Monitoring and enforcement as a second-order guidance problem

10 December 2020. JURIX 2020 @ Brno/Prague (virtual)

Giovanni Sileno[a] (g.sileno@uva.nl)
Alexander Boer[b], Tom van Engers[c]

[a] Informatics Institute, University of Amsterdam, the Netherlands
[b] KPMG, Amsterdam, the Netherlands
[c] Leibniz Institute, TNO/University of Amsterdam, the Netherlands

Research context: Digital Market-Places (DMPs) infrastructures

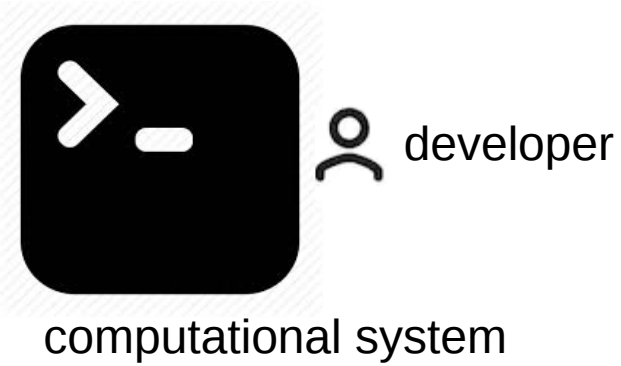
legal norms	*rules of "society"*	
DMP policy	*rules of the "game"*	***these are about what ought to be (but may be violated)***
agreements, contracts	*ad-hoc rules set amongst "players"*	
transactions	*"rules" of the infrastructure*	***these are about what may be (possibility)***

Data-sharing infrastructures as DMPs exhibit the double status of computational and socio-economic systems

The developer’s view: Control

- **Commander**
- Instructions → Operators

⇒ Controlled environment (internal)



The user’s view: Guidance

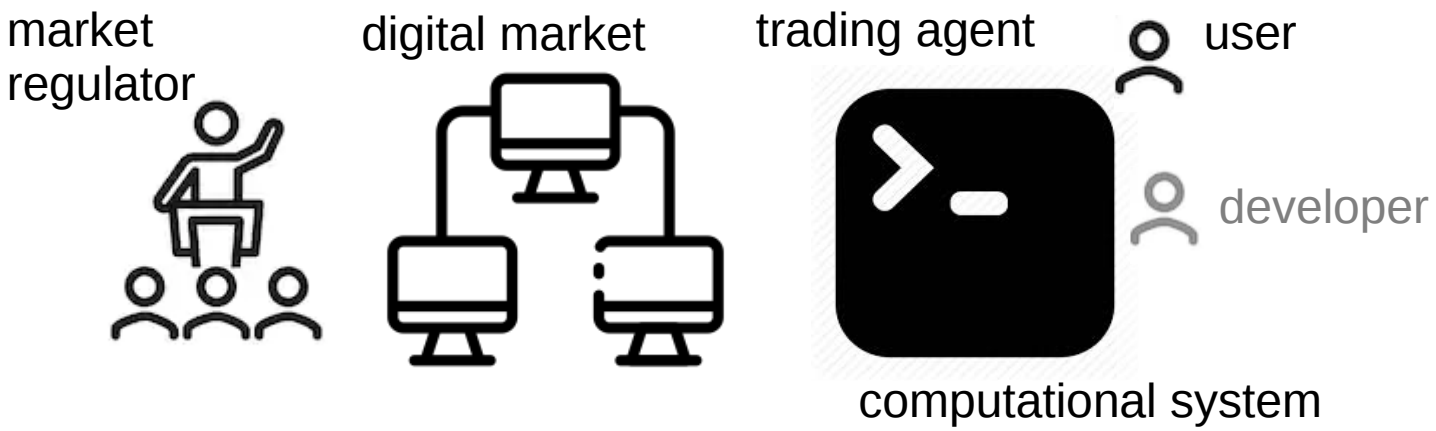
- **Decision-maker**
- Directives →
  - Commander
  - Instructions → Operators

➡ Partially-controlled environment (external, micro-level)

The “maintainer”’s view: Second-order guidance

- **Policy-maker**
- Policies → Decision-maker
  - Directives → Commander
    - Instructions → Operators

