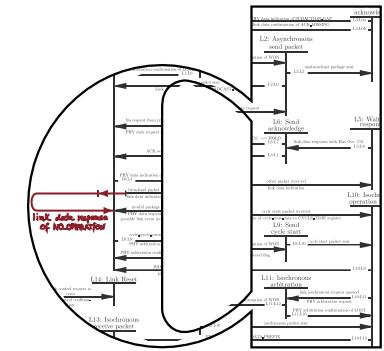




Ethics for Nerds

An Advanced Course in Computer Science
Summer Semester 2020

Ethics 2
Normative Ethics and Moral Theories



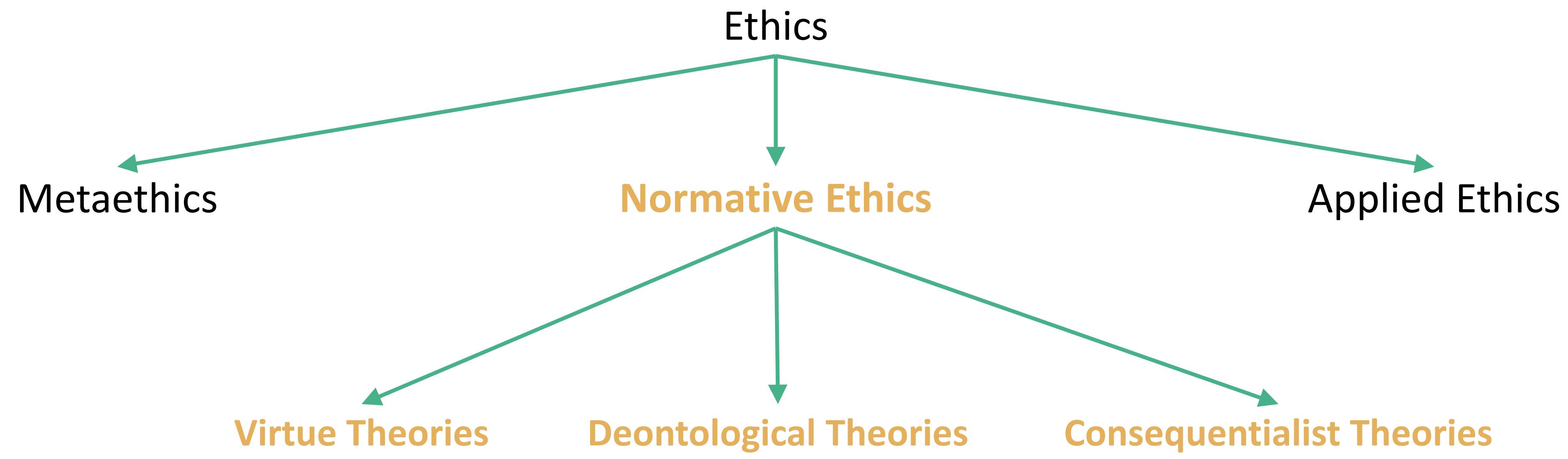
Prof. Holger Hermanns,
Kevin Baum, Sarah Sterz

THE CENTRAL QUESTION

What is right?

(according to different views)

FIELDS OF ETHICS



A Tweet to Arms



Donald J. Trump
@realDonaldTrump

LIBERATE MICHIGAN!

5:22 nachm. · 17. Apr. 2020 · Twitter for iPhone

42.921 Retweets **211.841 „Gefällt mir“-Angaben**

<https://twitter.com/realDonaldTrump/status/1251169217531056130>



Senator Dayna Polehanki
@SenPolehanki

Directly above me, men with rifles yelling at us. Some of my colleagues who own bullet proof vests are wearing them. I have never appreciated our Sergeants-at-Arms more than today. #mileg



6:38 nachm. · 30. Apr. 2020 aus Lansing, MI · Twitter for iPhone

15.231 Retweets **47.695 „Gefällt mir“-Angaben**

<https://twitter.com/SenPolehanki/status/1255899318210314241>

A Tweet to Arms



Trump tweets “LIBERATE MICHIGAN”.

Armed protestors enter Michigan Statehouse.

If Trump did not tweet that,
there would not have been armed protestors.

More on the story: <https://www.bbc.com/news/world-us-canada-52496514>

A Tweet to Arms



Trump tweets “LIBERATE MICHIGAN”.

↑
Agent
Armed protestors enter Michigan Statehouse.

If Trump did not tweet that,
there would not have been armed protestors.

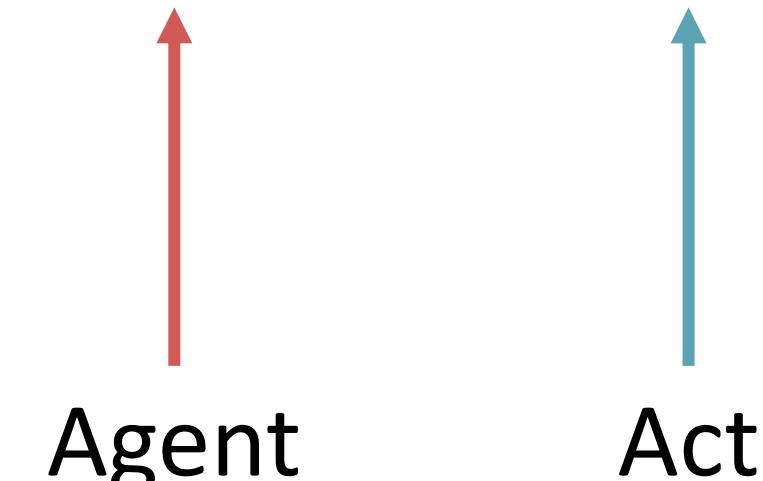
More on the story: <https://www.bbc.com/news/world-us-canada-52496514>

A Tweet to Arms



Trump tweets “LIBERATE MICHIGAN”.

