

***Shifting Responsibilities: ethical challenges related to the use of reasoning bypassing technologies***

christine.clavien@unige.ch

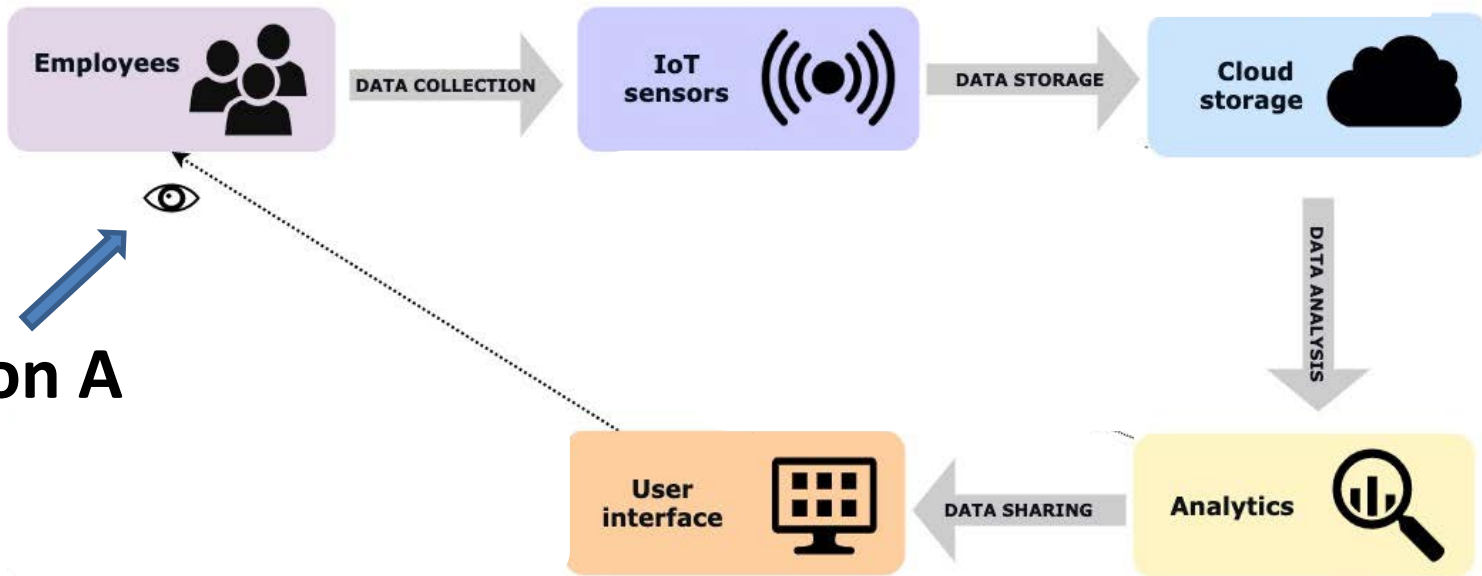
Bigorio jan 2023



## Scenario 1: occupational health

Aim: nudge employees to adopt a better posture at work

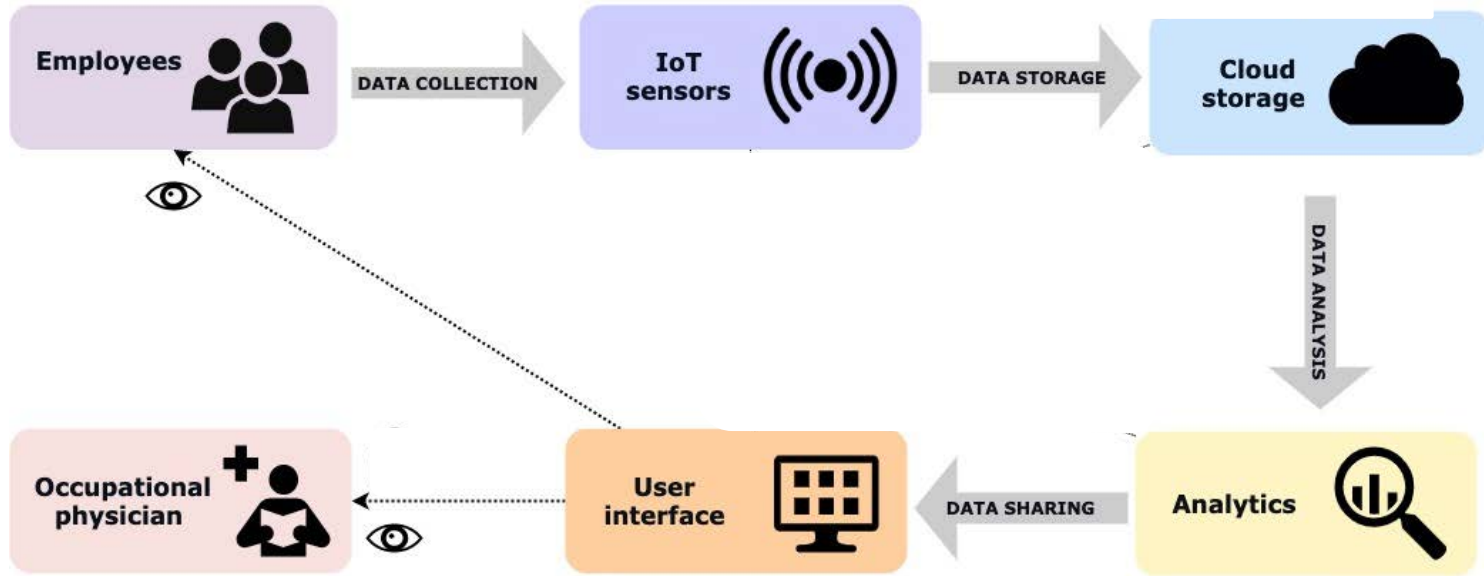
### Option A





## Scenario 1: occupational health

Aim: nudge employees to adopt a better posture at work



## Option B



## Scenario 2: prison cafeteria

Aim: nudge prisoners to eat more healthy food

**B**



**A**

odourless meat-disgust spray





Aim: reduce the risk of offending behavior among rehabilitated prisoners

## Scenario 3: neurostimulation



### «intelligent brain implanted device»

Functions:

**A**

- **Automated:** sends a signal in case of undesired impulse
- **Manual:** provides localized brain stimulation

**B**

Functions:

- **Automated:** sends a signal in case of undesired impulse
- **Automated:** provides localized brain stimulation

## Shared characteristics

All three scenarios are deployed for good purposes: health promotion (purposes shared with the target population)



**high-tech chair:** the company's CEO aims at reducing employees' back pain and adverse health effects of wrong posture.



**meat-disgust drug:** the prison administrator aims at preserving prison inmates from illnesses caused by unhealthy food intake.



**neurostimulation:** the head of the rehabilitation program aims at avoiding re-occurrence of compulsive aggression and successfully rehabilitate the former aggressive sex offenders.

## Shared characteristics

All three interventions are science based and use technologies that bypasses rational control : they target simple mechanisms or brain automatism known to influence action choices.



**high-tech chair:** studies have shown that timely warning or reminders tend to successfully enforce behaviours that are aligned with peoples' preferences; it a subtle push that helps people changing from procrastinators to actors.



**meat-disgust drug:** the drug is known to activate neurological paths that generates a feeling of disgust.



**neurostimulation:** targeted brain stimulation is known to suppress aggressive urges.

## Difference between options A & B

Difference in individuals' liberty to consent to the use of the technology: not all possibly attractive choice options are available.



