$

This truth condition explicitly defines the comparison of the current choice with other choices possible in that situation. In particular, if and only if the chance of obtaining φ for the current choice is higher than for the other choices possible in the given situation, the current choice is an attempt for φ. The 'side condition' says that there actually must be a choice alternative with a strictly lower chance of success.

Proposition 4.1 *Each instance of any of the following formula schemas is valid in the logic determined by the semantics of definition 4.1.*

(Cons)	$\neg[\{ag\}\ \mathit{xatt}]\bot$
(D)	$[\{ag\}\ \mathit{xatt}]\neg\varphi \rightarrow \neg[\{ag\}\ \mathit{xatt}]\varphi$
(Indep-Att)	$\Diamond[\{ag1\}\ \mathit{xatt}]\varphi \wedge \Diamond[\{ag2\}\ \mathit{xatt}]\psi \rightarrow$
	$\Diamond([\{ag1\}\ \mathit{xatt}]\varphi \wedge [\{ag2\}\ \mathit{xatt}]\psi)$
(Sure-Att)	$[\{ag\}\ \mathit{xstit}]\varphi \wedge \Diamond\neg[\{ag\}\ \mathit{xstit}]\varphi \rightarrow$
	$[\{ag\}\ \mathit{xatt}]\varphi$

The D-axiom says that the same choice cannot be at the same time an attempt for φ and $\neg\varphi$. This is due to the presence of the 'side condition' in definition 4.1. The side condition says that a choice can only be an attempt if there is at least one alternative choice with a strictly lower chance of success. Now we see immediately why de D-axiom holds: this can never be the case for complementary effects, since these have also complementary probabilities. In *stit* theory, side conditions are used to define 'deliberative' versions of *stit* operators [9]. And indeed the same intuition is at work here: a choice can only be an attempt if it is 'deliberate'.

The (Indep-Att) schema says that attempts of different agents are independent. Attempts are independent, because maximizing choice probabilities from the perspective of one agent is independent from maximizing choice probabilities from the perspective of some other agent.

Finally, the (Sure-Att) schema reveals the relation between the *stit* operator of our base language and the attempt operator. We already saw that we can associate the base operator $[\{ag\}\ \mathsf{xstit}]\varphi$ with a probabilistic *stit* operator with a chance of success of 1. Now, if such a choice qualifies as an attempt, it can only be that there is an alternative to the choice with a probability strictly lower than 1 (due to the side condition in definition 4.1). In the base language we can express this as the side condition $\Diamond\neg[\{ag\}\ \mathsf{xstit}]\varphi$ saying that φ is not ensured by ag's choice. This results in the property (Sure-Att) that says that if ag ensures φ with a chance of success of 1, and if ag could also have refrained (i.e., ag took a chance higher than 0 for $\neg\varphi$), then ag *attempts* φ. This again reveals the relation between the notion of attempt and the notion of 'deliberate choice' from the philosophical *stit* literature [9].

5 Conclusion and Discussion

This paper starts out by defining a base *stit* logic, which is a variant on Broersen's XSTIT. However, we define the semantics in terms of h-effectivity functions, which does more justice to the nature of the structures interpreting the language. We show completeness relative to this semantics. Then we proceed by generalizing the central *stit* operator of the base language to a probabilistic variant. The original operator comes out as the probabilistic operator assigning a chance 1 to success of a choice. In a second step we use the machinery used to define the probabilistic *stit* variant to define a notion of attempt. An attempt of agent ag is modeled as a 'kind' of maximal expected utility: agent ag attempts φ if and only if it performs a choice that is optimal in the sense that the sum of probabilities of the opponent choices ensuring φ would have been lower for any alternative choice by ag. So, an attempt for φ is a choice most likely leading to φ given ag's subjective probabilities about what other agents choose simultaneously.

There are several opportunities for future work. Among them are axiomatizations for the probabilistic *stit* and attempt operators. Since the probabilistic *stit* operator introduces probabilities explicitly in the object language, axiomatization is expected to be difficult. In case of the attempt operator, probabilities are implicit. An interesting observation is that this seems to resemble the reasoning of agents like ourselves that estimate optimality of their choices based on chances that are seldomly made explicit.

Finally, an interesting route for investigation is the generalization of the theory in this paper to group choices of agents. If a group makes a choice, we may assume all kinds of conditions on the pooling of

information within the group. This means that the chances that agents assign to choices made by agents within the group are generally different than the chances they assign to choices by agents outside the group. How this pooling of information takes form in a setting where beliefs are modeled as subjective probabilities is still an open question to us.

References

[1] R. Alur, T.A. Henzinger, and O. Kupferman. Alternating-time temporal logic. *Journal of the ACM*, 49(5):672–713, 2002.

[2] Philippe Balbiani, Andreas Herzig, and Nicolas Troquard. Alternative axiomatics and complexity of deliberative stit theories. *Journal of Philosophical Logic*, 2007.

[3] N. Belnap, M. Perloff, and M. Xu. *Facing the future: agents and choices in our indeterminist world*. Oxford University Press, 2001.

[4] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*, volume 53 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press, 2001.

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

[6] Jan Broersen, Andreas Herzig, and Nicolas Troquard. A normal simulation of coalition logic and an epistemic extension. In Dov Samet, editor, *Proceedings Theoretical Aspects Rationality and Knowledge (TARK XI), Brussels*, pages 92–101. ACM Digital Library, 2007.

[7] W. Conradie, V. Goranko, and D. Vakarelov. Algorithmic correspondence and completeness in modal logic I: The core algorithm SQEMA. *Logical Methods in Computer Science*, 2(1):1–26, 2006.

[8] E.A. Emerson. Temporal and modal logic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science, volume B: Formal Models and Semantics*, chapter 14, pages 996–1072. Elsevier Science, 1990.

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

[10] Wojciech Jamroga. A temporal logic for markov chains. In *AAMAS '08: Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems*, pages 697–704, 2008.

[11] Barteld P. Kooi and Allard M. Tamminga. Conflicting obligations in multi-agent deontic logic. In L. Goble and J.-J.Ch. Meyer, editors, *Proceedings 8th International Workshop on Deontic Logic in Computer Science (DEON'06)*, volume 4048 of *Lecture Notes in Computer Science*, pages 175–186. Springer, 2006.

[12] E. Lorini and A. Herzig. A logic of intention and attempt. *Synthese*, 163(1):45–77, 2008.

[13] Marc Pauly. A modal logic for coalitional power in games. *Journal of Logic and Computation*, 12(1):149–166, 2002.

[14] D. Vanderveken. Attempt, success and action generation: A logical study of intentional action. In D. Vanderveken, editor, *Logic, Thought and Action*, pages 316–342. Springer, 2005.