

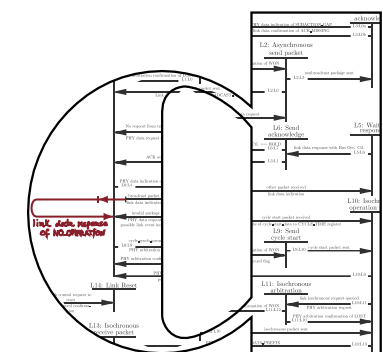


# Ethics for Nerds

An Advanced Course in Computer Science  
Summer Semester 2020

Ethics 4.1  
Deontology

Kantianism

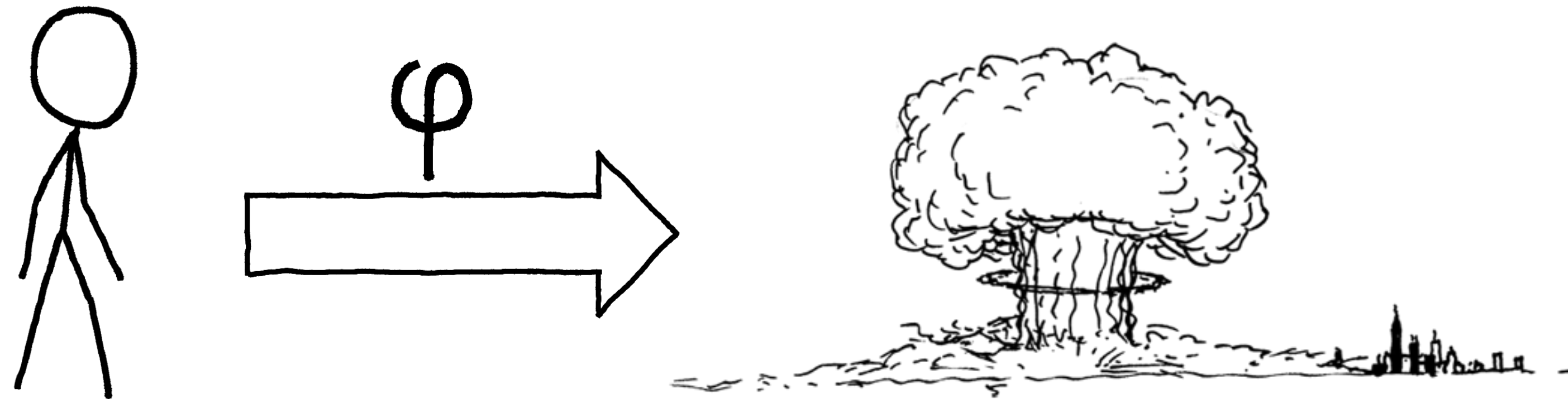


Prof. Holger Hermanns,  
Kevin Baum, Sarah Sterz



UNIVERSITÄT  
DES  
SAARLANDES

## A Classical Description



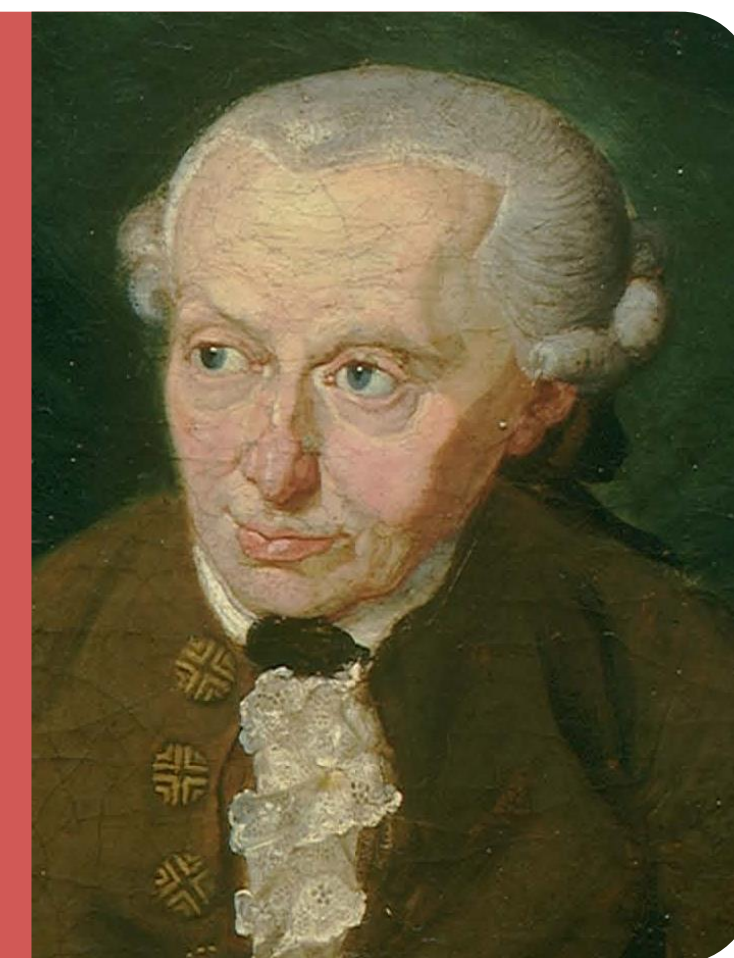
An **agent** performs an **act**. An **act** has **consequences**.

### Deontological Theories

most famously: Immanuel Kant

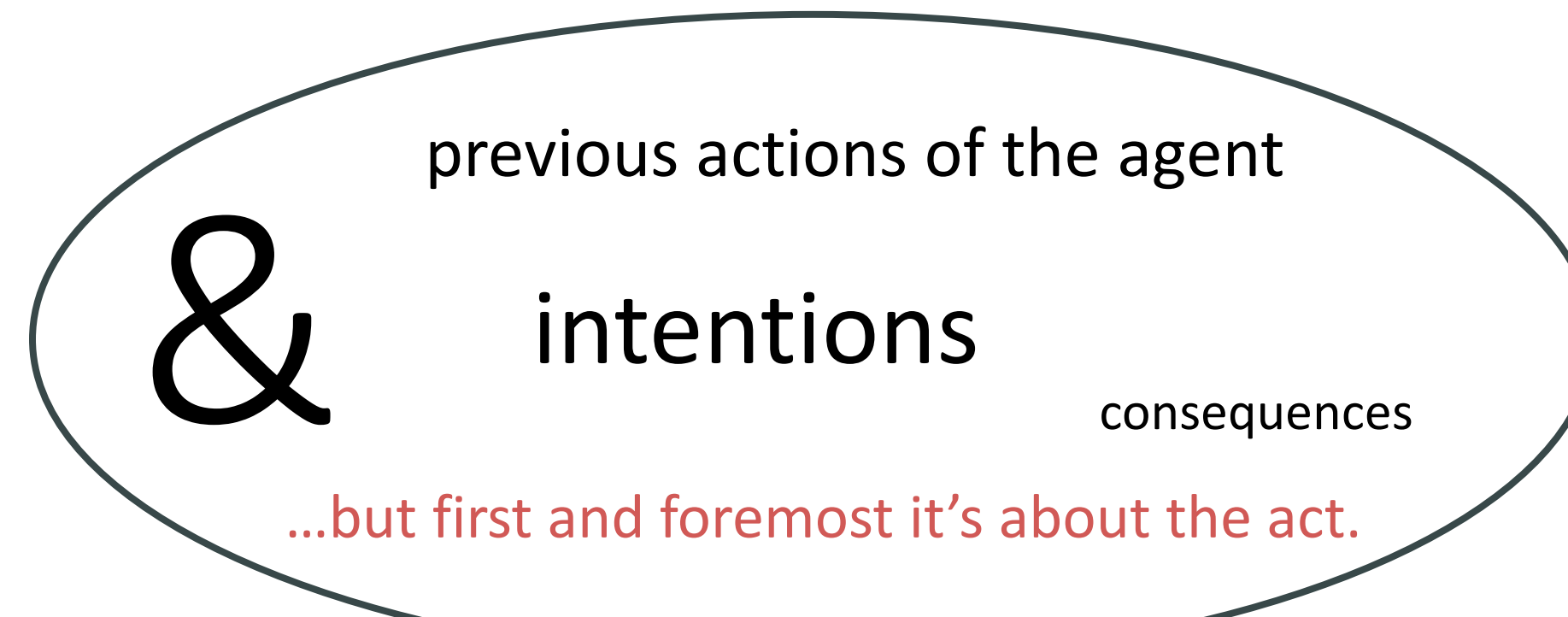
1748–1832

*Grundlegung zur Metaphysik der Sitten*



# THE FAMILIES OF THEORIES OF NORMATIVE ETHICS

## A Classical Description



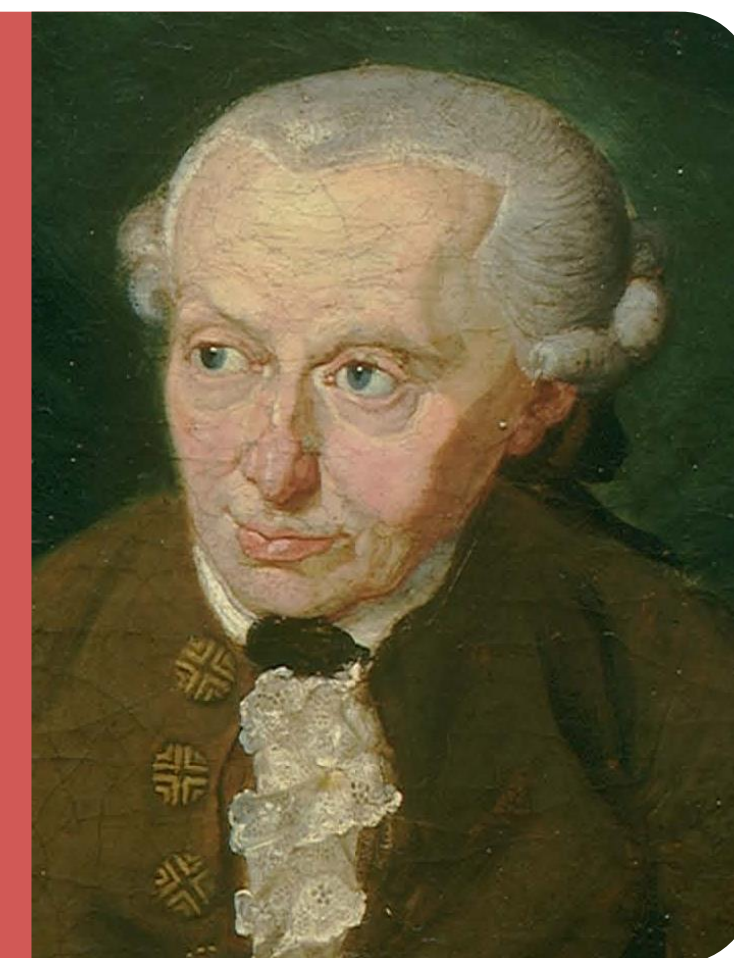
An **agent** performs an **act**. An **act** has **consequences**.

### Deontological Theories

most famously: Immanuel Kant

1748–1832

*Grundlegung zur Metaphysik der Sitten*



The moral/normative/deontic status of an action solely depends on its adherence to a rule or rules.

## Standard 'framework':

Agent  $A$  ought to perform one of the right actions.

An action  $\phi$  is a right iff  $\phi$ ing adheres to/follows/respects **certain (universal) rules**.

## Which rules?

Golden Rule

Categorical Imperative

Contractualistic rules

...

## Golden Rule

*not so golden after all*

### positive or directive form

One should treat others as one would like others to treat oneself.