**high-tech chair:** employees have no choice but using the chair and the information produced by the software is sent to the company's occupational physician without their consent



**meat-disgust drug:** prisoners have no way to avoid the drug if they want to eat.



**neurostimulation:** sex offenders have no choice but to accept the automated neurostimulation implant if they want to be admitted in the rehabilitation program



## Which option is ethically more acceptable?

- A**      The technology bypassing rational control is avoidable and consented by the targeted individuals
  
- B**      Autonomy of choice is constrained...

In short, whenever possible, we should preserve targeted individuals' autonomy of decision, avoid to design interventions that fail to treat the individuals as moral agents because it undermines their autonomy and moral responsibility.

# The focusing effect

The tendency to give too much importance to one (relevant) element or aspect of a situation or event, and correlatively fail to see or acknowledge the relevance of other elements.

**The focusing effect in normative ethics:** tendency to paying **exaggerated attention to human rational decision-making**, which is also illustrated by concerns for autonomy or freedom of action.



Partner pro & cons list	
Pros	Cons
Is so manly and stylish	Has no humor
Is handsome even when he sleeps with open mouth	Thought «Monty Python» is a python named «Monty»
Has a great deep romantic voice	Enjoys car racing and talks of cars all the time
Has an incomparable way to look at me	Not sure he is faithful
Is great to take out to parties	

# Ethical issues

## Structural feature

exploitation of biases & heuristics that bypass rational decision-making



## Risk

manipulation (threat to autonomous decision-making, rationality, liberty of choice)

## Remedies?

- shared preference justification
- consent justification
- empowerment to resist (e.g. via transparency)
- show the overriding importance of a conflicting good or principle → autonomy would be *infringed* but not *violated*

JOURNAL OF MORAL EDUCATION, 2018  
<https://doi.org/10.1080/03057240.2017.1408577>

 Routledge  
Taylor & Francis Group

 Check for updates

**Ethics of nudges: A general framework with a focus on shared preference justifications<sup>5</sup>**

Christine Clavien

Institute Ethics History Humanities, University of Geneva, Genève, Switzerland

## Structural feature

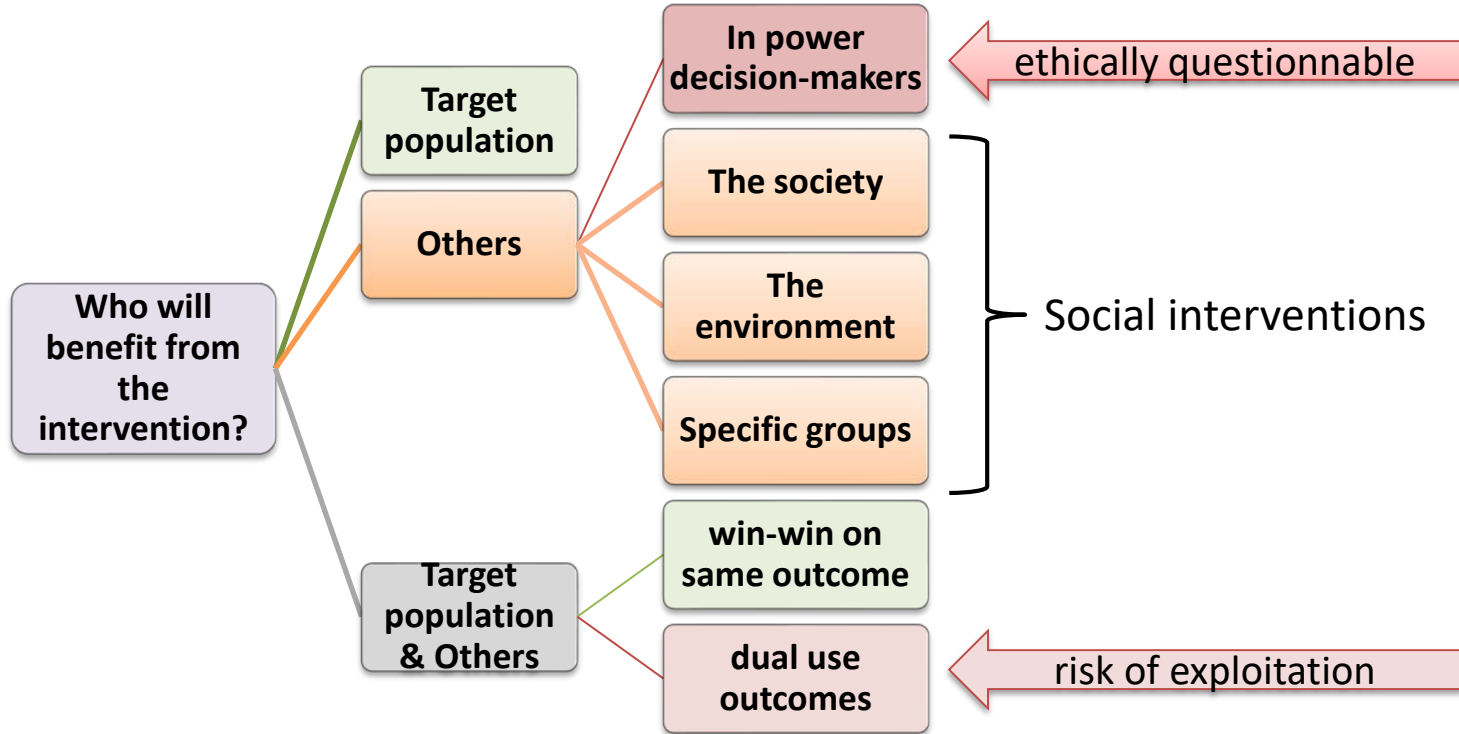
asymmetry of power  
between decision  
maker and target  
population



