Norms, Intentions and Actions

Dominik Klein, Alessandra Marra TiLPS

Applied Logic Seminar TU Delft

May 26, 2014

Dominik Klein, Alessandra Marra: Norms, Intentions and Actions

External and internal approaches

Our approach

The Framework

Future Work

Dominik Klein, Alessandra Marra: Norms, Intentions and Actions

Norms, intentions and actions

Aim:

investigate the notion of agency and its logic

In particular:

 model the change that the acceptance of a norm triggers in an agent's intentions and actions

Current frameworks

Current frameworks in the logic of agency:

- external perspective: stit-logics (e.g., Horty and Belnap 1995, Horty 2001)
- internal perspective: intentions-based logics (e.g., Veltman 2012)

Stit (=seeing to it that) logics

- Indeterministic representation of the course of events (forward branching time structure)
- agent's actions constrain the course of events
- an agent sees to it that φ if she acts in such a way that φ is guaranteed

Stit frames

A Stit-frame $\mathcal{F} = \langle \mathit{Tree}, <, \mathit{Agent}, \mathit{Choice} \rangle$:

- Tree is a nonempty set of moments
- ▶ < is a tree-like ordering on *Tree* (i.e, $\forall m_1, m_2, m_3 \in Tree$, if $m_1 < m_3$ and $m_2 < m_3$, then either $m_1 = m_2$ or $m_1 < m_2$ or $m_2 < m_1$)
- Agent is a set of agents
- Choice is a function mapping each agent α and moment m into a partition Choice^m_α of the histories H_(m)

where:

► a history h is a maximal set of linearly ordered moments from Tree and H_(m) = {h | m ∈ h}

Deliberative stit

- A Stit-model $\mathcal{M} = \langle \mathcal{F}, \mathbf{v} \rangle$:
 - ▶ *F* is a stit-frame
 - $v : At \rightarrow \mathcal{P}(Tree \times H)$ the atomic valuation

Deliberative stit

A Stit-model $\mathcal{M} = \langle \mathcal{F}, \mathbf{v} \rangle$:

- *F* is a stit-frame
- $v : At \rightarrow \mathcal{P}(Tree \times H)$ the atomic valuation

Truth conditions for the deliberative stit:

- $\mathcal{M}, m/h \models [\alpha \ dstit]\phi$ iff:
 - $\mathcal{M}, m/h' \models \phi$ for each $h' \in Choice_{\alpha}^{m}(h)$ (positive requirement)
 - ► $\exists h'' \in H_{(m)}$ such that $\mathcal{M}, m/h'' \not\models \phi$ (negative requirement)

The stit picture

- Agent α's action picks out a partition cell: {{h₁, h₂}, {h₃}}
- at m/h₁ agent α guarantees that p
- [α dstit] ϕ true at m/h_2



Stit setting

- External representation of the course of events
- an agent's action constrains the course of events to lie within some subset of the possible histories still available
- [α dstit]φ: agent α's action guarantees that φ (where guaranteeing=necessitation of effects)

Stit setting

- External representation of the course of events
- an agent's action constrains the course of events to lie within some subset of the possible histories still available
- [α dstit]φ: agent α's action guarantees that φ (where guaranteeing=necessitation of effects)

Hence:

agency represented purely in terms of outcomes of actions

Limitations

- Stit-logics are static
- purely external perspective (Belnap's slogan: Leave the mind out!)

Limitations

- Stit-logics are static
- purely external perspective (Belnap's slogan: Leave the mind out!)

However:

- reasoning about what an agent does/can do is typically dynamic
- agent's intentions also count!

Especially when dealing with the case of the agent's acceptance of a norm, looking only at the outcomes of the agent's actions is **not** sufficient.