### negative or prohibitive form

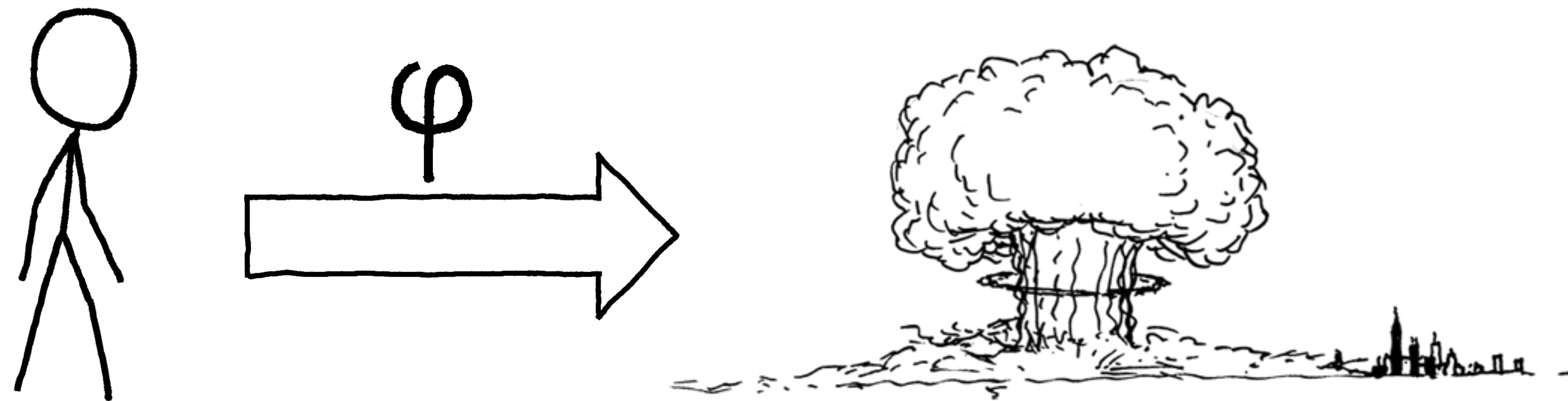
One should not treat others in ways that one would not like to be treated.

### Problem:

- either **too subjective**  
(what is good for me does not need to be good for others)
- or **underspecified**  
(if only certain ways of treating others are covered by the Golden Rule, it should tell us which one)



## A Classical Description



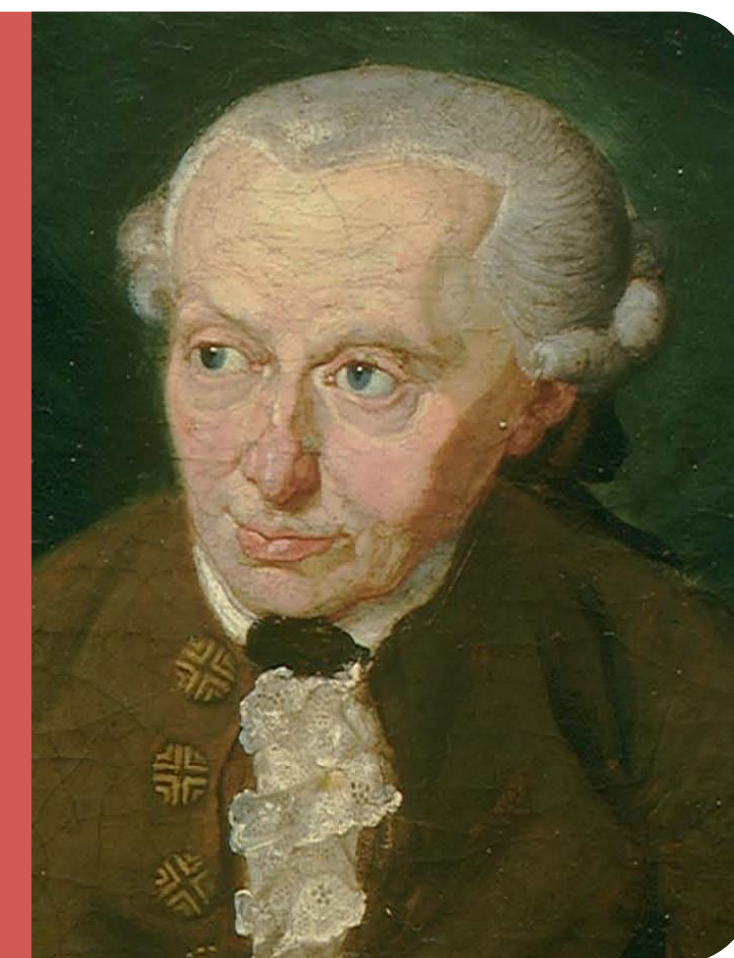
An **agent** performs an **act**. An **act** has **consequences**.

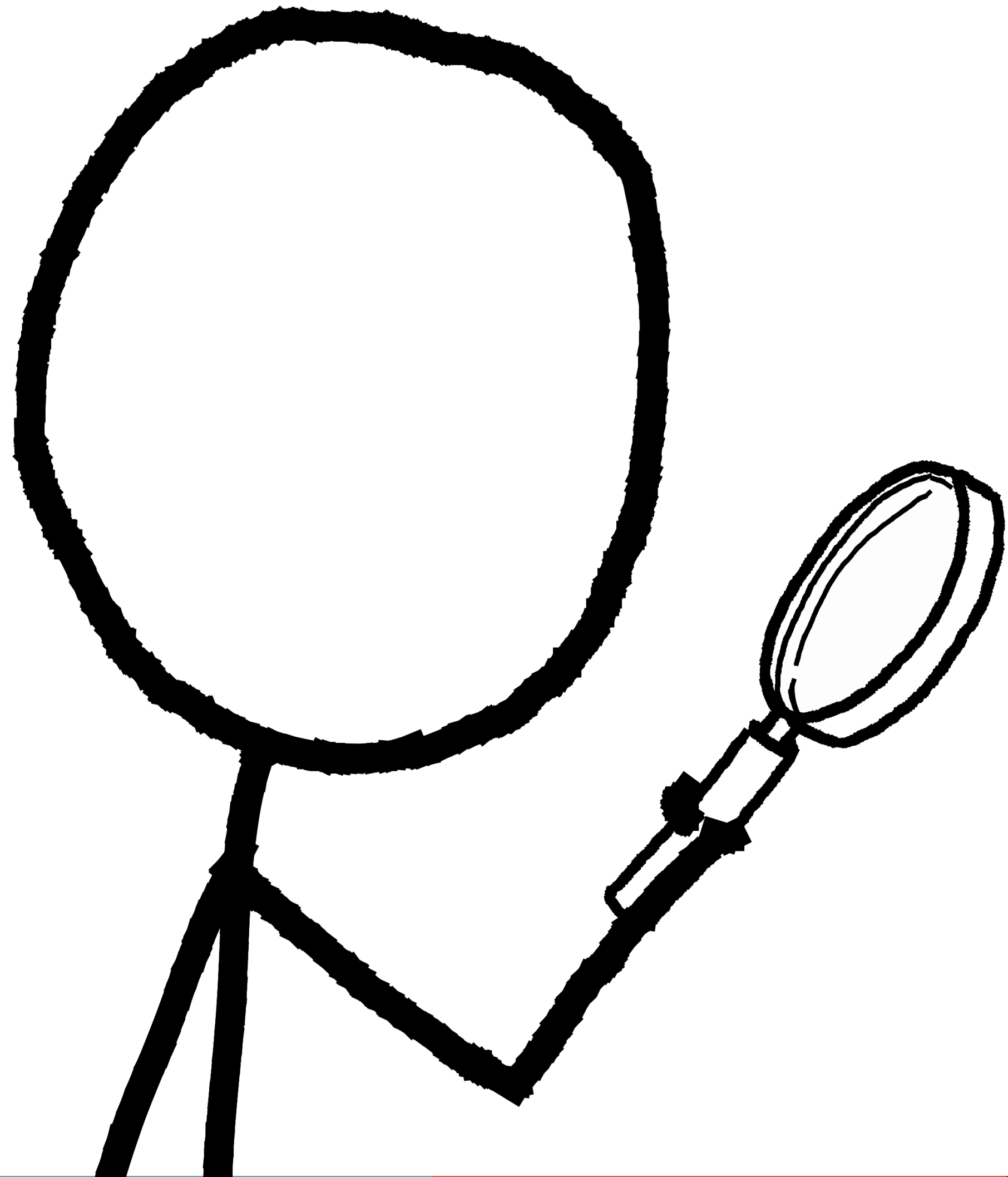
### Deontological Theories

most famously: Immanuel Kant

1748–1832

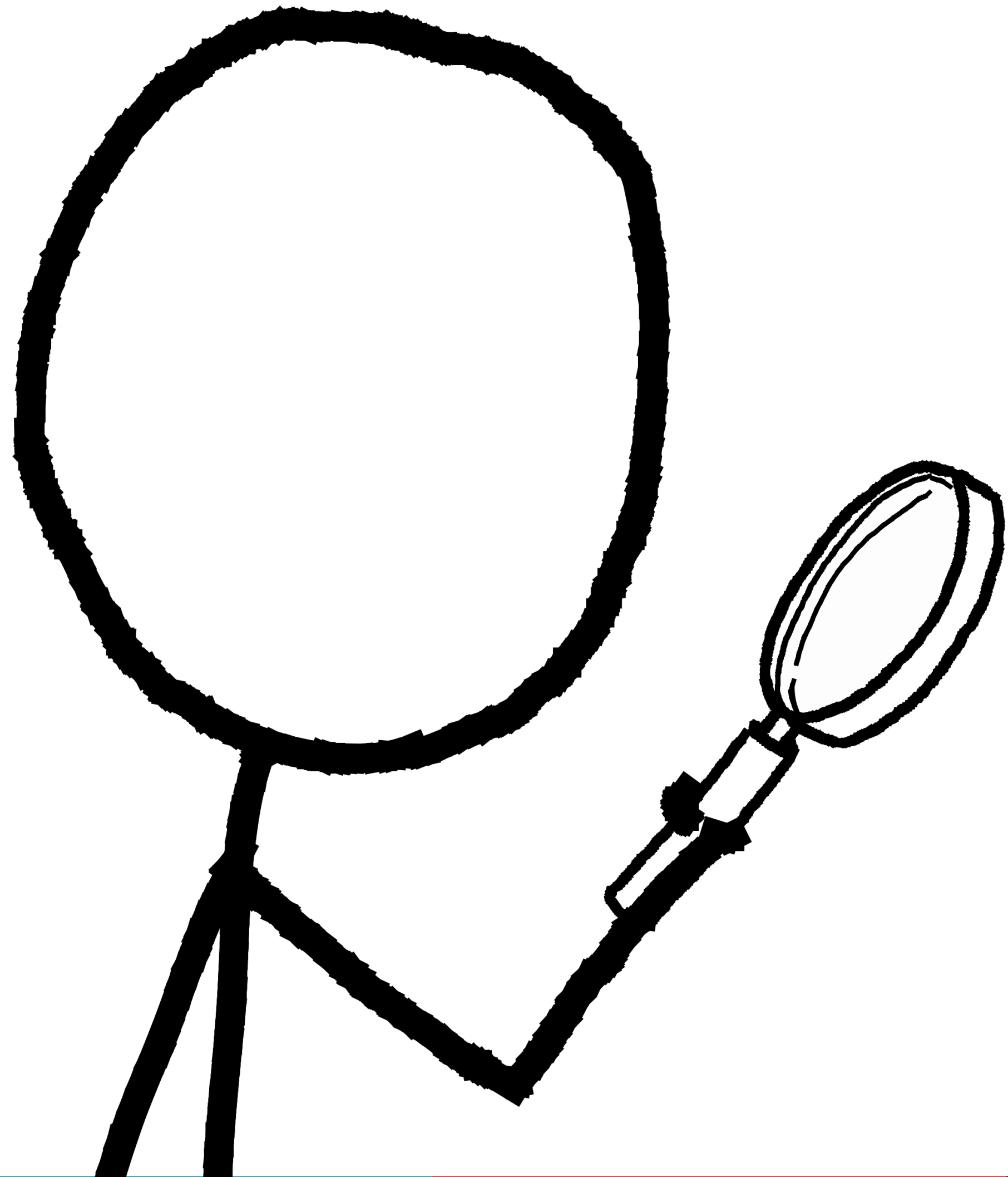
*Grundlegung zur Metaphysik der Sitten*





Kant was looking for the fundamental principle of morals.

