

Sociotechnical Systems and Social Norms

Artificial Intelligence: Multiagent Systems – II

Nirav Ajmeri

School of Computer Science
University of Bristol

TB1 2023/24

Outline

- 1 Review
- 2 Sociotechnical Systems
- 3 Social Norms

Agents and Multiagent Systems

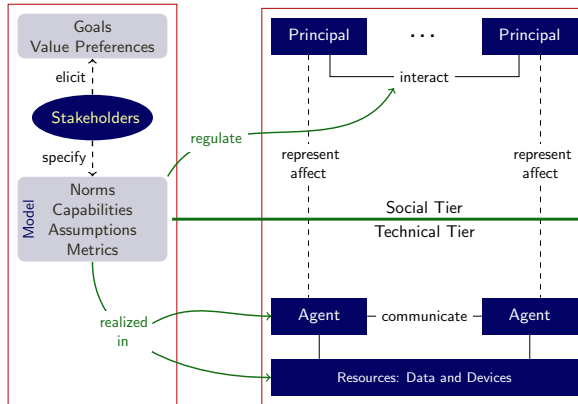
- Definitions
- Properties of an intelligent agent
- Properties or attributes of the environment

Outline

- 1 Review
- 2 Sociotechnical Systems
- 3 Social Norms

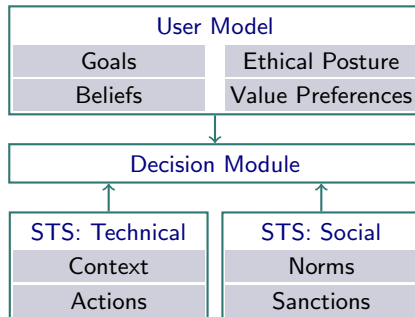
Sociotechnical System (STS)

- An STS includes social entities—*principals* (humans and organizations) and other *stakeholders*—and technical entities—*agents* and *resources*



STS: A Decentralized Multiagent System

- An STS is not a separate running entity but is realized through the interactions of principals, agents, and resources featuring in that STS



Outline

- 1 Review
- 2 Sociotechnical Systems
- 3 Social Norms**

Norms Govern the Interactions between Principals

Computational entities with well-defined semantics

- Norm: A *directed* social expectation between principals (stakeholders)
 - Types: Commitment, prohibition, authorization, power, ...
 - Standards of correctness
 - *Prima facie*, satisfaction is ethically desirable and violation undesirable
- Accountability: A principal can call another to account for its actions
 - Derives from norms
 - Provides an opportunity for principals to explain their actions
 - Leading to *prima facie* judgments being reconsidered

Social Norms (Singh 2013)

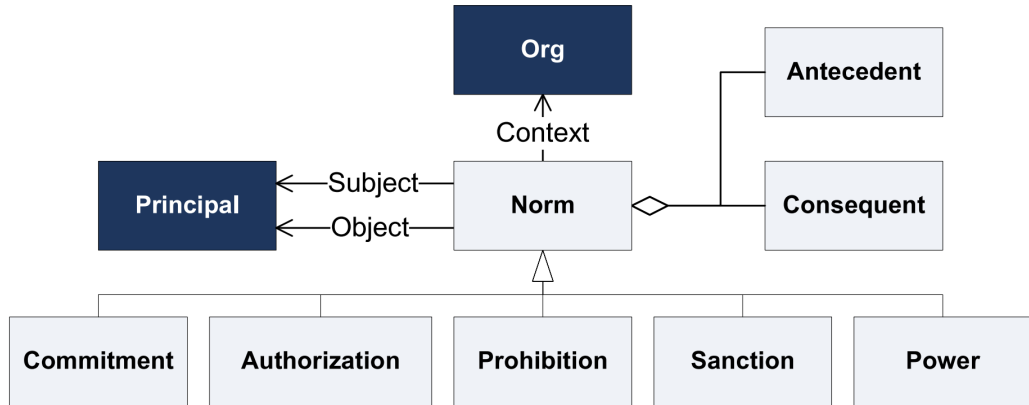
Norms govern the interactions between principals

Formally, a norm is a tuple $\langle n, \text{SBJ}, \text{OBJ}, \text{ant}, \text{con} \rangle$, where

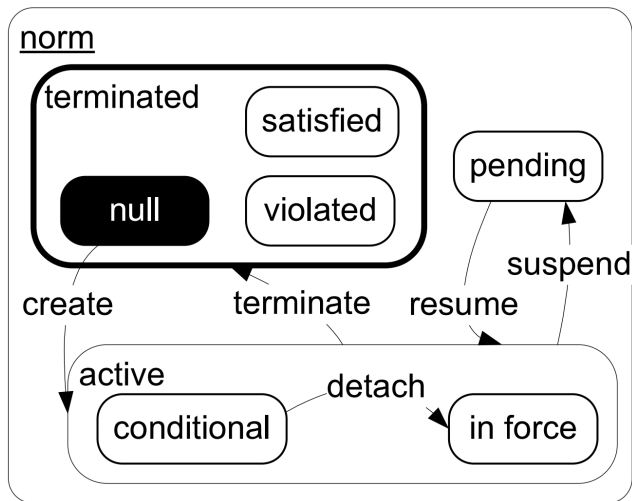
- n , its type, is one of $\{\text{commitment}, \text{prohibition}, \text{authorization}, \text{power}, \text{sanction}\}$;
- $\text{SBJ} \in \mathbb{R}$ is its subject;
- $\text{OBJ} \in \mathbb{R}$ is its object;
- $\text{ant} \in \text{Expr}$ is its antecedent; and
- $\text{con} \in \text{Expr}$ is its consequent.