→ Partially-controlled environment (external, macro-level)



The “maintainer”’s view: Second-order guidance

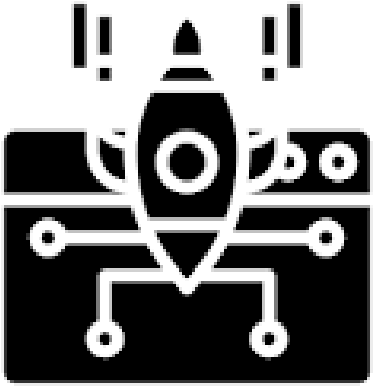
- **Policy-maker**
- Policies → Decision-maker
  - Directives → Commander
  - Instructions → Operators

→ Partially-controlled environment (external, macro-level)

Second-order guidance depends on **adoption**.
Enforcement measures are (some of) the means by which the policy-maker can influence adoption.

Example of
"second-order" guidance **problem**

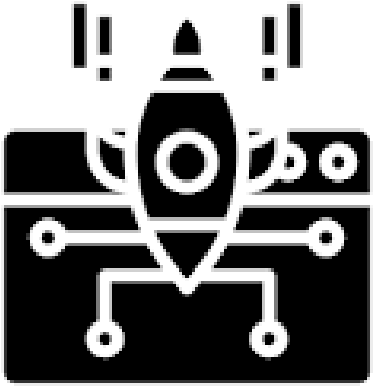
Cyber-attack scenario



- If you suffer of a cyber-attack,
  share the information with the consortium
- If you are notified of cyber-attack,
  start defensive maneuvers

*Inspired by the SARNET project.

Cyber-attack scenario



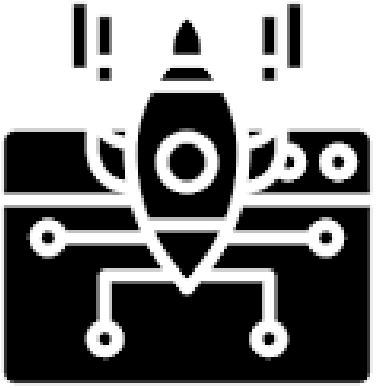
- If you suffer of a cyber-attack,
  share the information with the consortium
- If you are notified of cyber-attack,
  start defensive maneuvers

Sharing may be detrimental if the released data has competitive value

Defensive maneuvers may carry costs for the service provider

*Inspired by the SARNET project.

Cyber-attack scenario



- If you suffer of a cyber-attack,
  share the information with the consortium
- If you are notified of cyber-attack,
  start defensive maneuvers

Sharing may be detrimental if the released data has competitive value

Defensive maneuvers may carry costs for the service provider

What enforcement measures to apply?

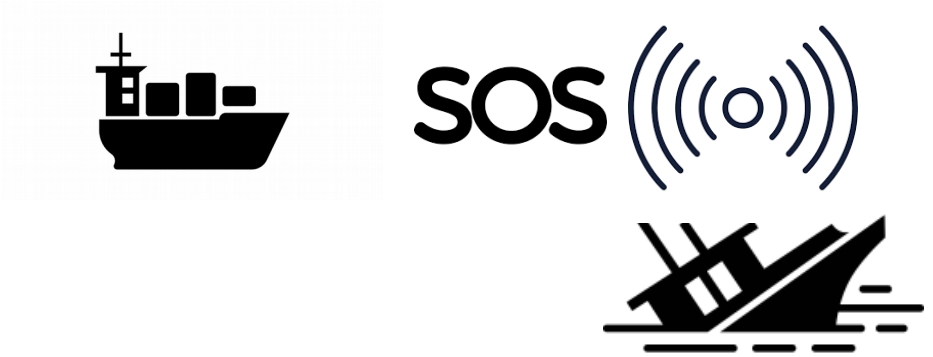
*Inspired by the SARNET project.