## Risk

abuse of power

# Nudging for the benefit of whom ?



## The risk of disproportionate responsibility attributed to patients for their illnesses



> [Camb Q Healthc Ethics](#). 2020 Apr;29(2):175-191. doi: 10.1017/S0963180119000975.

### The Undeserving Sick? An Evaluation of Patients' Responsibility for Their Health Condition

[Christine Clavien](#), [Samia Hurst](#)

PMID: 32159486 DOI: [10.1017/S0963180119000975](#)

Patients with unhealthy lifestyles are disproportionately blamed for their diseases

And more generally, humans tend to attribute disproportionate responsibility in case of bad consequences of behavioral choices

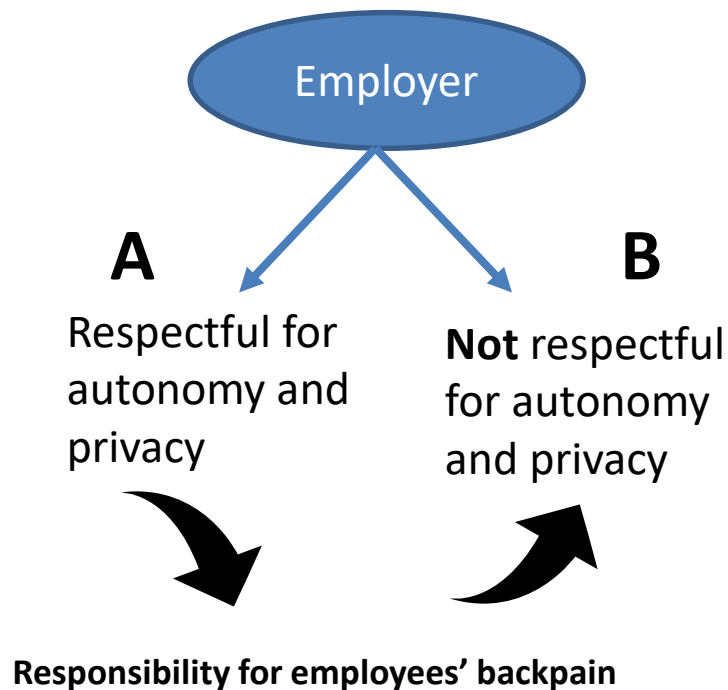
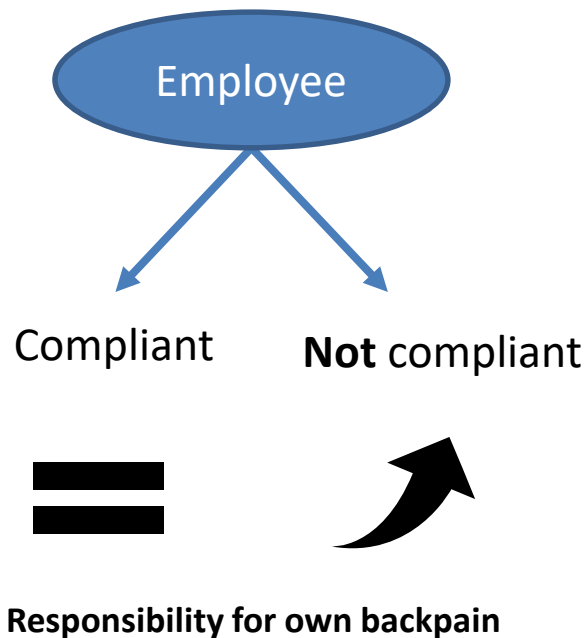