He roughly found that **an action is right if and only if it is motivated by moral duty.**



Kant was looking for the fundamental principle of morals.

He roughly found that **an action is right if and only if it is motivated by moral duty.**

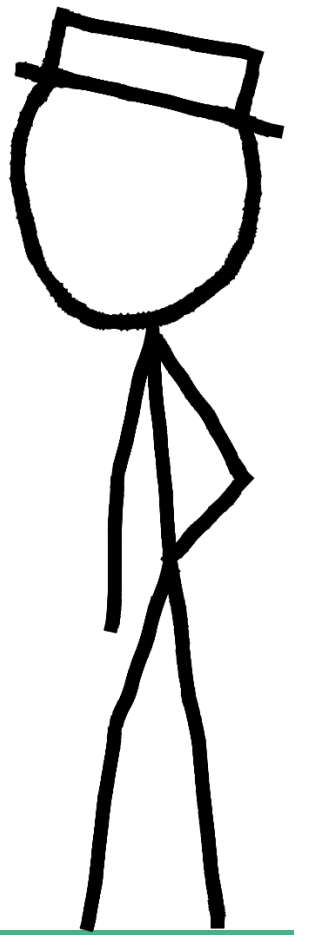




Merchant A:

is honest with their customers  
**but only out of self-interest**

acting merely in accordance with duty

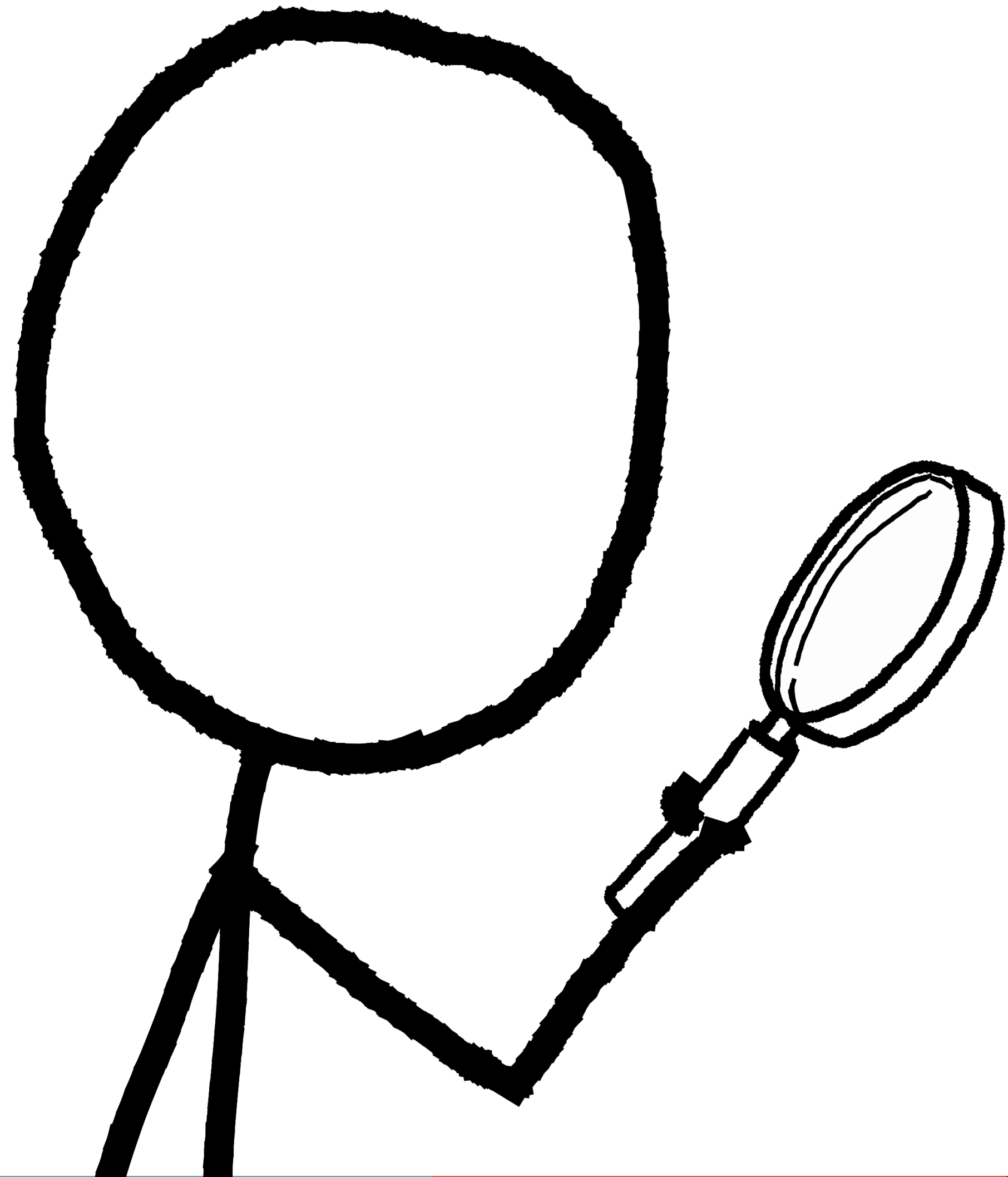


Merchant B:

is honest with their customers  
**because this is what moral duty demands**

acting from the motive of duty

everything less than this is too little for Kant



Kant was looking for the fundamental principle of morals.

He roughly found that **an action is right if and only if it is motivated by moral duty.**

reason (and only reason) can tell us what moral duty demands,  
but Kant was nice enough to spell it out for us:

## The Categorical Imperative

which is the fundamental principle of morals according to Kant

# THE CATEGORICAL IMPERATIVE

## formula of universality \*

Act only according to that maxim whereby you can at the same time will that it should become a universal law.

## formula of the law of nature

Act as if the maxims of your action were to become through your will a universal law of nature.

## formula of humanity \*

Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end.

## formula of autonomy

Act so that through your maxims you could be a legislator of universal laws.

## kingdom-of-ends formula

Act according to maxims of a universally legislating member of a merely possible kingdom of ends.

**all allegedly equivalent**

# THE CATEGORICAL IMPERATIVE

## perfect duties

- no latitude in how we want to fulfil them
- must never be violated

### Examples:

- duty not to kill
- duty not to lie

## imperfect duties

- some latitude in how we want to fulfil them
- roughly: do overall need to be fulfilled, but if I sometimes act against them, that's ok

### Examples:

- duty to develop our talents
- duty to help others



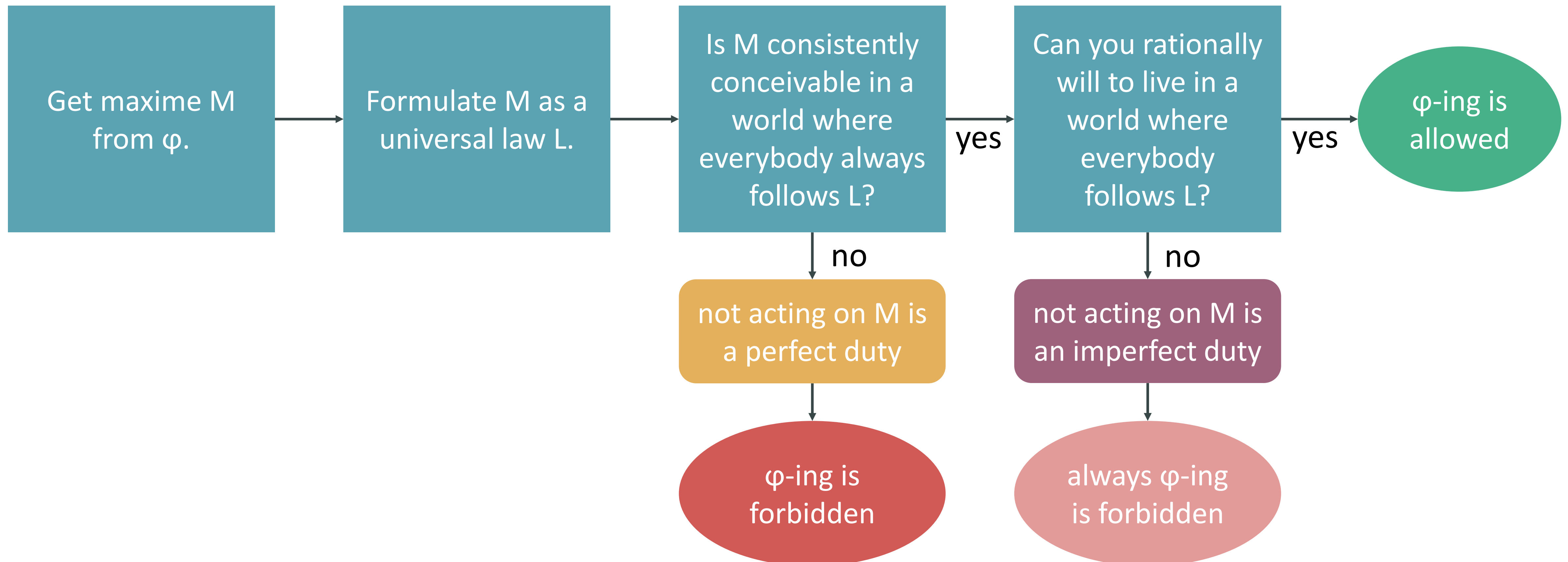
# FORMULA OF UNIVERSALITY

formula of universality



Act only according to that maxim whereby you can at the same time will that it should become a universal law.

From the formula of universality, you can derive the deontic status of action  $\varphi$  as follows:



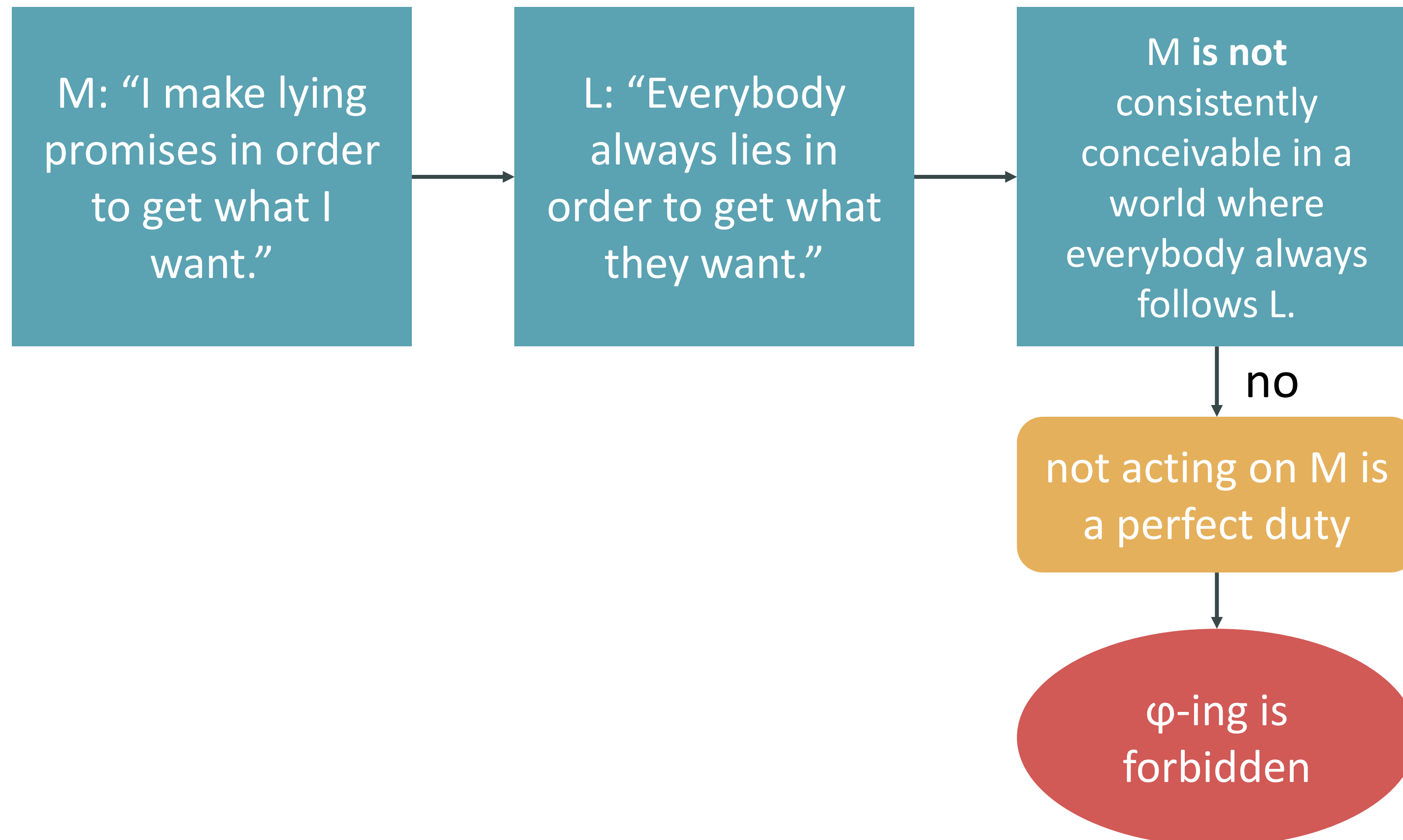
# FORMULA OF UNIVERSALITY

formula of universality



Act only according to that maxim whereby you can at the same time will that it should become a universal law.