Types of enforcements

Function of norms

- One of the functions of norms is to express **relative preferences** that should guide behaviour

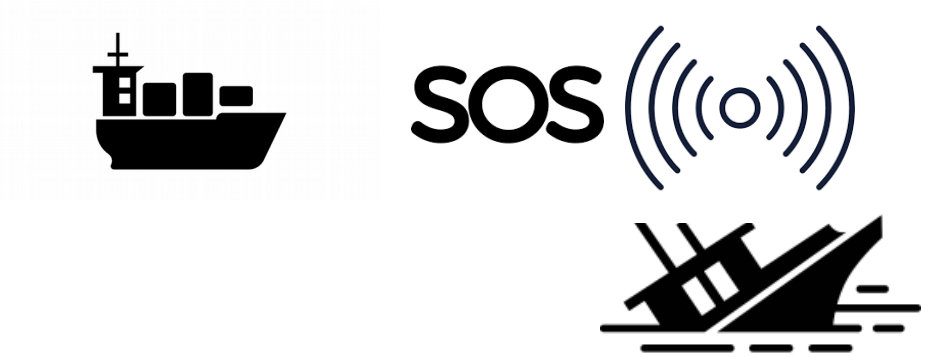
In context C, action A is preferred to its omission.



Function of norms

- One of the functions of norms is to express **relative preferences** that should guide behaviour

In context C, action A is preferred to its omission.



- Existence of a collective value function, or more plausibly, of a partial order:

$$C \rightarrow \nu_*(A) > \nu_*(\text{not } A) \quad \Rightarrow \quad C \rightarrow A >_{\nu_*} \text{not } A$$

collective value function

partial order

Norms per type of enforcement

- Relative expression of preference can be practically implemented in two forms:

punishment or penalty

Deontic directive

In context C, X has the duty of A, otherwise she will obtain P.

reward

Potestative directive

In context C, X has the power to obtain R by performing A.

Norms per type of enforcement

- Relative expression of preference can be practically implemented in two forms:

punishment or penalty

Deontic directive

In context C, X has the duty of A, otherwise she will obtain P.

reward

Potestative directive

In context C, X has the power to obtain R by performing A.

By whom?
Implicit reference to some *enforcer*

Formally, punishments and rewards are indistinguishable!

- A contract can be written as:
  - a price of $100 and a **penalty for late performance** of $9
  - a price of $91 and a **bonus for timely performance** of $9.

- In both cases the delivering party
  - takes $100 if it completes performance on time
  - takes $91 if it completes it late.

Formally, punishments and rewards are indistinguishable!

- A contract can be written as:
  - a price of $100 and a **penalty for late performance** of $9
  - a price of $91 and a **bonus for timely performance** of $9.

- In both cases the delivering party
  - takes $100 if it completes performance on time
  - takes $91 if it completes it late.

Are we missing something?