 Compare: Today (Friday) you eat fish for dinner because you have only fish left in the fridge vs. you eat fish for dinner because you are following a Christian norm

Veltman's intentions-based logic

- dynamic semantics: update operations over states
- states represent cognitive states of an agent
- In particular:
 - ► "!\$\phi\$, if accepted, induces a change of intentions in the agent's cognitive state"

Plans and realizations

A cognitive state S is a triple $\langle W, P, R \rangle$:

- if w ∈ W then, for all an agent in that cognitive state knows, w might be the actual world
- if $w \in W$ then P(w) is the plan the agent has developed for w
- w ∈ W and v ∈ R(w) then v is a possible successor of w.
 Every successor of w realizes one of the options of the plan for w

Plans and realizations, cont'd

In particular:

- A plan is a set of consistent to-do lists, none of which is a proper subset of another
- ► the acceptance of !\$\phi\$ triggers a change in the agent's plan (i.e., in the to-do lists)
- ► the agent is then committed to that plan: the set of possible successors is such that φ is realized

A simple example: update of the minimal state S with !p and then with !q



Dominik Klein, Alessandra Marra: Norms, Intentions and Actions

A simple example: update of the minimal state S with !p and then with !q



A simple example: update of the minimal state S with !p and then with !q



Limitations

- every accepted norm has to be realized immediately (one-step realizations)
- no distinction between stand and one-time norms

Limitations

- every accepted norm has to be realized immediately (one-step realizations)
- no distinction between stand and one-time norms

However:

- flexibility in planning is more realistic
- stand norms like Don't kill! are different from one-time norms Bring the trash out by tomorrow morning!

A logic for norms, intentions and action

We combine external and internal perspective on agency, in order to model the changes that the acceptance of a norm triggers

A logic for norms, intentions and action

We combine external and internal perspective on agency, in order to model the changes that the acceptance of a norm triggers

 cf. Dignum and al. 1996; Broersen 2001 where agent's intentions are simply accessibility relations – the relationship between agent's intentions and actions remains unexplained

A logic for norms, intentions and action

We combine external and internal perspective on agency, in order to model the changes that the acceptance of a norm triggers

Main characteristics:

- agent' actions represented in a tree-like structure (vectors are actions, nodes are states of affairs)
- agent's intentions are represented in to-do lists attached to nodes
- ▶ !φ if accepted- triggers a change in the agent's intentions and, consequently, in her actions
- we look at the nodes (for norms: Meinong/Chisholm reduction)

The Setting



The Setting



- No names for actions, all evaluations at nodes
- only one agent
- no epistemic uncertainty
- no uncertain success
- abstract representation of the world
- Conditions to be relaxed later

Formal Definition



For the rest of the talk fix a set of atoms At. An *action tree* is a 4-tuple $T = \langle W, w_0, \prec, V \rangle$ where

- W is a set of worlds with $w_0 \in W$
- *Val* : At $\rightarrow \mathcal{P}(W)$ the atomic valuation
- ▶ \prec is a tree order on W with root w_0

















Crucial Distinction:



Crucial Distinction:

1 Norm(, command or goal) = linguistic item


Crucial Distinction:

- 1 Norm(, command or goal) = linguistic item
- 2 Obligation = Semantic item, parts of the tree to land up in

Obligations



Definition

i) An *obligation* in an action tree T is a subtree S of T

 Obligations correspond to consistent, extendable actions the agent can perform

Dominik Klein, Alessandra Marra: Norms, Intentions and Actions

Obligations



Definition

i) An *obligation* in an action tree T is a subtree S of T such that for every $v \in S$ that has a successor in T it also has a successor in S

 Obligations correspond to consistent, extendable actions the agent can perform

Dominik Klein, Alessandra Marra: Norms, Intentions and Actions

Obligations



Definition

- i) An *obligation* in an action tree T is a subtree S of T such that for every $v \in S$ that has a successor in T it also has a successor in S
- ii) For any subtree S of T let S be the tree derived from S by iteratedly removing dead ends. (Technically: S = ∪{h ⊆ S|h a maximal history of T}
 - Obligations correspond to consistent, extendable actions the agent can perform

Dominik Klein, Alessandra Marra: Norms, Intentions and Actions

Obligations, cont'd

 Obligations correspond to consistent, extendable actions the agent can perform

Obligations, cont'd

- Obligations correspond to consistent, extendable actions the agent can perform
- But where do obligations come from? Norms

How to talk about Norms (and Commands)?