$\varphi$ : promise something that you do not intent to keep in order to get money that you urgently need



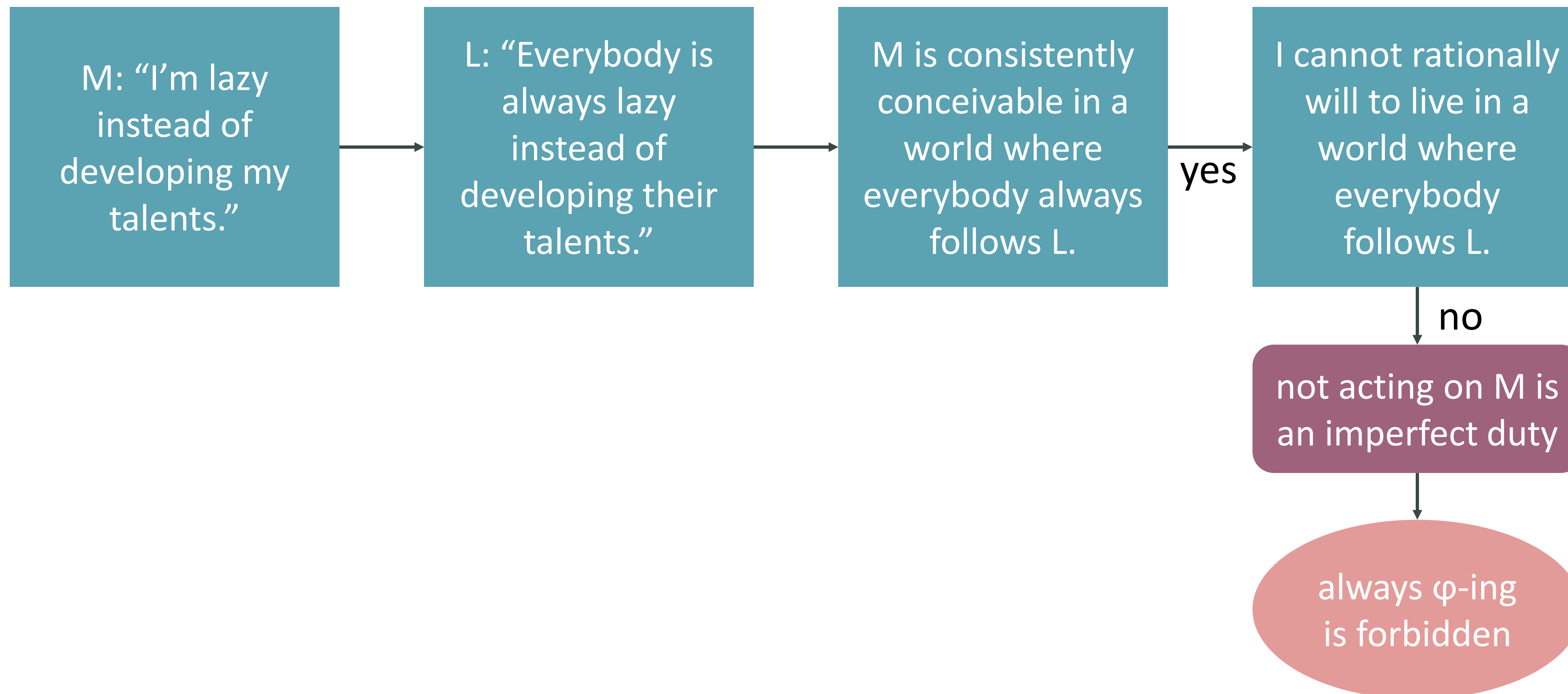
## FORMULA OF UNIVERSALITY

formula of universality



Act only according to that maxim whereby you can at the same time will that it should become a universal law.

$\varphi$ : being lazy and not practicing the violin (given that you are somewhat talented)



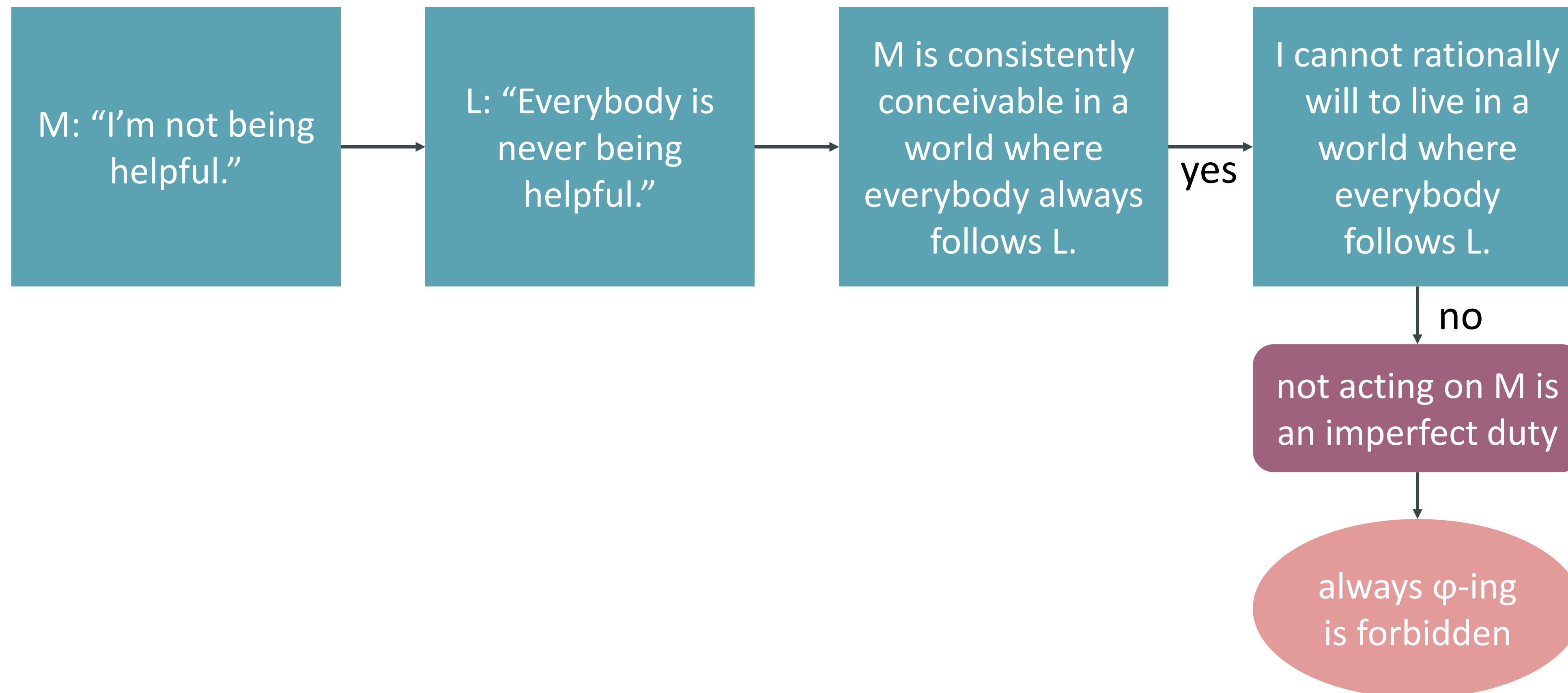
# FORMULA OF UNIVERSALITY

formula of universality



Act only according to that maxim whereby you can at the same time will that it should become a universal law.

$\varphi$ : not helping a neighbour



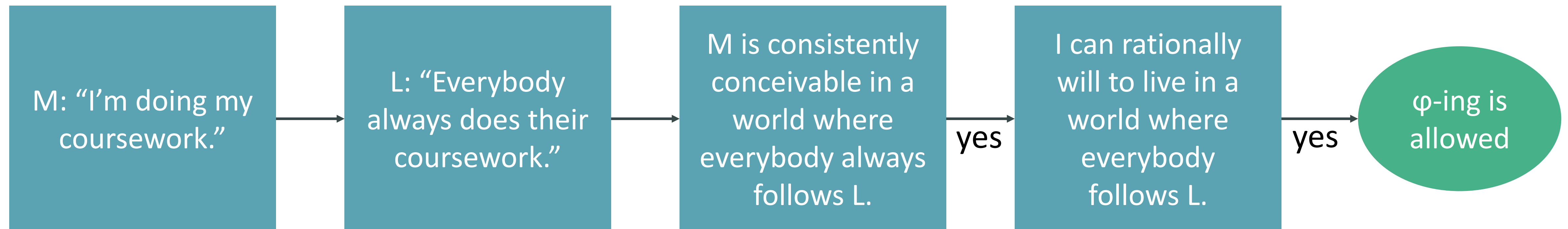
# FORMULA OF UNIVERSALITY

formula of universality



Act only according to that maxim whereby you can at the same time will that it should become a universal law.

$\varphi$ : doing my coursework





# THE CATEGORICAL IMPERATIVE

## formula of universality \*

Act only according to that maxim whereby you can at the same time will that it should become a universal law.

## formula of the law of nature

Act as if the maxims of your action were to become through your will a universal law of nature.

## formula of humanity \*

Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end.

## formula of autonomy

Act so that through your maxims you could be a legislator of universal laws.

## kingdom-of-ends formula

Act according to maxims of a universally legislating member of a merely possible kingdom of ends.

**all allegedly equivalent**

# FORMULA OF HUMANITY

## the formula of humanity



Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end.