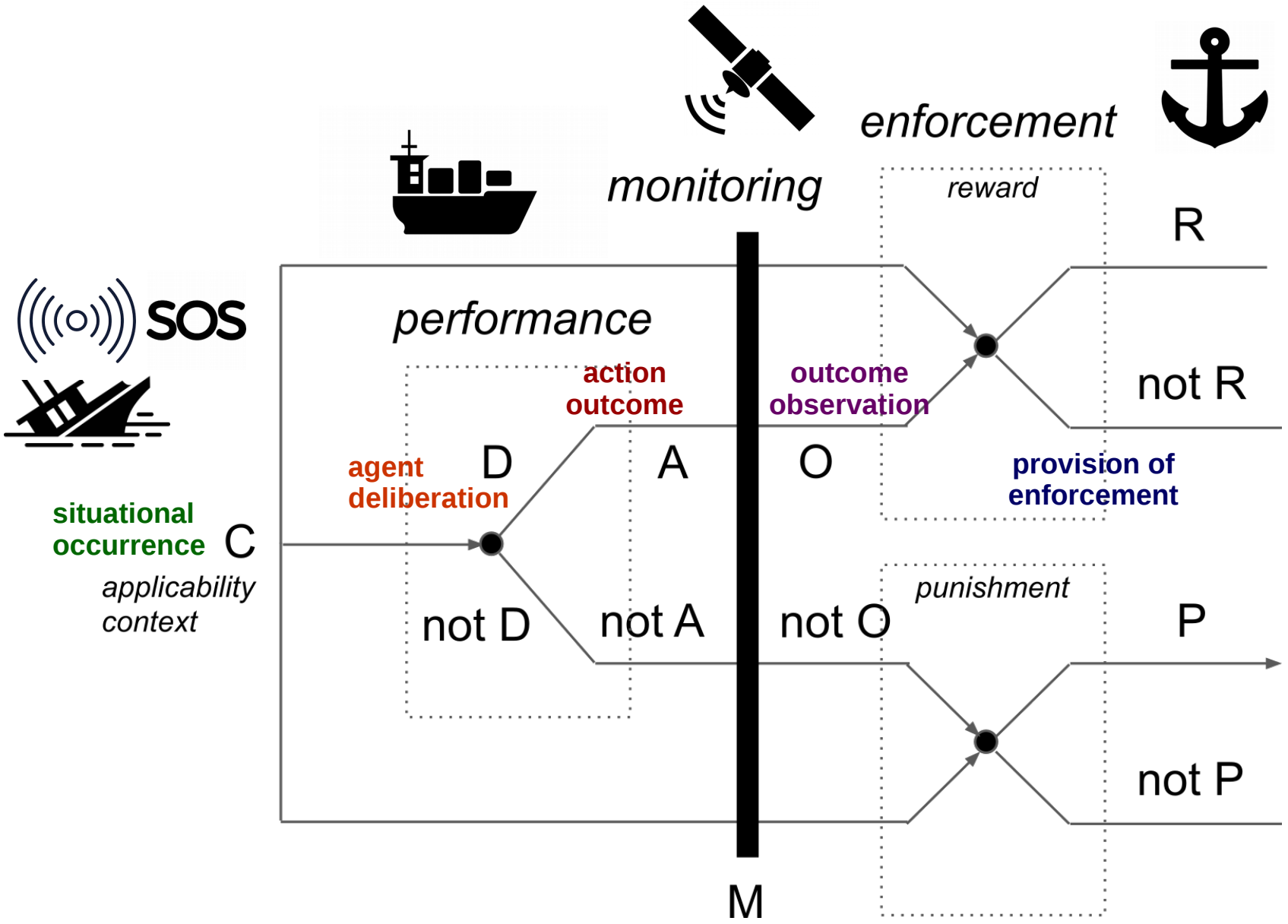
Monitoring requires resources!

(people, expertise, attention, time...)

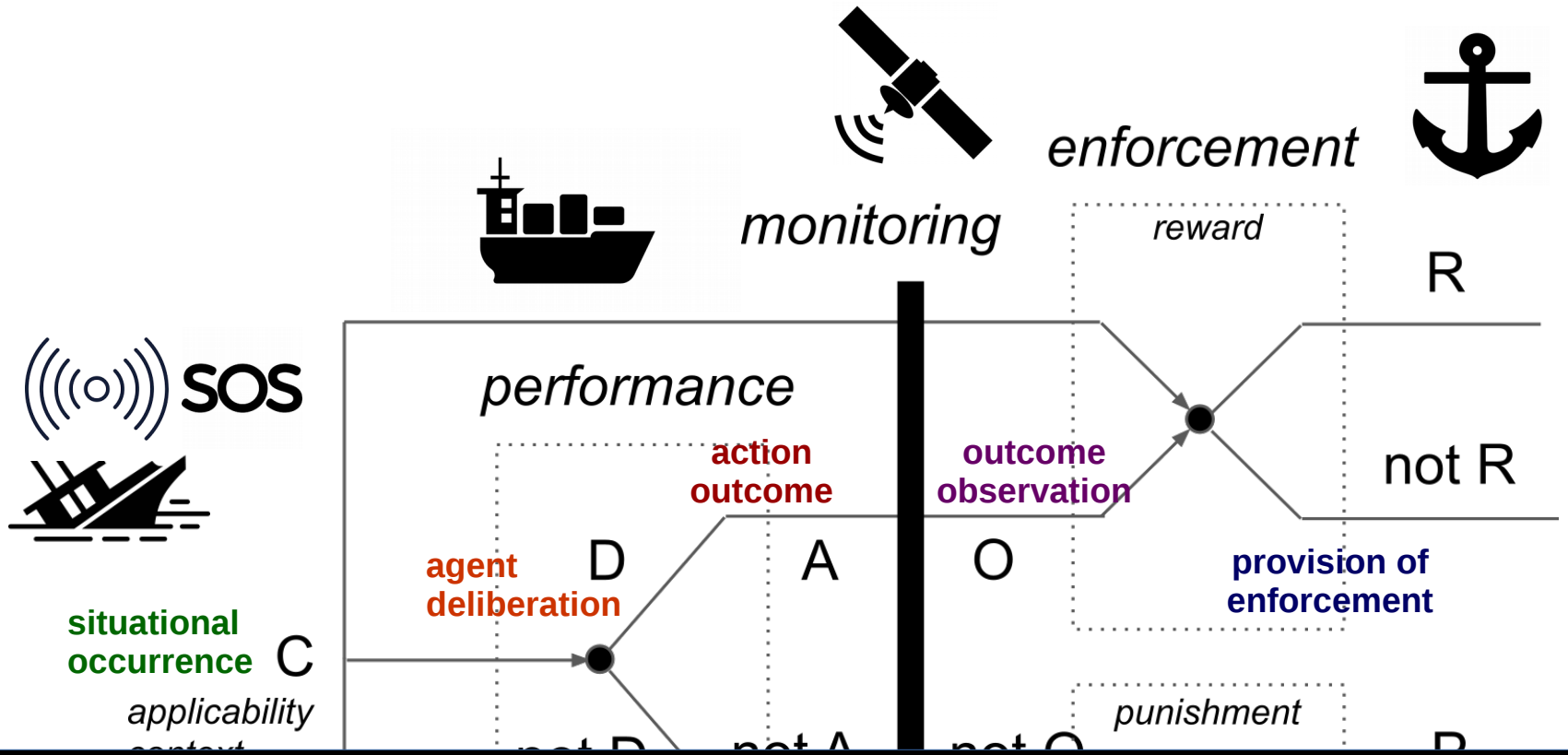
Monitoring requires resources and can be difficult!