We write a norm as $n(\text{SBJ}, \text{OBJ}, \text{ant}, \text{con})$

Logical Form of a Norm



Norm Lifecycle



Types of Social Norms

Commitment, prohibition, and authorization

- Commitment: The subject is committed to the object that when the antecedent holds, subject will bring about the consequent
A meeting attendee is committed to other attendees that he or she will keep his or her phone on silent during the meeting
- Prohibition: The subject is prohibited by the object that when the antecedent holds, subject will not bring about the consequent
A library visitor is prohibited by the library to answer any phone calls when the visitor is in the silent reading area of the library
- Authorization:
The subject is authorized by the object that when the antecedent holds, subject can bring about the consequent
A library staff member is authorized by the library to make any personal phone calls during lunch hours

Types of Social Norms

Power and sanction

- Power: The object empowers the subject to bring about the consequent by bringing about the antecedent. A power is the ability to alter the norms between two or more principals
A library staff member is empowered to admit new people into the library by creating their accounts
- Sanction: The object would sanction the subject by bringing about the consequent provided the antecedent holds

Requirements of a Healthcare STS

Healthcare emergency scenario

- R-Publish: Patient's personally identifying information (PHI) should not be published online under any circumstances
- R-External: Except in emergencies, hospital physicians should not share a patient's PHI with outside physicians
- R-Family: In emergencies, hospital physicians may share patient's PHI with family members to inform family members or gather new information to help with treatment

Requirements of a Healthcare STS

Exercise: Think about the norm types for each requirement in the healthcare emergency scenario

Normative Specification of an STS

Healthcare emergency scenario

Exercise: Refine this initial specification to satisfy the requirements

- R-Publish: Patient's personally identifying information (PHI) should not be published online under any circumstances
`prohibition(physician, hospital, true, publish_PHI_online)`
- R-External: Except in emergencies, hospital physicians should not share a patient's PHI with outside physicians
`prohibition(physician, hospital, true, share_PHI_outside_phy)`
- R-Family: In emergencies, hospital physicians may share patient's PHI with family members to inform family members or gather new information to help with treatment
`authorization(physician, hospital, true, share_PHI_family)`

Refining a Specification of an STS

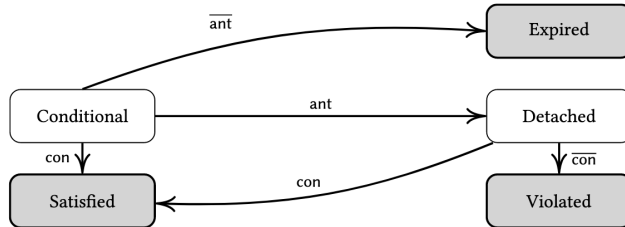
Healthcare emergency scenario

- R-Publish: Don't publish PHI
~~prohibition(physician, hospital, true, publish_PHI_online)~~
 • Further refinement: Include a mechanism to not allow publishing PHI online
- R-External: Except in emergencies, don't share PHI with outside physicians
~~prohibition(physician, hospital, true, share_PHI_outside_phy)~~
~~prohibition(physician, hospital, **emergency**, share_PHI_outside_phy)~~
- R-Family: Share PHI in emergencies
~~authorization(physician, hospital, true, share_PHI_family)~~
~~authorization(physician, hospital, **emergency**, share_PHI_family)~~
 • Further refinement: Include a commitment from physician to family
~~commitment(physician, family, emergency, share_PHI_family)~~

Lifecycle of a Commitment

A physician is practically committed to the hospital to operating upon patients in an emergency

`commitment(PHYSICIAN, HOSPITAL, emergency, operate)`

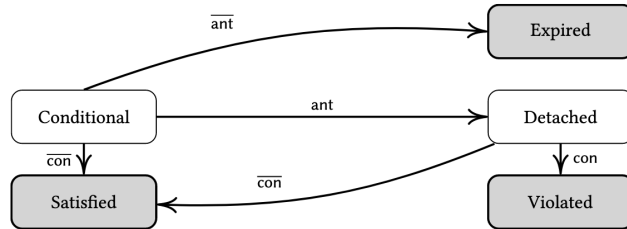


Accountable party is the subject of the norm. In the example commitment, if the physician fails to operate upon patients, the commitment is violated

Lifecycle of a Prohibition

Physician is prohibited by the hospital from disclosing a patient's PHI to others
(share_PHI_thirdparty)

$p(\text{PHYSICIAN}, \text{HOSPITAL}, \text{true}, \text{share_PHI_thirdparty})$

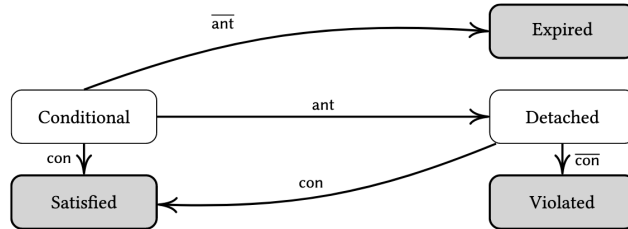


Accountable party is the subject of the norm. The prohibition in the example is unconditional because its antecedent is true. If the patient's PHI is disclosed, the prohibition is violated

Lifecycle of a Authorization

Physician is authorized to access EHR as well as operate upon a patient when the patient's consent is obtained

$a(\text{PHYSICIAN}, \text{HOSPITAL}, \text{consent}, \text{EHR} \vee \text{OPERATE})$



Accountable party is the object of the norm. If the physician cannot access the patient's EHR or operate upon the patient when the authorization is detached, then the authorization is violated

Normative Conflict

Exercise: Think about cases when norms conflict

Norm Emergence

Exercise: Think how norms could emerge or come into force