φ: giving the postman a letter

ways in which I treat the postman's humanity as a means:

- I use him as a means to get my letter delivered

*you are allowed to treat them (or rather: their humanity) as a means, as long as you also treat them (or rather: it) as an end*

ways in which I treat the postman's humanity as an end:

- if we suppose that the postman has used his rational capacities when they chose to be a postman, we treat their humanity as an end if we approach them in a way that is usual for their job (e.g. give them a letter)

*technically, it's not about the person, but about their humanity: our humanity is what make us human, especially certain rational capacities. But in practice we simplify this for the scope of this lecture to: we treat someone's humanity as a means/an end iff we treat the person as a means/an end.*

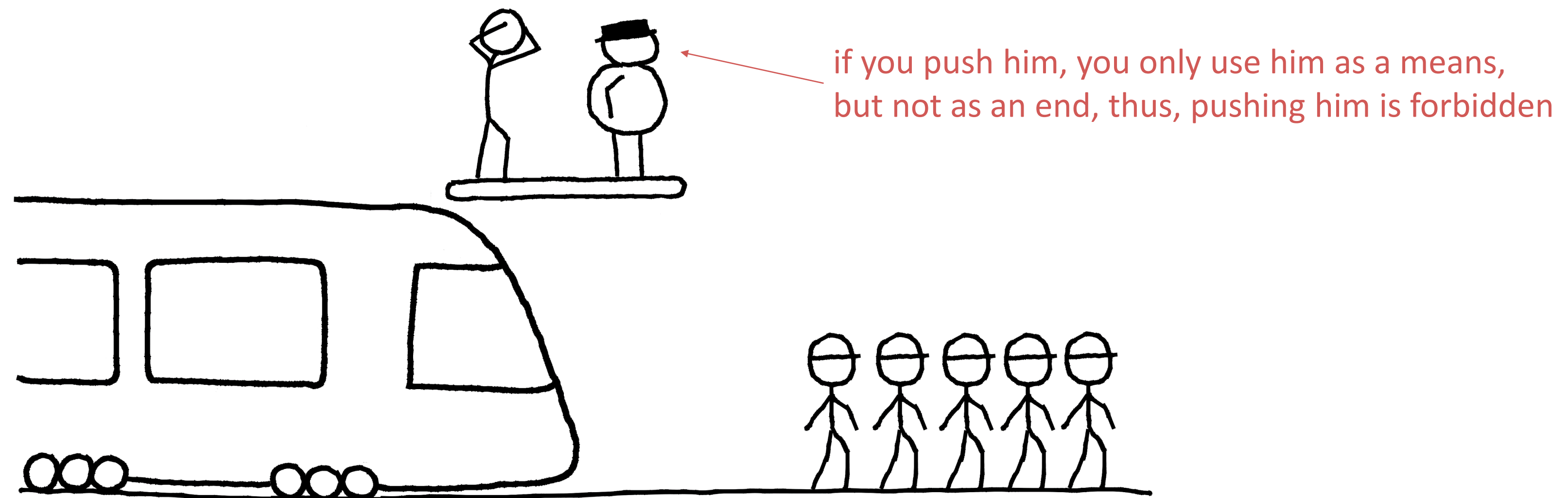
φ-ing is  
allowed

# FORMULA OF HUMANITY

## the formula of humanity \*

Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end.

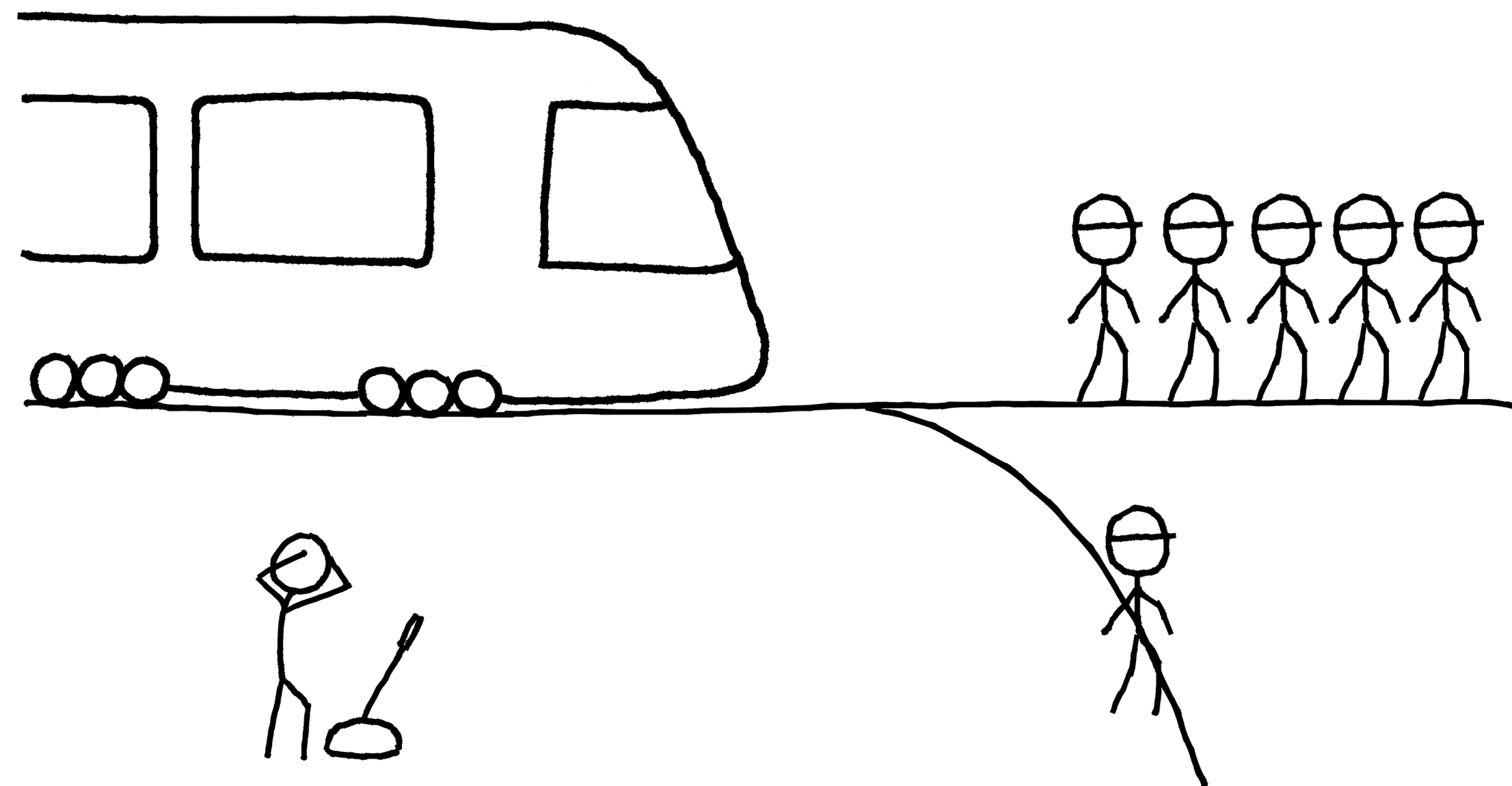
$\varphi$ : pushing the fat man



Act in such a way that you always treat humanity, whether in your own person or in the person of any other, never simply as a means, but always at the same time as an end.

# SHOULD I PULL THE LEVER?

$\varphi$ : pulling the lever



# PROBLEMS

general clarity and understandability

problem of formulating the maxime

(not useful as a decision procedure)

theory is in danger of contradicting itself

highly counterintuitive results!

*it's not permitted to lie to save  
someone from an unjust death*

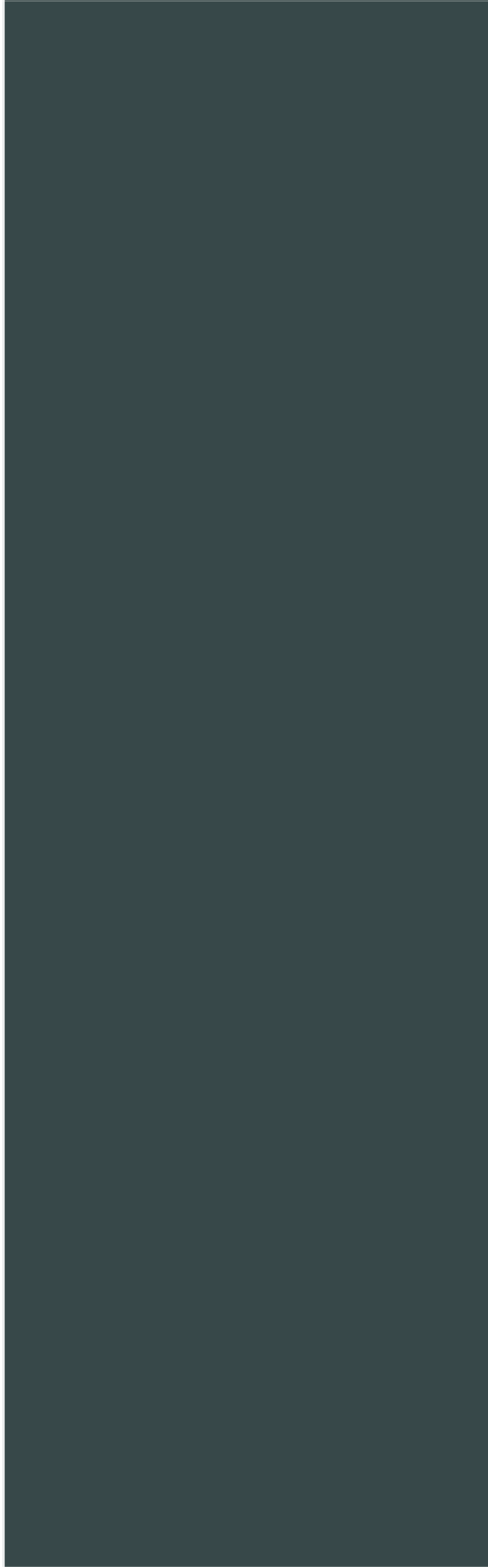
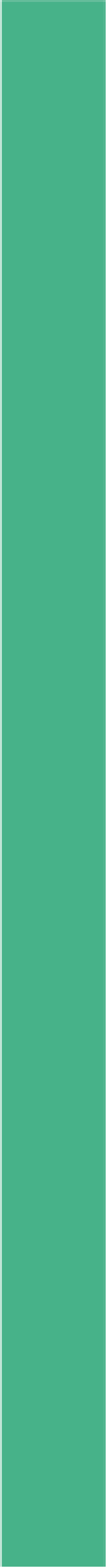
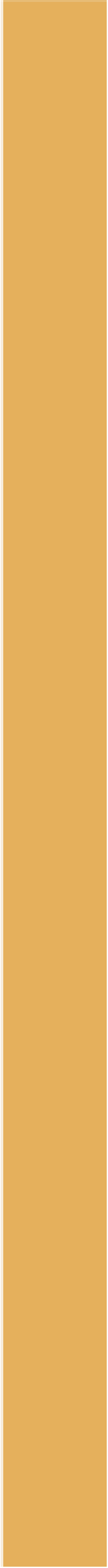
...



This video drew a lot on the SEP article on Kant's moral philosophy:

Johnson, Robert and Cureton, Adam, "Kant's Moral Philosophy", *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/spr2019/entries/kant-moral/>

If you want to get to know more, check it out!



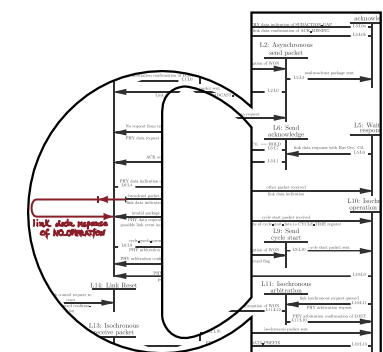


# Ethics for Nerds

An Advanced Course in Computer Science  
Summer Semester 2020

Ethics 4.2  
Deontology

Contractualism



Prof. Holger Hermanns,  
Kevin Baum, Sarah Sterz



UNIVERSITÄT  
DES  
SAARLANDES

The moral/normative/deontic status of an action solely depends on its adherence to a rule or rules.

## Standard 'framework':

Agent  $A$  ought to perform one of the right actions.

An action  $\phi$  is a right iff  $\phi$ ing adheres to **certain (universalizable) rules**.