## The wrong logic

- Unwanted situation: an illness it is likely to cause organizational difficulties at work (e.g. colleagues need to take up the job).
- Colleagues do not like such unwanted situations, especially if the situation (seemingly) could have been avoided.
- People feel the urge to find culprits for unwanted situations & tend to focus on some salient features of the unwanted situation at the expense of other relevant features (focusing effect again!).
- People tend to overestimate the effectiveness of technological devices.
- The chairs are installed with the explicit intention. So it seems that there is a solution to avoid the unwanted situation.
- The ill employee has refused the chair offered or has repetitively failed to comply with the chair instructions.
- Voluntary persistence in bad habit is an easy explanation.
- The ill employee will be likely (wrongly) attributed full responsibility for her back pain → the unlucky ill has de double pain of being sick and being hold responsible for it.



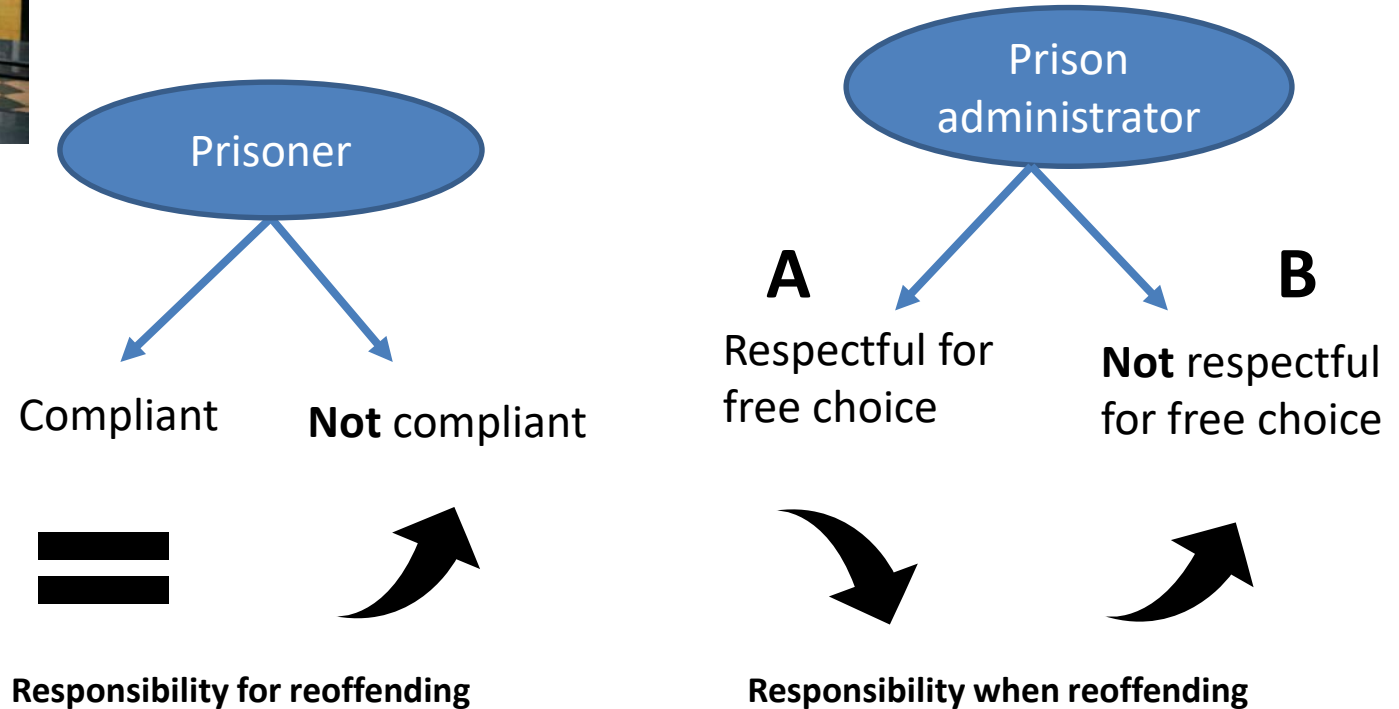
## Consented interventions involve shifting responsibilities





