Lecture 1: STIT Theory and the Challenges of Non-deterministic Actions

ESSLLI’18: Multi-Agent Deontic Logic:
Reasoning About Normative Multi-Agent Systems
Gabriella Pigozzi & Leon van der Torre

1 Introduction

STIT logic [Horty2001] is a logic that allows to express non-deterministic actions.¹

The idea that an agent $\alpha$ sees to it that $A$ is taken to mean that the truth of the proposition $A$ is guaranteed by an action, or choice, of $\alpha$. [Horty2001, p.12]

Definition 1 (Branching time frames $\mathcal{F}$) A branching time frame is a structure $\mathcal{F}$ of the form $\langle \text{Tree}, < \rangle$, with Tree a nonempty set of moments, and $<$ a transitive, irreflexive, and treelike ordering on Tree.

Definition 2 (Branching time models $\mathcal{M}$) A branching time model is a structure $\mathcal{M}$ of the form $\langle \mathcal{F}, v \rangle$, with $\mathcal{F} = \langle \text{Tree}, < \rangle$ a branching time frame and $v$ an evaluation function mapping each sentence letter from the background language into a set $m/h$ pairs from Tree.

Definition 3 (Satisfaction relation) Let $m/h$ be an index and $v$ the evaluation function from a branching time model $\mathcal{M}$:

- $M, m/h \models A$ iff $m/h \in v(A)$, for $A$ an atomic formula
- $M, m/h \models A \land B$ iff $\mathcal{M}, m/h \models A$ and $\mathcal{M}, m/h \models B$
- $M, m/h \models \neg A$ iff $\mathcal{M}, m/h \not\models A$
- $M, m/h \models P A$ iff $\exists m' \in h$ s.t. $m' < m$ and $\mathcal{M}, m'/h \models A$ (P is the operator that represents past tense and can be read as “it was the case that”.)
- $M, m/h \models F A$ iff $\exists m' \in h$ s.t. $m < m'$ and $\mathcal{M}, m'/h \models A$ (F is the operator that represents simple future tense and can be read as “it will be the case that”.)

Definition 4 (General deontic stit frames/models) A general deontic stit frame is a structure of the form $\langle \text{Tree}, <, \text{Agent}, \text{Choice}, \text{Value} \rangle$, with Tree and $<$ as in branching time frames, Agent a nonempty set of agents, and Choice a function mapping each agent $\alpha$ and moment $m$ into a partition $H_m$ subject to the requirements of independence of agents and no choice between undivided histories. Finally, Value is a function associating each moment $m$ from Tree with a mapping $Value_m$ from $H_m$ into a set of values partially ordered by $\leq$. 
The stit operator used in Horty’s book is known as “Chellas stit” and represented as $\text{cstit}$.

**Definition 5 (Evaluation $\text{cstit}$)** $\mathcal{M}, m/h \models [\alpha \text{cstit} : A]$ iff $\text{Choice}_\alpha^m(h) \subseteq |A|_m^\mathcal{M}$.

That is, $[\alpha \text{cstit} : A]$ is true at a moment history pair $m, h$ if and only if $A$ is true at all moment-history pairs through $m$ that belong to the same action as $m, h$ (see Figure 2).

**Definition 6 (Evaluation $\Diamond A$)** $\mathcal{M}, m/h \models \Diamond A$ iff there is some history $h' \in H_m$ s.t.

1. $\mathcal{M}, m/h' \models A$
2. $\mathcal{M}, m/h'' \models A$ for each history $h'' \in H_m$ s.t. $\text{Value}_m(h') \leq \text{Value}_m(h'')$

That is, $\Diamond A$ is true at a moment history pair $m, h$ if and only if there is some history $h'$ through $m$ such that $A$ is true at all pairs $m, h''$ for which the history $h''$ has a utility at least as high as $h'$.

---

1Definitions and examples in this handouts are taken from [Horty2001].
On the STIT-model of Figure 3 we have:

- $\mathcal{M}, m/h_3 \models A$ directly from the valuation of atomic propositions on moment-history pairs
- $\mathcal{M}, m/h_3 \models F \neg A$ (the proposition $\neg A$ is true later on, at moment $n$, on the history $h_3$ through $m$)
- $\mathcal{M}, m/h_3 \models [\alpha \ cstit : A]$, because $A$ holds for all histories through $m$ that belong to the same action as $h_3$ (i.e. action $K_3^n$)
- $\mathcal{M}, m/h_3 \models \bigcirc A$ and $\mathcal{M}, m/h_3 \models \bigcirc [\alpha \ cstit : A]$. These two propositions are true for the same reason: the history $h_4$ through $m$ has the highest utility and satisfies both $A$ and $[\alpha \ cstit : A]$ at $m$.

![Figure 3: Example of utilitarian STIT model](image)

2 The gambling problem: ought-to-do vs. ought-to-be statements

What are the relations between what an agent ought to do and what ought to be the case? Are ought-to-do statements special kinds of ought-to-be statements? Can ‘agent $\alpha$ ought to see to it that $A$’ be modelled by the formula $\bigcirc [\alpha \ cstit : A]$ (‘it ought to be that agent $\alpha$ sees to it that $A$’)? Horty answers no and the example that justify this is the ‘gambling example’, where an agent faces the choice between gambling to double or lose five dollar (action $K_1$) and refraining from gambling (action $K_2$) (Figure 4).

3 Moral luck and the driving example

The gambling problem may be seen as a kind of moral luck: whether we obtain the utility of 10 or 0 is not due to our actions, but due to luck. So: How to deal with moral luck in normative reasoning?

“[T]wo drivers are travelling toward each other on a one-lane road, with no time to stop or communicate, and with a single moment at which each
must choose, independently, either to swerve or to continue along the road. There is only one direction in which the drivers might swerve, and so a collision can be avoided only if one of the drivers swerves and the other does not; if neither swerves, or both do, a collision occurs. This example is depicted in Figure 5, where $\alpha$ and $\beta$ represent the two drivers, $K_1$ and $K_2$ represent the actions available to $\alpha$ of swerving or staying on the road, $K_3$ and $K_4$ likewise represent the swerving or continuing actions available to $\beta$, and $m$ represents the moment at which $\alpha$ and $\beta$ must make their choice. The histories $h_1$ and $h_3$ are the ideal outcomes, resulting when one driver swerves and the other one does not; collision is avoided. The histories $h_2$ and $h_4$, resulting either when both drivers swerve or both continue along the road, represent non-ideal outcomes; collision occurs. The statement $A$, true at $h_1$ and $h_2$, expresses the proposition that $\alpha$ swerves.” [Horty2001, p.119]
4 Procrastination: actualism vs possibilism

Jackson and Pargetter [Jackson and Pargetter1986] gave the example of Professor Procrastinate:

“Professor Procrastinate receives an invitation to review a book. He is the best person to do the review, has the time, and so on. The best thing that can happen is that he says yes, and then writes the review when the book arrives. However, suppose it is further the case that were to say yes, he would not in fact get around to writing the review. Not because of incapacity or outside interference or anything like that, but because he would keep on putting the task off. (This has been known to happen.) This although the best thing that can happen is for Procrastinate to say yes and then write, and he can do exactly this, what would happen in fact were he to say yes is that he would not write the review. Moreover, we may suppose, this latter is the worst thing which may happen.

Actualism and possibilism are two views about the way in which an agent’s future choices might be thought to impact her current oughts. Roughly, actualism is the view that an agent’s current actions are to be evaluated against the background of the actions he is actually going to perform in the future. Possibilism is the view that an agent’s current actions are to be evaluated against the background of the actions that he might perform in the future, the available future actions.

References
