Putting Norms in Context

René Mellema

rene.mellema@cs.umu.se
@remellema

UMEÅ UNIVERSITY

What are norms?

Much debate over definition

- Rule like structures
- Make behaviour:
  - Obligatory
  - Forbidden
  - Permissible
- Can be broken

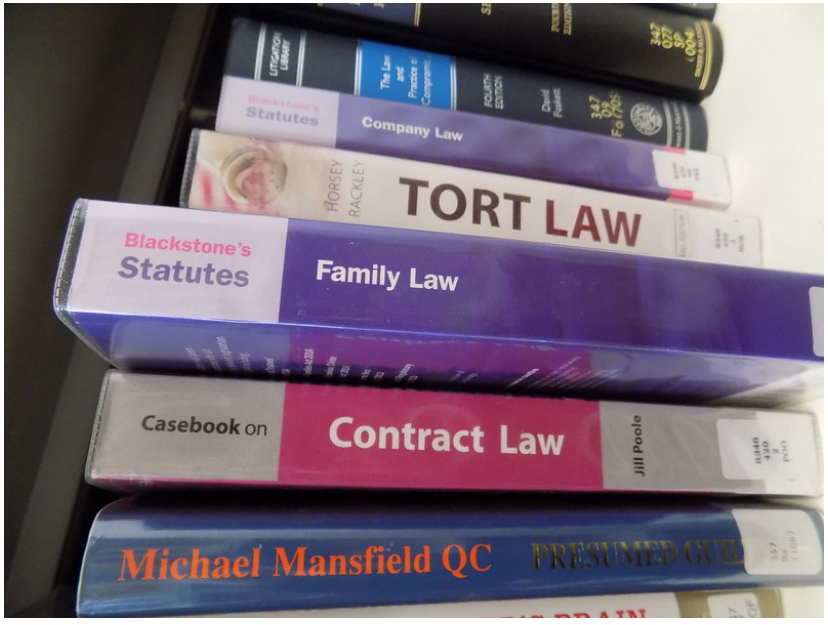


Image from: http://www.dpp-law.com/

Examples of norms

- Legal/Formal norms
  - Laws of a country
  - Rules of a university
- Moral norms
  - Codes of ethics
  - Rules for good living
- Social norms
  - Rules of etiquette
  - Rules within friend groups

Why norms?

- Control over heterogeneous agents
  - More flexible than constraints
- Norms have a motivational component
  - Social standing
  - Promotes values
  - Identity
- Clearer for communication with stakeholders
  - Social simulation
  - Human computer interaction

What is in a norm?

- Violation condition
- Activation condition
- Deactivation condition
- Deadline
- Governed agents
- Repair
- Punishment

Image from: www.Pixel.la on Flickr

Current approaches

- Implementations either
  - Only norm focussed (Metanorms)
  - Norms sort of hidden away (ASSOCC)
- Formalizations based on simple examples

Formalizations

- Normally given in deontic logic
  - Logic of obligation and permission
  - $O(\varphi)$: It is obligated that/to do φ
  - $F(\varphi)$: It is forbidden that/to do φ
  - $P(\varphi)$: It is permitted that/to do φ
- Worlds/states are labelled as bad if there is a norm violation there
  - Worlds labelled in the logic allows reduction into Temporal/Dynamic logic