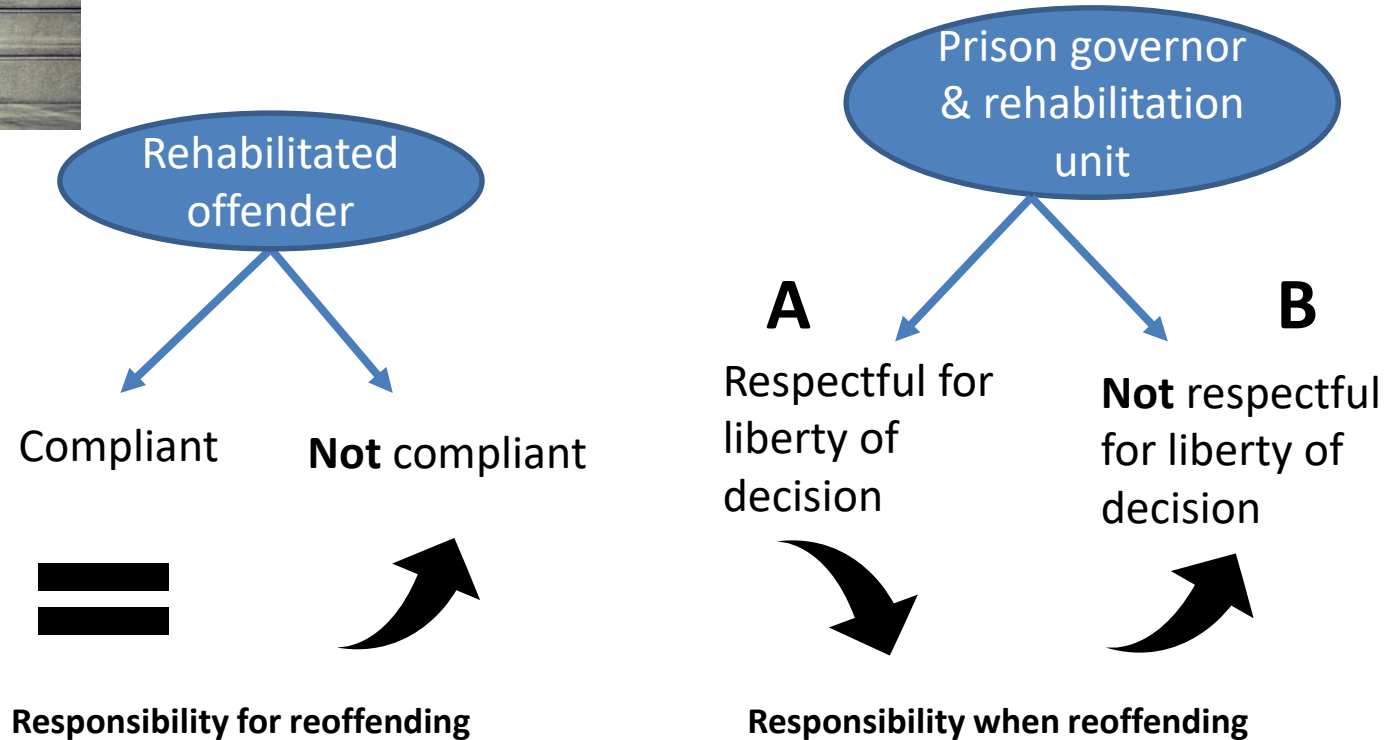


## Consented interventions involve shifting responsibilities

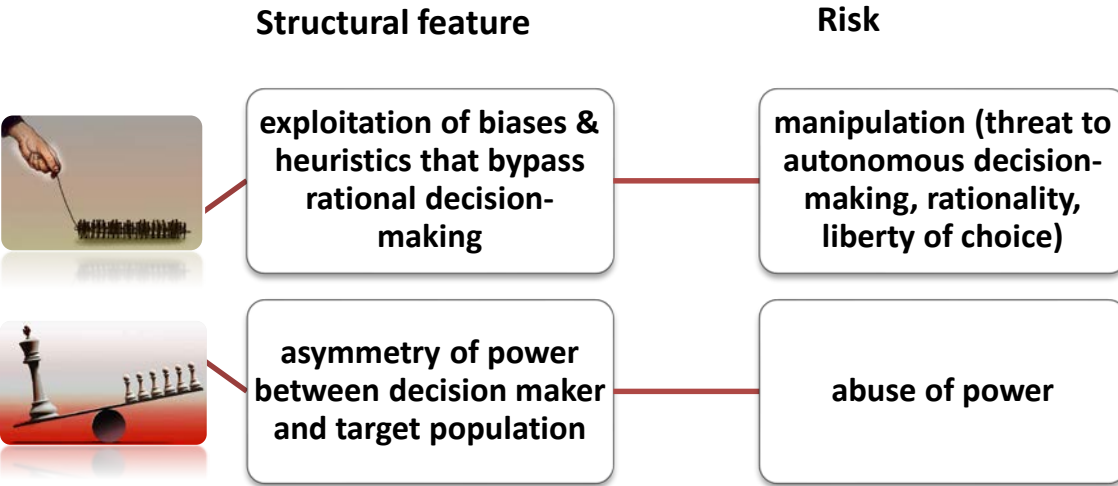




## Consented interventions involve shifting responsibilities



# Ethical issues



Are there situations in which abuse of power on the targeted population could be convincingly counter-balanced with other moral values?

## Structural feature

asymmetry of power  
between decision  
maker and target  
population



## Risk

abuse of power

## Remedies?

Asymmetry of power involves increased responsibility on part of decision-makers: the necessity to take care of vulnerabilities. For this, we need to develop specific codes of conduct.

At minima, decision-makers are expected to actively

1. identify dual use situations
2. Identify target population' vulnerabilities related to the dual use situations (what wrong could be done to them?)
3. Take care of these vulnerabilities

## Concluding remark on the focusing effect in normative ethics

The risk if we focus too much on autonomy concerns or respect for rational decision-making is to overlook...

- the problem of shifting responsibilities in relation to dual use interventions
- nudgers “additional” responsibility to protect vulnerable target populations