## Which rules?

Golden Rule

Categorical Imperative

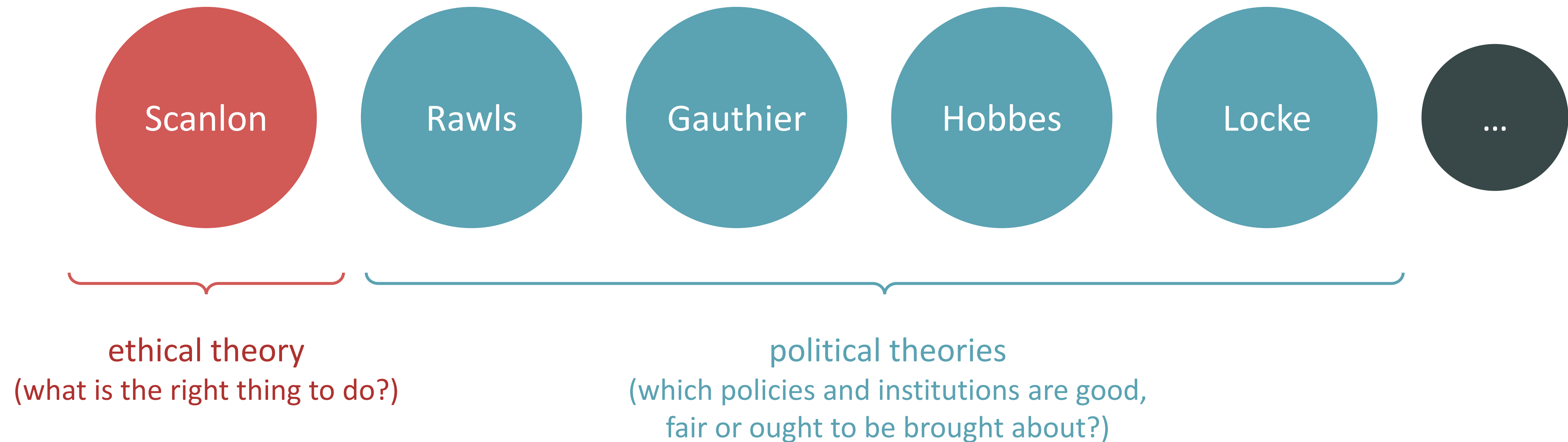
Contractualistic rules

...

## Contractualism

Rough core idea: Everybody ought to act upon principles which are (in a very specific way) very reasonable.

famous:

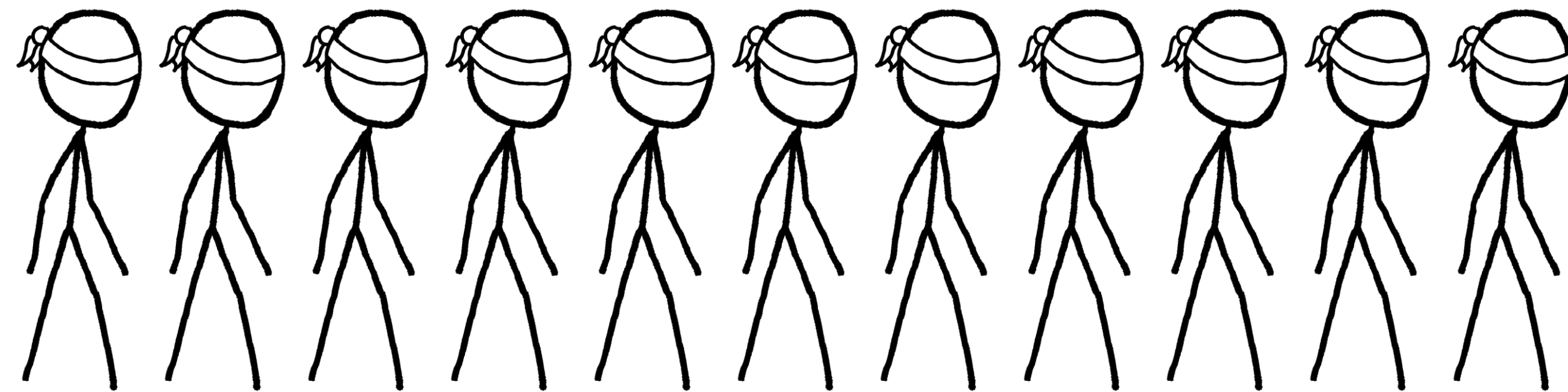




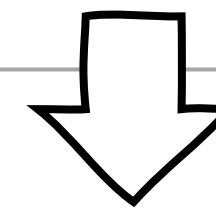
# RAWL'S ORIGINAL POSITION AND THE VEIL OF IGNORANCE

they formulate a contract without knowing anything about their future self (physique, gender, ability, social class, intelligence, sexual orientation, ...)

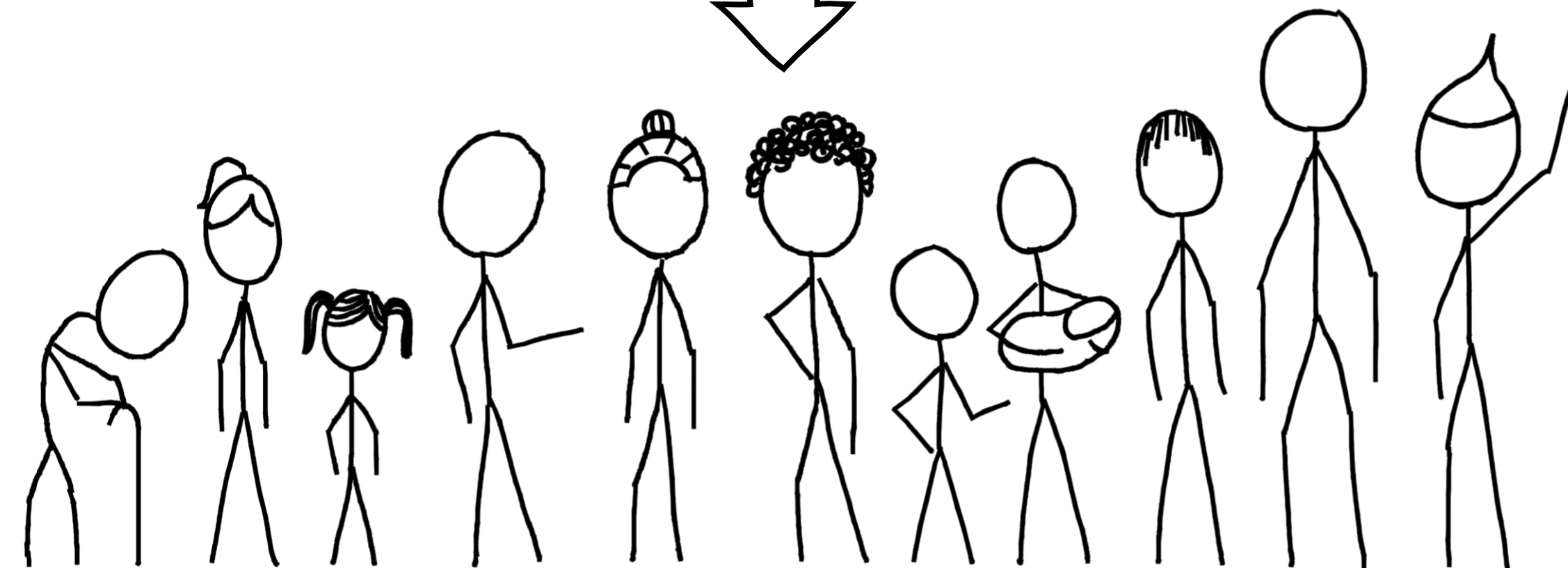
what they decide upon would be perfectly fair and impartial



people in the original position behind the veil of ignorance



they should implement/follow rules that the people in the original position would approve of



people in the real world

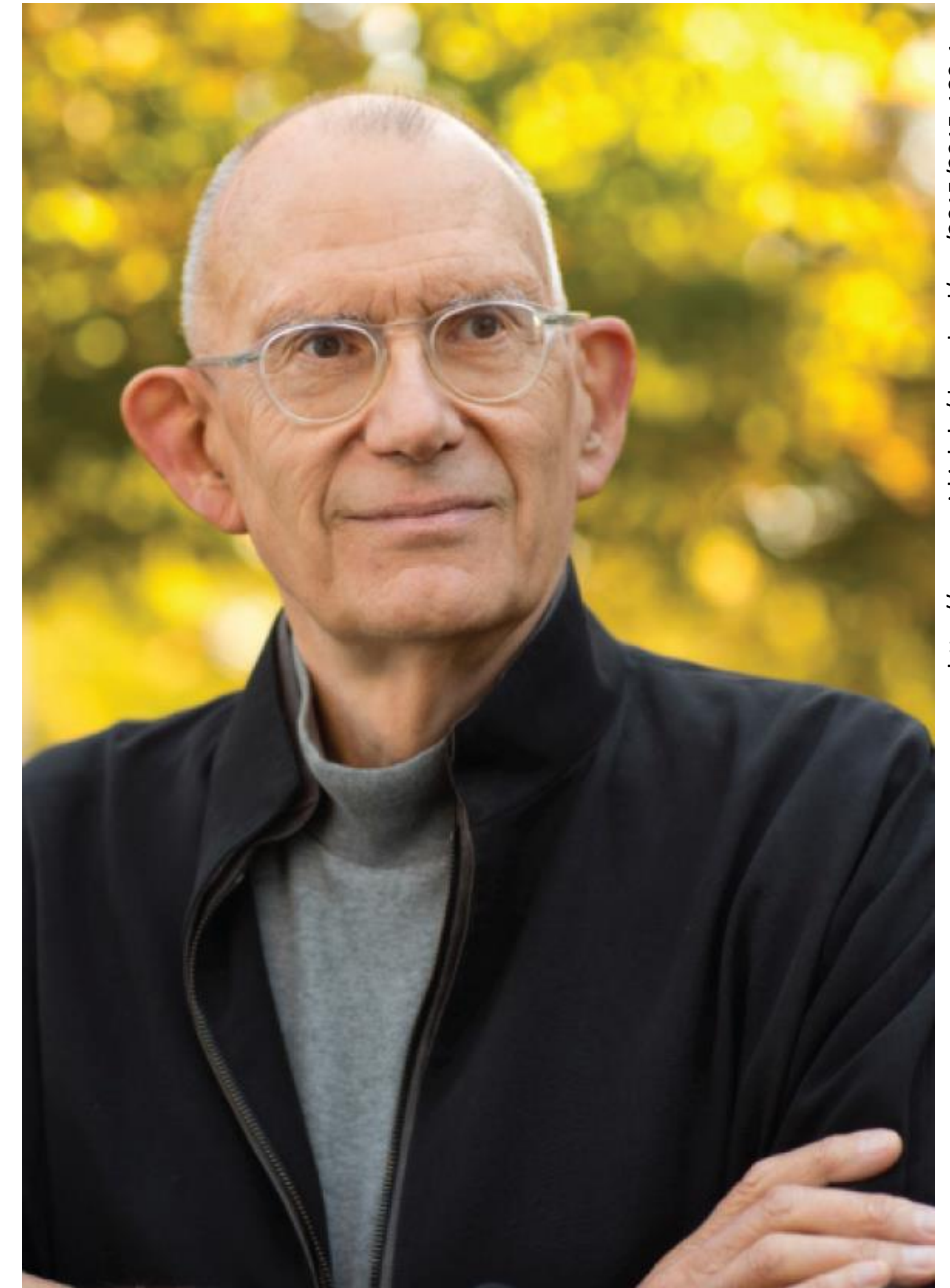
this is a political theory not an ethical theory, but it could be tweaked to be one

## Scanlon's Contractualism (simplified)

simplification for the  
purpose of this lecture

“An act is wrong if[f] its performance under the circumstances would be disallowed by [...all sets] of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced, general agreement.”

from Scanlon, 1998: *What We Owe to Each Other*



<http://www.uni-kiel.de/download/pm/2015-180-1.png>

## Scanlon's Contractualism (simplified)

"An act is wrong if[f] its performance under the circumstances would be disallowed by [...all sets] of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced, general agreement."

from Scanlon, 1998: *What We Owe to Each Other*

## Informed, unforced, and general agreement?

- informed: everyone knows and understands all principles in the set
- unforced: nobody is forced by anything else than their reason to accept a set of principles
- general: if a set were agreed upon, everybody would always follow all the principles in this set



## Scanlon's Contractualism (simplified)

"An act is wrong if[f] its performance under the circumstances would be disallowed by [...all sets] of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced, general agreement."

from Scanlon, 1998: *What We Owe to Each Other*

## Why all sets?

- there are (sets of) principles, which are mutually exclusive, but could still both be allowed
- example:
  - Implementing left-hand traffic is allowed.
  - Implementing right-hand driving is allowed.

though, not both can be allowed at the same time and in the same place.