(discriminating true positives from false positives/fakes)

Variables in the interaction



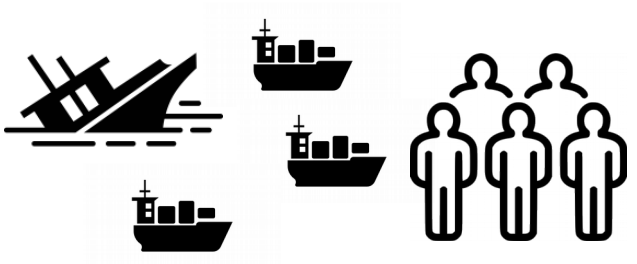
Variables in the interaction



The model can be easily enriched with non-linear, circular, non-additive relationships, complex internal models and dynamic aspects (e.g. agent adaptation to norms).

OBJECTIVE: going beyond static payoff tables.

Simplified economic flows



Authority	Agent X (addressee)	Collectivity
Monitoring cost: $m_p \cdot P(M) \cdot N$	Certification cost: c_r	Aggregated effects
Punishment benefit: $-p \cdot N_P$	Punishment cost: p	of performance:
Reward cost: $r \cdot N_R$	Reward benefit: $-r$	$(1 - \mathrm{PNC^e}) \cdot P(C) \cdot N \cdot e_*$
	Non-normative effects	Aggregated effects
	of performance: e_X	of non-performance:
	Non-normative effects	$\mathrm{PNC^e} \cdot P(C) \cdot N \cdot f_*$
	of non-performance: f_X	

Costs per transaction
(including amortized costs)

Number
of agents

(aggregated)
potential of non-compliance

Observations on Sustainability

$$(1-\mathsf{PNC}^{\mathsf{e}})\cdot e_{*}-\mathsf{PNC}^{\mathsf{e}}\cdot f_{*}\geq m_{p}\cdot\frac{P(M)}{P(C)}-p\cdot P(P|\text{not }A)\cdot\mathsf{PNC}^{\mathsf{e}}+r\cdot P(R|A)\cdot(1-\mathsf{PNC}^{\mathsf{e}})$$

Observations on Sustainability

$$(1-\mathrm{PNC}^{e})\cdot e_{*}-\mathrm{PNC}^{e}\cdot f_{*}\geq m_{p}\cdot\frac{P(M)}{P(C)}-p\cdot P(P|\text{not }A)\cdot\mathrm{PNC}^{e}+r\cdot P(R|A)\cdot(1-\mathrm{PNC}^{e})$$

- Cases in which **sticks have to be preferred**:
  - If people are **generally compliant**, too many “carrots” make the system not sustainable.
  - Punishment works already if there is a **perceived threat of punishment**, in which case *P(M)* can be kept sufficiently low at some moments.

cf. Gerrit De Geest and Giuseppe Dari-Mattiacci. The Rise of Carrots and the Decline of Sticks.Universityof Chicago Law Review, 80(1):341–392, 2013.