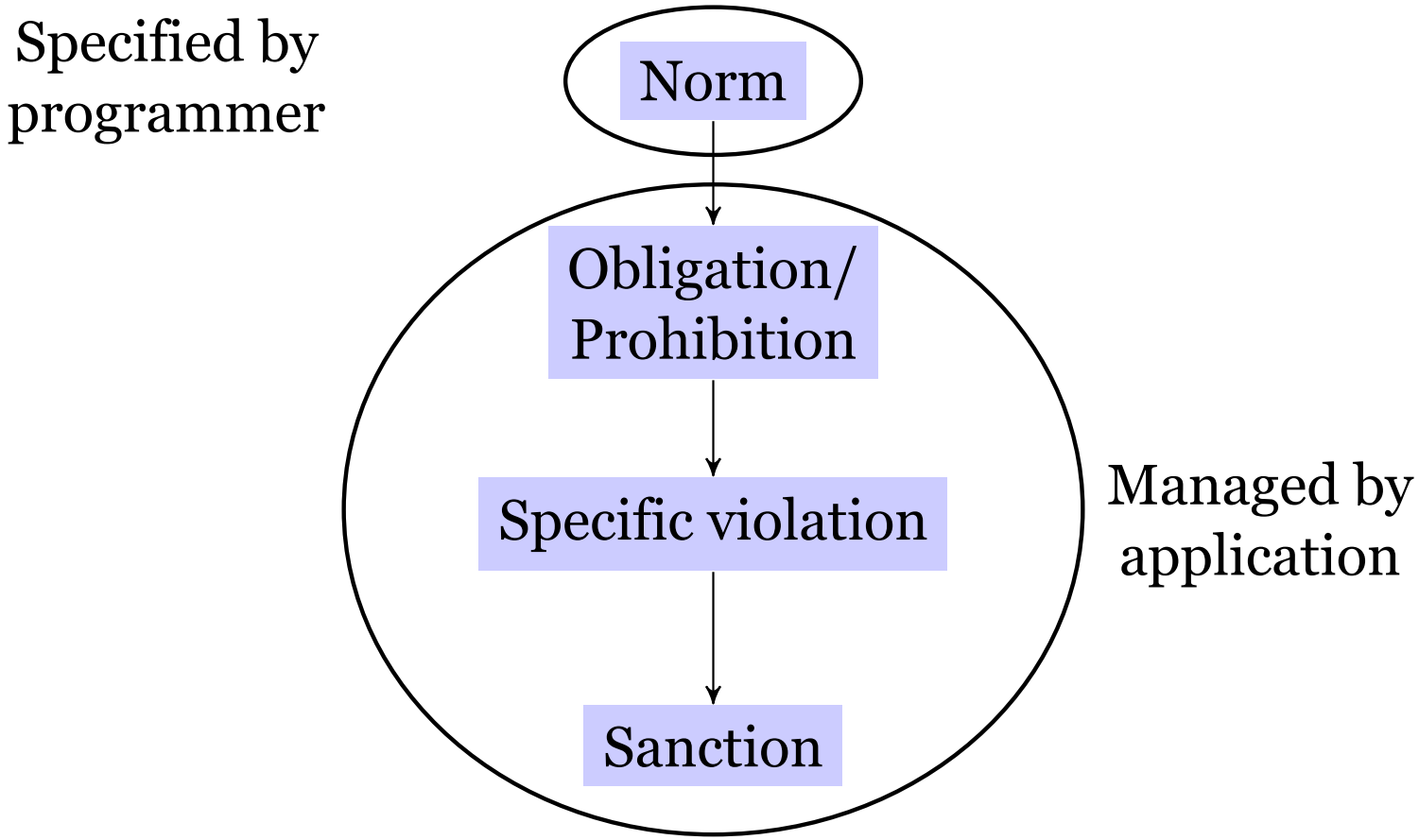
Problems

- World is either in violation or it is not
  - Only one violation of all norms at a time!
- All information about violation is lost
  - No link to punishment

Wishlist

- Specification of norms
- Violations are linked to consequences
  - Violations are unique
- Temporal reasoning

Solution


Specified by programmer
Norm
Obligation/ Prohibition
Specific violation
Sanction
Managed by application

Notation

$$
\begin{aligned}
E_a\varphi &:= a \text{ sees to it that } \varphi \\
X\varphi &:= \text{at that next step, } \varphi \text{ holds} \\
\varphi U\psi &:= \varphi \text{ holds at least until } \psi \\
t.\varphi(t) &:= \varphi \text{ holds when all occurences of } t \text{ are replaced by now} \\
V_{i,a,t} : \varphi &:= \text{agent } a \text{ has violated norm } i \text{ at time } t \text{ by doing } \varphi
\end{aligned}
$$

Current formalization

- Prohibition:

$$F_{i,a}(\varphi) \iff E_a\varphi \rightarrow X(t.V_{i,a,t} : \varphi)$$

- Obligation with deadline:

$$\begin{aligned} M, s \models O_{i,a}(\varphi < \delta) \iff & \text{ for all } \sigma \text{ with } \sigma_0 = s, \\ & \text{there exists } j > 0, M, \sigma_j \models \delta \text{ and for all } 0 \leq k < j : \\ & M, \sigma_k \models t.\neg V_{i,a,t} : \varphi \wedge \neg\delta \text{ and} \\ & (\text{there exists } 0 \leq k < j : \\ & \qquad M, \sigma_k \models E_a\varphi \rightarrow (t.\neg V_{i,a,t} : \varphi) U \delta \\ & \text{or (for all } 0 \leq k < j : \\ & \qquad M, \sigma_k \models \neg E_a\varphi \text{ and } M, \sigma_j \models t.V_{i,a,t} : \varphi)) \end{aligned}$$

When is it a norm?

$$\begin{aligned}\text{NORM}(\quad & label : i, \quad activation : \alpha, \quad deactivation : \beta, \\ & deadline : \delta, \quad condition : \varphi, \\ & repair : \rho, \quad punishment : \pi)\end{aligned}$$

$$M, s \models \text{NORM}(i, \alpha, \beta, \delta, \varphi, \rho, \pi) \iff \text{for all } a \in \mathcal{A},$$
$$M, s \models \alpha \rightarrow t_s.([O_{i,a}(\varphi < \delta)U\beta] \wedge$$
$$[t_v.V_{i,a,t_v} : \varphi \rightarrow (O_a(\rho) \wedge O_a(\pi))]U\beta)$$