- one time norms Bring the trash out!
- standing norms Don't kill!

How to talk about Norms (and Commands)?

- one time norms Bring the trash out!
- standing norms Don't kill!

Let $At = \{p_1 \dots p_n\}$ be a set of atomic propositions (*I kill, I bring the trash out, I drink a beer...*) The obligational language is given by

$$\varphi := \boldsymbol{p}^{\exists} | \boldsymbol{p}^{\forall} | \neg \boldsymbol{p}^{\exists} | \neg \boldsymbol{p}^{\forall} | \varphi \land \varphi | \varphi \lor \varphi$$

Norm φ gives rise to an obligation O_{φ}



Norm φ gives rise to an obligation O_{φ}



Inductive definition of O_{φ}

 $O_{p^{\exists}} = \bigcup \{h | h \text{ history of } T \text{ and some world in } h \text{ satisfies p} \}$

Norm φ gives rise to an obligation O_{φ}



Inductive definition of O_{φ}

 $O_{p^{\forall}} = \bigcup \{h | h \text{ history of } T \text{ and every world in } h \text{ satisfies p} \}$

Norm φ gives rise to an obligation O_{φ}



Inductive definition of O_{φ}

 $O_{\neg p^{\exists}}$ similar to $O_{p^{\forall}}$

Norm φ gives rise to an obligation O_{φ}



Inductive definition of O_{φ}

 $O_{\neg p^{\forall}}$ similar to $O_{p^{\exists}}$

Norm φ gives rise to an obligation O_{φ}



$$O_{\varphi ee \psi} = O_{\varphi} \cup O_{\psi}$$

Norm φ gives rise to an obligation O_{φ}



$$O_{arphiee\psi} = O_arphi \cup O_\psi$$

Norm φ gives rise to an obligation O_{φ}



$$O_{arphiee\psi} = O_arphi \cup O_\psi$$

Norm φ gives rise to an obligation O_{φ}



$$O_{\varphi \wedge \psi} = \overline{O_{\varphi} \cap O_{\psi}}$$

Norm φ gives rise to an obligation O_{φ}



$$O_{arphi\wedge\psi}=\overline{O_arphi\cap O_\psi}$$

Norm φ gives rise to an obligation O_{φ}



$$O_{arphi\wedge\psi}=\overline{O_arphi\cap O_\psi}$$

Norm φ gives rise to an obligation O_{φ}



$$O_{arphi\wedge\psi}=\overline{O_arphi\cap O_\psi}$$

Planning to Fulfill a Norm

- Norms (Goals, Commands) = Abstract Formulas that could or could not be satisfied
- Accepting a Norm = Internalizing it on a To-Do List

Planning to Fulfill a Norm

- Norms (Goals, Commands) = Abstract Formulas that could or could not be satisfied
- Accepting a Norm = Internalizing it on a To-Do List



Obligations and Intentions

- Once accepted, an agent intends to fulfill his norms/goals/commands
- ► Mental representation of accepted commitments: To-Do Lists To-Do List of the form (p, true, ∃)(q, false, ∀)...