Observations on Sustainability

$$(1-\mathrm{PNC}^{\mathrm{e}})\cdot e_{*}-\mathrm{PNC}^{\mathrm{e}}\cdot f_{*}\geq m_{p}\cdot\frac{P(M)}{P(C)}-p\cdot P(P|\mathrm{not}\,A)\cdot\mathrm{PNC}^{\mathrm{e}}+r\cdot P(R|A)\cdot(1-\mathrm{PNC}^{\mathrm{e}})$$

- Cases in which **carrots have to be preferred**:
  - **singling out** problem: unequal distribution of burden across agents (*P(C) ~ 0*)
  - **specification problem**: difficult definition of the expected behaviour, which increases m_p in order to have adequate increase of *P(not O|not A)*.

cf. Gerrit De Geest and Giuseppe Dari-Mattiacci. The Rise of Carrots and the Decline of Sticks.Universityof Chicago Law Review, 80(1):341–392, 2013.

Observations on Sustainability

$$(1-\mathrm{PNC}^{\mathrm{e}})\cdot e_{*}-\mathrm{PNC}^{\mathrm{e}}\cdot f_{*}\geq m_{p}\cdot\frac{P(M)}{P(C)}-p\cdot P(P|\text{not }A)\cdot\mathrm{PNC}^{\mathrm{e}}+r\cdot P(R|A)\cdot(1-\mathrm{PNC}^{\mathrm{e}})$$

- Cases in which **carrots have to be preferred**:
  - when **agents are deemed by default non-compliant**.

cf. Alexander Boer. Punishments, rewards, and the production of evidence. In Legal Knowledge and Infor-mation Systems Conference: JURIX 2014, FAIA 271, pages 97–102.

Observations on Sustainability

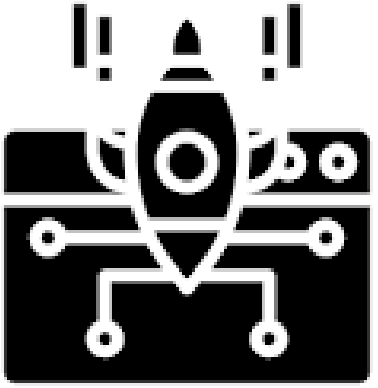
$$(1-\mathrm{PNC}^{\mathrm{e}})\cdot e_{*}-\mathrm{PNC}^{\mathrm{e}}\cdot f_{*}\geq m_{p}\cdot\frac{P(M)}{P(C)}-p\cdot P(P|\text{not }A)\cdot\mathrm{PNC}^{\mathrm{e}}+r\cdot P(R|A)\cdot(1-\mathrm{PNC}^{\mathrm{e}})$$

- Cases in which **carrots have to be preferred**:
  - when **agents are deemed by default non-compliant**.
    - increasing punishment is an alternative, but a rational choice for the agent would be to attempt ***avoidance*** behaviour (i.e. avoiding applicable conditions)
    - If applicability cannot be escaped, avoidance goes at meta-level, contesting the authority issuing the norm (eroding consensus)

cf. Alexander Boer. Punishments, rewards, and the production of evidence. In Legal Knowledge and Infor-mation Systems Conference: JURIX 2014, FAIA 271, pages 97–102.

Back to the initial problem...