Implementation

- Only need to know what an agent does/changes to determine norm compliance
- Violations can be stored in i.e. a database
- Activation can be regulated by checking for activation/deactivation
- Consequences of norm breaking put in terms agent understands

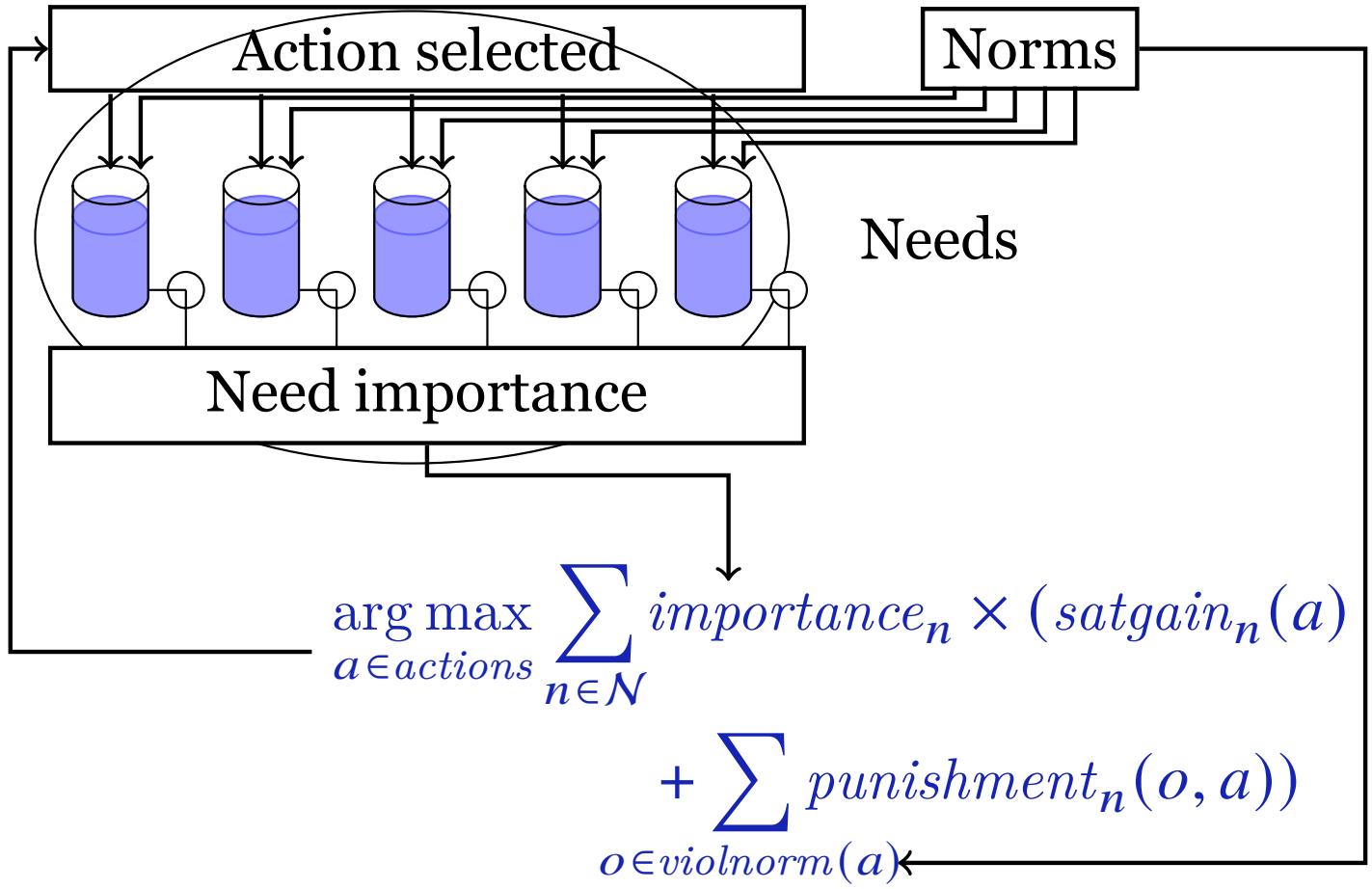
Consequences for ASSOCC

- Apply ideas to an existing implementation, ASSOCC
  - Agent-based Social Simulation of the Coronavirus Crisis
- Norms currently implemented implicitly
- Examples of norms in simulation:
  - Stay at home orders
  - Social distancing

Consequences for ASSOCC

- Implementation already tracks activation/deactivation
- Only prohibitions, so no deadlines
- Repair and condition are each others negation
- Punishment is internal and direct

Consequences for ASSOCC



Conclusion

- Norms can be a powerful tool for dealing with autonomous agents
- Current approaches are limited
- Need to keep track of violations

Next steps

- Studying the formalization more
- Build implementation in a social simulation
- Apply to norm learning/emergence/internalization