## Scanlon's Contractualism (simplified)

"An act is wrong if[f] its performance under the circumstances would be disallowed by [...all sets] of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced, general agreement."

from Scanlon, 1998: *What We Owe to Each Other*

## What is reasonable rejection?

I can reasonably reject a set of principles

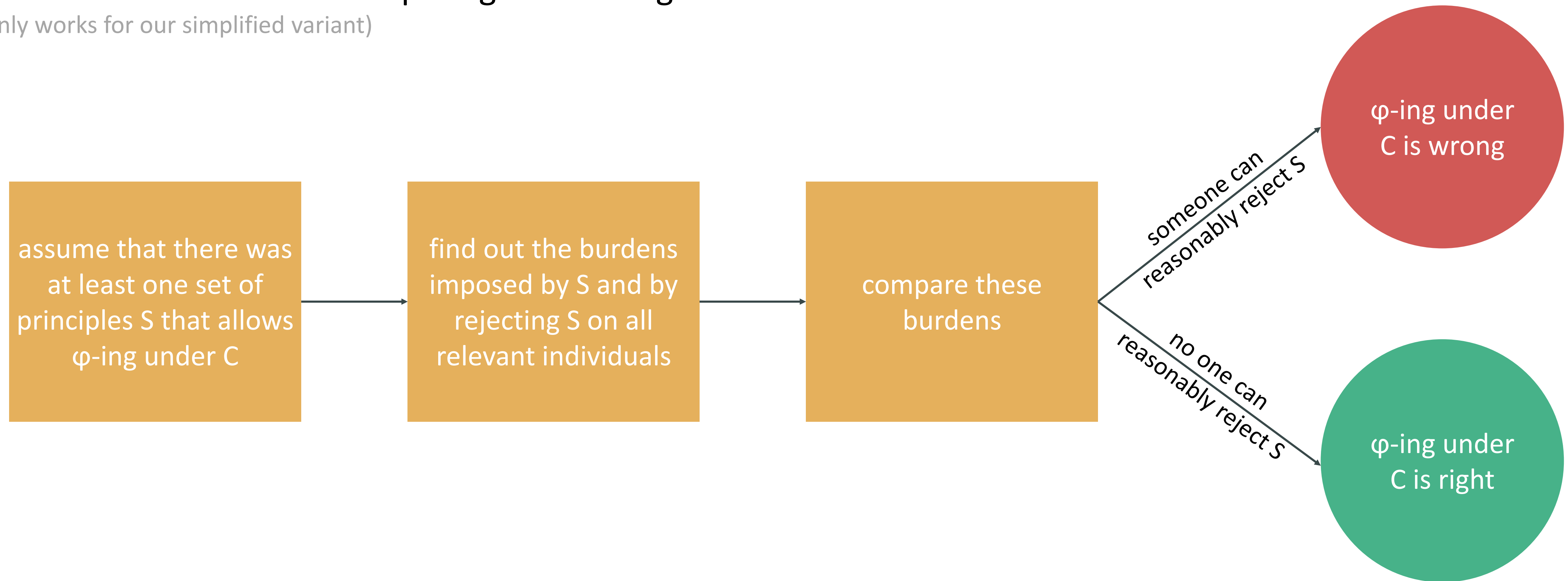
- if it imposes a burden on me
- and**
- if my rejecting it does not impose a much greater burden on someone else

**or**

- if the set distributes the burdens in a way that is unacceptable (because it, e.g., is sexist, racist, or otherwise unfair)

Find out whether an action  $\phi$  is right or wrong under circumstance C:

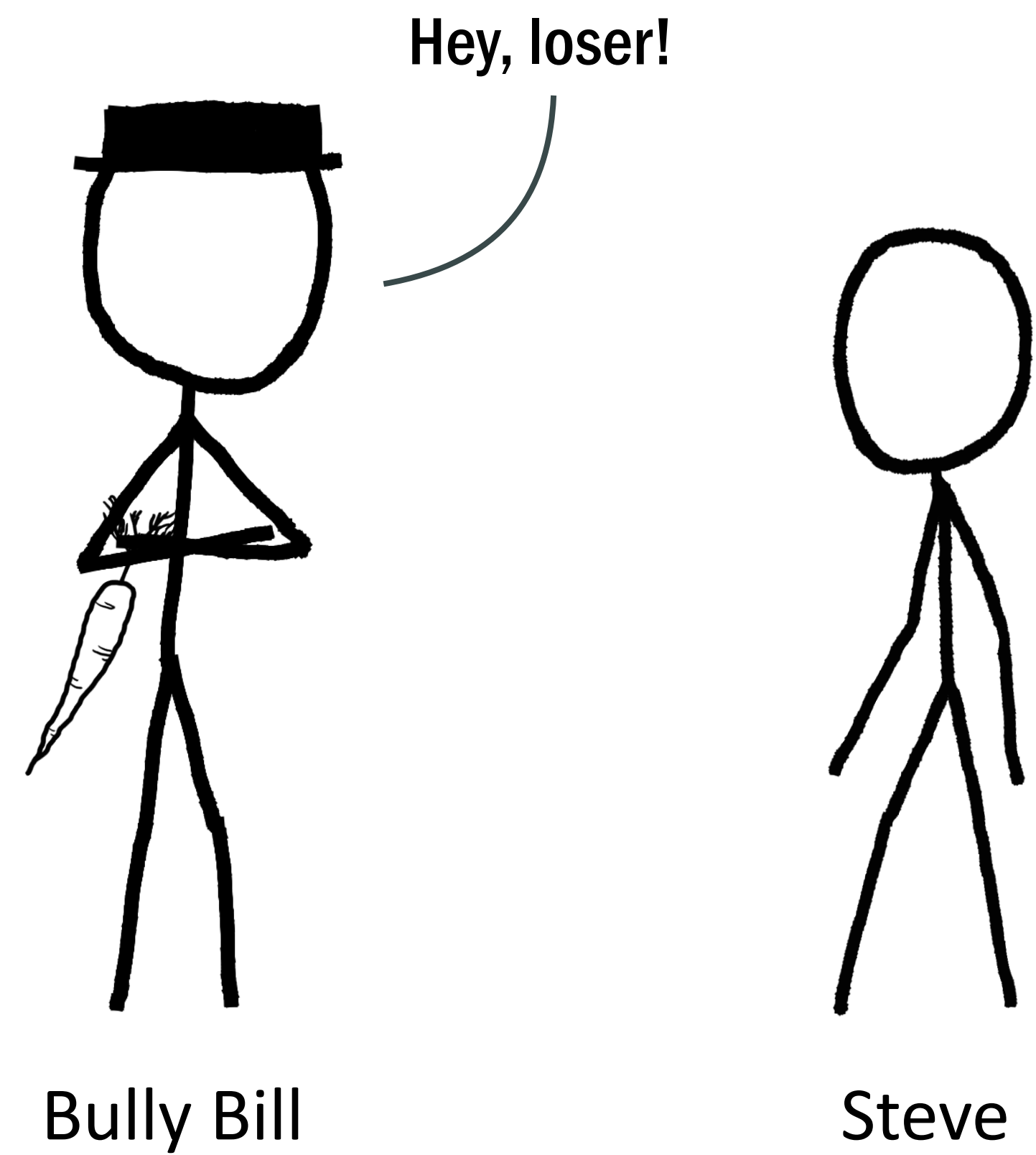
(only works for our simplified variant)





EXAMPLE

$\varphi$ : stealing someone’s lunch to bully them



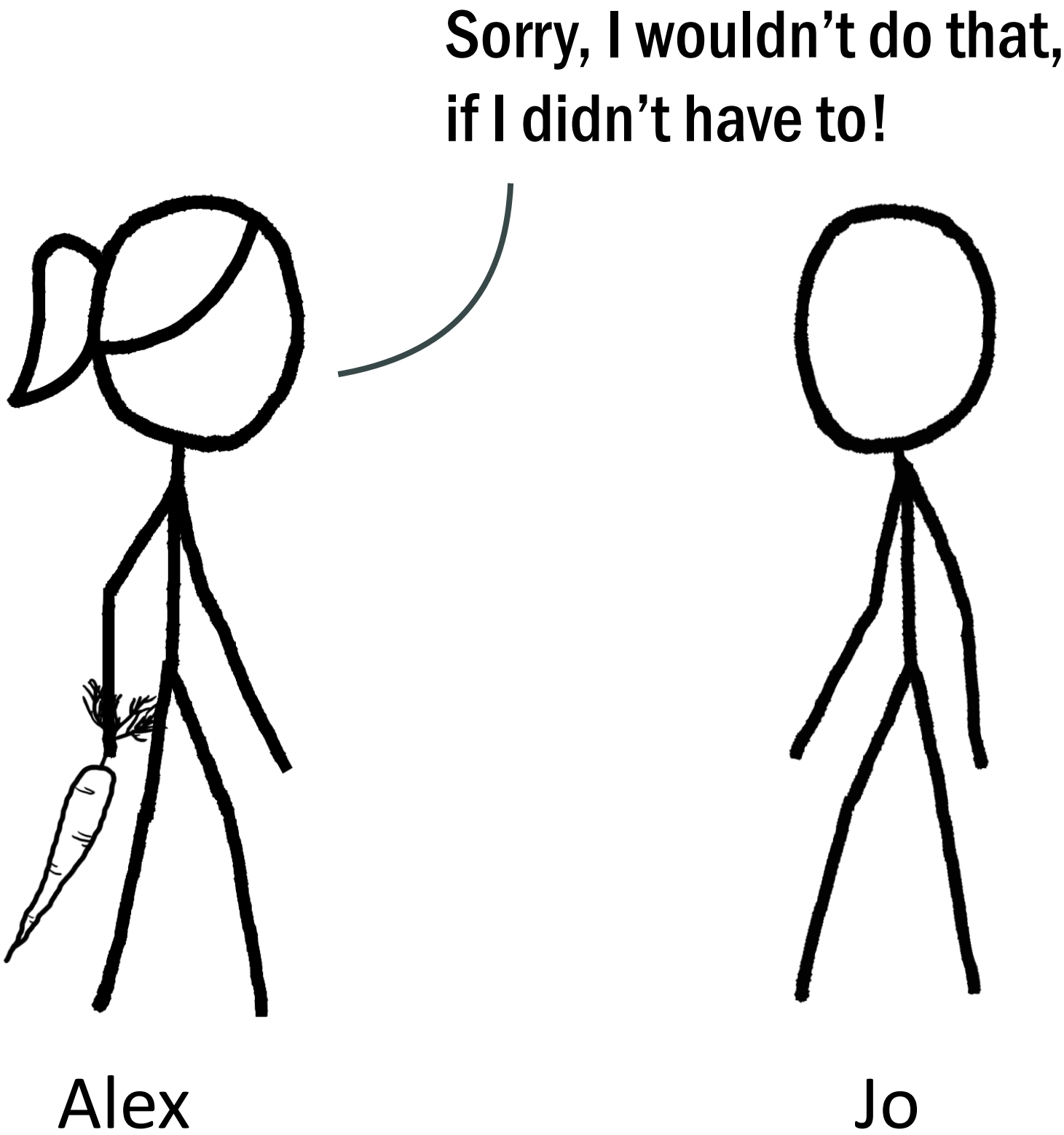
- 1. we assume that there was a set of principles S that allows Bill to steal Steve’s lunch
- 2. find out the burdens:

	Steve	Bill	relevant others?
burden that S places on...	large	none	other bullies and other people who are bullied
burden that Steve’s rejection of S places on...	none	small (at most)	(but everything works analogously so we do not have to discuss them separately)

- 3. compare the burdens: can Steve reasonably reject S? Yes!
- It is wrong for Bill to  $\varphi$ .