Obligations and Intentions

- Once accepted, an agent intends to fulfill his norms/goals/commands
- ► Mental representation of accepted commitments: To-Do Lists To-Do List of the form (p, true, ∃)(q, false, ∀)...

true	false
p∃	$q^{orall}$

Obligations and Intentions

- Once accepted, an agent intends to fulfill his norms/goals/commands
- ► Mental representation of accepted commitments: To-Do Lists To-Do List of the form (p, true, ∃)(q, false, ∀)...

true	false
p∃	$q^{orall}$

 Problem of Free Choice: Bring the trash out or Wash the Dishes

false

Several To-Do-Lists, one per choice, give rise to a Plan



Definition

i) A to-do list is a Set D $D \subseteq At \times \{true, false\} \times \{\forall, \exists\}$

Definition

- i) A to-do list is a Set D $D \subseteq At \times \{true, false\} \times \{\forall, \exists\}$
- ii) A to-do list is *consistent* if it doesn't contain contradicting commitments,

Definition

i) A to-do list is a Set D $D \subseteq At \times \{true, false\} \times \{\forall, \exists\}$

ii) A to-do list is *consistent* if it doesn't contain contradicting commitments, that is pairs of commitments

 $\langle p, \mathit{true}, \forall \rangle$ and $\langle p, \mathit{false}, \forall \rangle$

Definition

i) A to-do list is a Set D $D \subseteq At \times \{true, false\} \times \{\forall, \exists\}$

ii) A to-do list is *consistent* if it doesn't contain contradicting commitments, that is pairs of commitments

 $\begin{array}{l} \langle \textit{p},\textit{true},\forall\rangle \text{ and } \langle \textit{p},\textit{false},\forall\rangle \\ \langle \textit{p},\textit{true},\forall\rangle \text{ and } \langle \textit{p},\textit{false},\exists\rangle... \end{array}$

Definition

i) A to-do list is a Set D $D \subseteq At \times \{true, false\} \times \{\forall, \exists\}$

ii) A to-do list is *consistent* if it doesn't contain contradicting commitments, that is pairs of commitments

 $\langle p, true, \forall \rangle$ and $\langle p, false, \forall \rangle$ $\langle p, true, \forall \rangle$ and $\langle p, false, \exists \rangle$...

iii) A plan P is a set of to-do lists such that $D \subsetneq D'$ for all $D, D' \in P$

Attentie Difference between plans

- ► {∅} Empty To-Do List = Do whatever you want
- \emptyset Empty Plan = State of Violation

►
$$\varphi$$
 of the form p^{\exists} resp. p^{\forall}
 $P \uparrow \varphi = \langle p, true, \exists \rangle$ resp. $\langle p, true, \forall \rangle$

Finally – A planning tree

Definition

An NIA tree is a 6-tuple $T = \langle W, w_0, \prec, V, \mathcal{O}, P \rangle$ where

- $T = \langle W, w_0, \prec, V \rangle$ is an action tree
- O is a set of Obligations in T
- $P: W \rightarrow \{Plans\}$ attaches a plan to each node

Finally – A planning tree

Definition

An NIA tree is a 6-tuple $T = \langle W, w_0, \prec, V, \mathcal{O}, P \rangle$ where

- $T = \langle W, w_0, \prec, V \rangle$ is an action tree
- O is a set of Obligations in T
- $P: W \rightarrow \{Plans\}$ attaches a plan to each node
- Plans should be executable
- To-Do lists track how plans are gradually fulfilled
- Once satisfied, remove one-time norms from to-do-lists
- All one-time norms should be satisfied at the end of time
- Standing norms should never be violated

Definition An NIA -tree T is coherent iff



Definition An NIA -tree T is coherent iff



Definition

An NIA -tree T is coherent iff

i) Success: If w is a leaf of T then no $D \in P(w)$ contains one-time norms

Definition

An NIA -tree T is coherent iff

i) Success: If w is a leaf of T then no $D \in P(w)$ contains one-time norms

ii) Gradual Fulfilment If $v \prec w \in T$ then

$$\begin{split} \mathsf{P}(\mathsf{w}) &\subseteq \min(\{D - \{\langle \mathsf{p}, \mathsf{true}, \exists \rangle | \mathsf{v} \in \mathsf{Val}(\mathsf{p})\} \\ &- \{\langle \mathsf{p}, \mathsf{false}, \exists \rangle | \mathsf{v} \notin \mathsf{Val}(\mathsf{p})\} | D \in \mathsf{P}(\mathsf{v})\}) \end{split}$$