Cyber-attack scenario



- ***If you suffer of a cyber-attack, share the information with the consortium***

- Beginning of the attack:

  P(attack) low → **singling out** problem

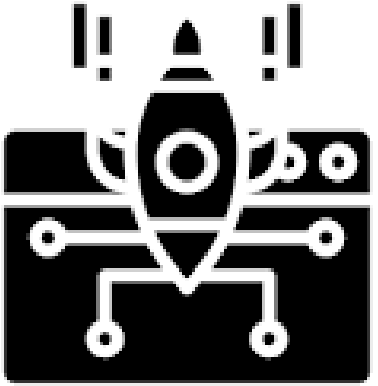
  unknown attack → **specification** problem

  → *"carrots"*

Sharing may be detrimental if the released data has competitive value

*Inspired by the SARNET project.

Cyber-attack scenario



- ***If you suffer of a cyber-attack, share the information with the consortium***

Sharing may be detrimental if the released data has competitive value

- Beginning of the attack:

  P(attack) low → **singling out** problem

  unknown attack → **specification** problem

  → *"carrots"*

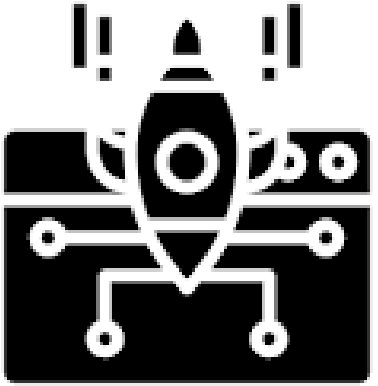
- Generalized attack

  higher P(attack)

  known attack → *"sticks"*

*Inspired by the SARNET project.

Cyber-attack scenario



- ***If you suffer of a cyber-attack, share the information with the consortium***

Sharing may be detrimental if the released data has competitive value

- Beginning of the attack:

  P(attack) low → **singling out** problem

  unknown attack → **specification** problem

  → *"carrots"*

- Generalized attack

  higher P(attack)

  known attack → *"sticks"*

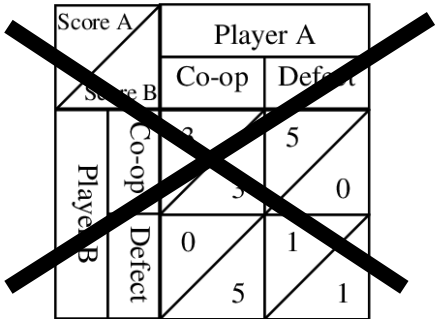
- If releasing information too expensive for the individual

  expected general non-compliance

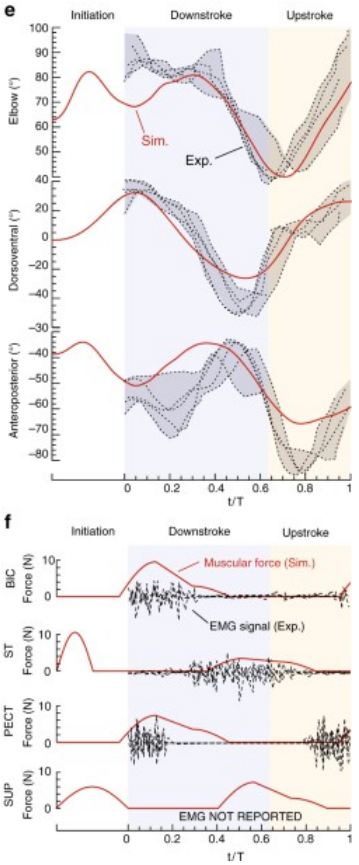
  → *"carrots"*

*Inspired by the SARNET project.

Conclusion



- Our research targets aspects of social-technical systems that *cannot* be treated by game-theoretical approaches based on **static pay-off tables**.
- With adequate values for the environmental parameters, and sound models (including non-linear, circular, etc.), the proposed template can be used to suggest **policy parameters** for ***monitoring*** and ***enforcement*** by means of optimization by **simulation** techniques,



GOAL: an integrated design platform for policy-making.

Monitoring and enforcement as a second-order guidance problem

10 December 2020. JURIX 2020 @ Brno/Prague (virtual)

Giovanni Sileno[a] (g.sileno@uva.nl)
Alexander Boer[b], Tom van Engers[c]

[a] Informatics Institute, University of Amsterdam, the Netherlands
[b] KPMG, Amsterdam, the Netherlands
[c] Leibniz Institute, TNO/University of Amsterdam, the Netherlands