EXAMPLE

$\varphi$ : stealing someone’s lunch to in order to make sure that one does not starve



- 1. we assume that there was at least one set of principles S that allows Alex to steal Jo’s lunch
- 2. find out burdens:

	Jo	Alex	relevant others?
burden that S places on...	large	none	other starving people and other people who have food (but everything works analogously so we do not have to discuss them separately)
burden that Jo’s rejection of S places on...	none	extreme	

- 3. compare the burdens: can Jo reasonably reject S? No!  
→ It is right for Alex to  $\varphi$ .

## Caution!

The theory is about whether

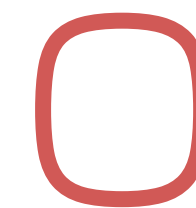
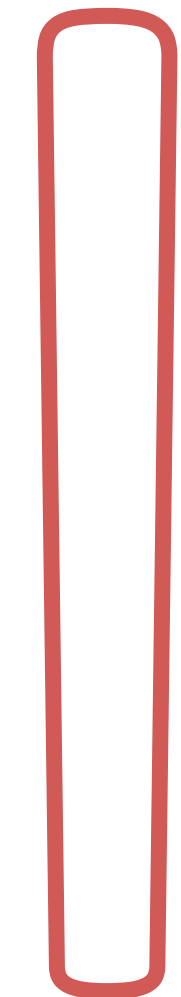
✓ someone *can reasonably* reject a *set* of principles.

This is not the same as whether

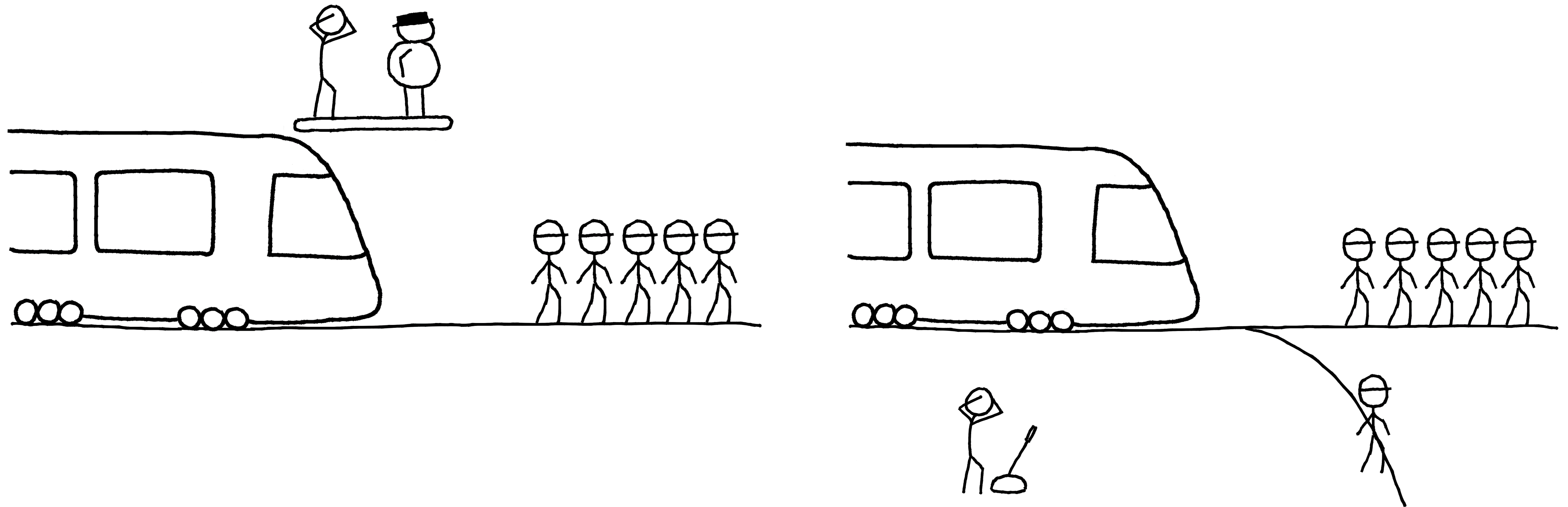
X someone does reasonably reject a set of principles,

X someone can reject a set of principles, or

X someone can reasonably reject a principle.



## EXAMPLE



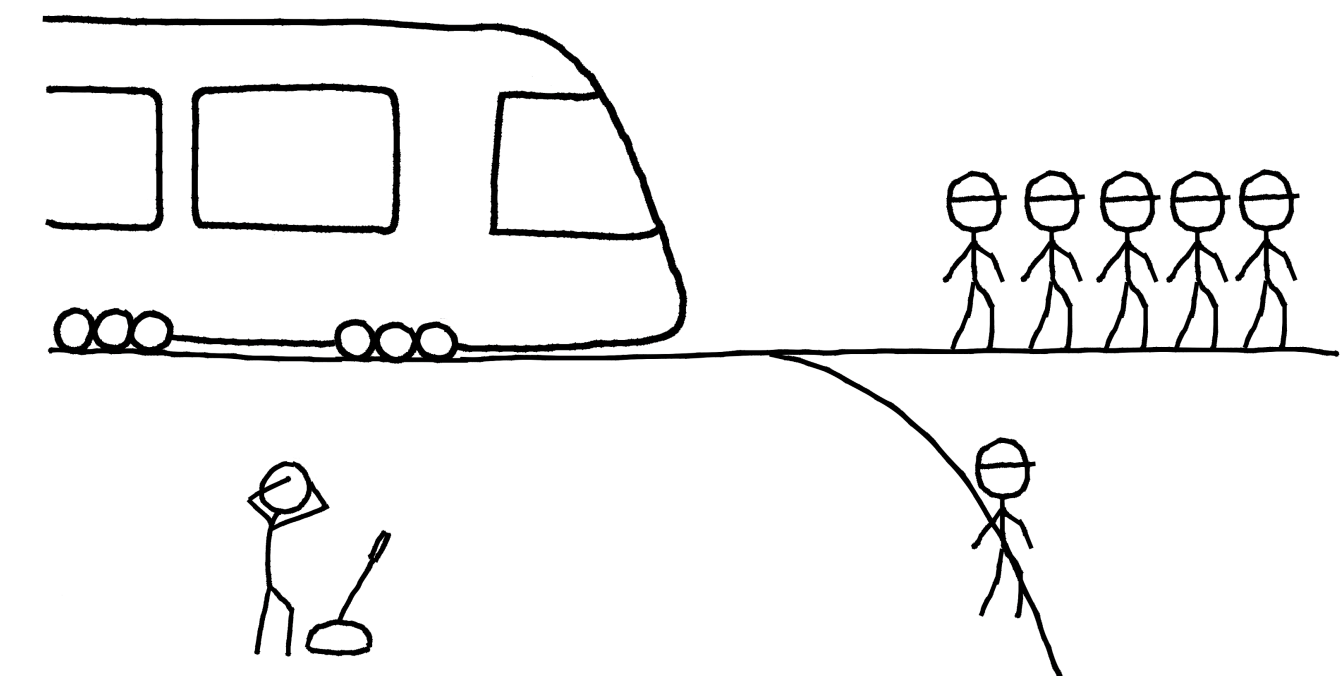
## Scanlon's Contractualism (simplified)

"An act is wrong if[f] its performance under the circumstances would be disallowed by [...all sets] of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced, general agreement."

from Scanlon, 1998: *What We Owe to Each Other*

**What if everybody has the same burden?**

→ Do a lottery, that is weighted with the number of people!



It is right to pull the lever with a probability of  $\frac{5}{6}$  and to not pull the lever with a probability of  $\frac{1}{6}$ .

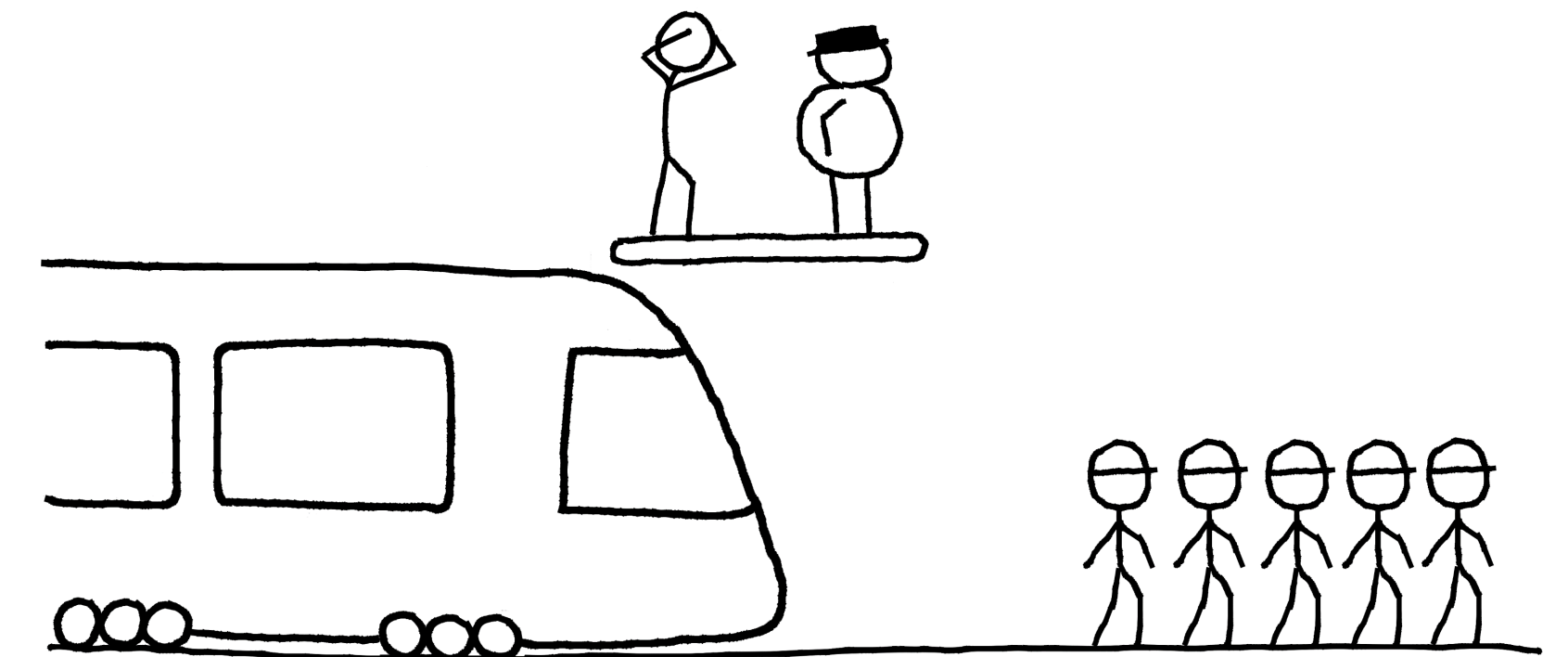
## EXAMPLE

### Scanlon's Contractualism (simplified)

"An act is wrong if[f] its performance under the circumstances would be disallowed by [...all sets] of principles for the general regulation of behaviour that no one could reasonably reject as a basis for informed, unforced, general agreement."

from Scanlon, 1998: *What We Owe to Each Other*

Can the fat man reasonably reject?



Does pushing the fat man place a significant additional burden on him?  
(e.g. by disrespecting him)

yes

pushing  
is wrong

no

we need  
a lottery  
again



# PROBLEMS

Ashford, Elizabeth and Mulgan, Tim, "Contractualism", *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.),  
<https://plato.stanford.edu/archives/sum2018/entries/contractualism/>

over-  
demandingness?

dictatorship of  
the one who is  
worst off?

redundant?

ignores  
animals and  
future people?

'too  
tidy'/one  
thought to  
many?

...

lottery  
counterintuitive?