"Fulfilled existential obligations get removed"

Definition cont'd

iii) Non violation If $v \prec w \in T$ then

 $P(w) \subseteq \{D \in P(v) | \langle p, false, \forall \rangle \notin D \text{ if } w \in Val(p), \\ \langle p, true, \forall \rangle \notin D \text{ if } w \notin Val(p) \}$

"The present state doesn't violate standings norms"

iv) Efficiency If $v \prec w$ then for every $D \in \mathcal{P}(v)$ there is some $D' \in \mathcal{P}(w)$ such that D' is obtained from D by removing one-time norms that are satisfied.

"All To-Do lists refer to some future branch"

Remark: Given $P(w_0)$, there is an algorithm to make the function P coherent (Upward-Downward Procedure)

Learning and accepting norms: intentions and actions

Let ${\cal T}$ be an NIA -tree. Accepting the norm $!\varphi$ updates ${\cal T}$ to ${\cal T}'$ with

- No change in the underlying action tree
- $\blacktriangleright \mathcal{O}' = \mathcal{O} \cup \mathcal{O}_{\varphi}$
- ► To obtain P' update P(w₀) with φ and run the down-up-down algorithm to make P' consistent

We say that ...

- ► ... the obligation that φ was supported if $\mathcal{O}' = \mathcal{O}$ Write: $T \models O \varphi$
- ► ... the agent already incorporated φ if P' = P Write: T ⊨!φ









A Theorem

Theorem

Let T be an NIA -tree obtained by starting with no obligations and empty to-do lists through iterated updating with norms $!\varphi$. Then:

$$\bigcap_{O\in\mathcal{O}}O=\{w\in W|P(w)\neq\emptyset\}$$

An agent following his to-do list will end up with the maximal set consistent with all Obligations around.

Extensions and Future Developments

- Introduce Payoffs for the Agents (Intrinsic Interest vs. Outside Conditions)
- Attach Plans to non-optimal worlds

Contrary to Duty: *Go home over Easter. If you don't go home at least go to church* Conflicting Obligations

- Introduce epistemic uncertainty (possibly with belief order)
 - Move by Nature
 - Uncertain Success of Actions
 - Distinguish between not accepting φ and failing to meet
 O_φ

Contrary to Duty

Go home over Easter. If you don't go home at least go to church



Contrary to Duty

Go home over easter. If you don't go home at least go to church



Contrary to Duty

Go home over easter. If you don't go home at least go to church



Conclusions

- Agency as complex notion which involves agent's actions and intentions
- Formal framework which takes into account both external and internal dimensions
- Application to the process of acquisition of norms
- Dynamic approach which distinguishes between different kinds of norms (stand vs. one-time) and allows for flexibility in the process of fulfilling norms
- Future directions: epistemic uncertainty, multi-agent setting, etc.

References

- Broersen et al., Resolving Conflicts Between Beliefs, Obligations, Intentions, and Desires. Symbolic and Quantitative Approaches to Reasoning and Uncertainty, Springer, 2001.
- Dignum et al., A Modal Approach to Intentions, Commitments and Obligations. In Deontic logic, agency and normative systems, Springer, 1996.
- Horty J.F., Belnap N., The Deliberative Stit: A Study of Action, Omission, Ability, and Obligation. Journal of Philosophical Logic, Vol.24, No.6, 1995.
- Horty J.F., Agency and Deontic Logic. Oxford University Press, 2001.
- Veltman F., Or else, what? Imperatives on the borderline of semantics and pragmatics. Institute for Logic, Language and Computation, Lego Seminar, Amsterdam, 2012.