Agent

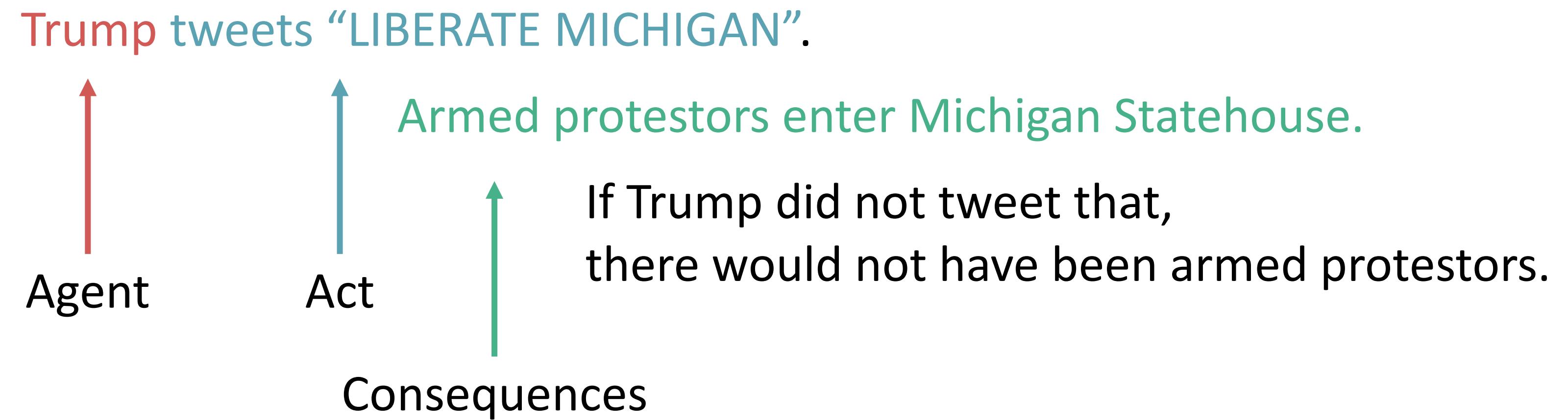


Armed protestors enter Michigan Statehouse.

If Trump did not tweet that,
there would not have been armed protestors.

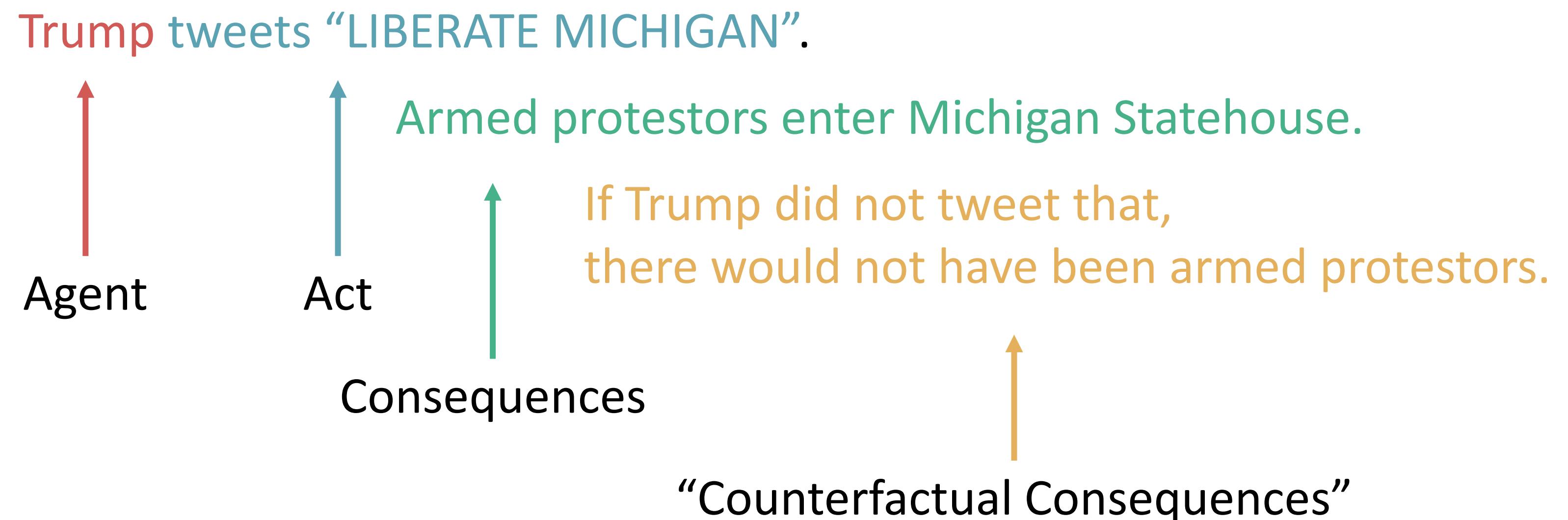
More on the story: <https://www.bbc.com/news/world-us-canada-52496514>

A Tweet to Arms



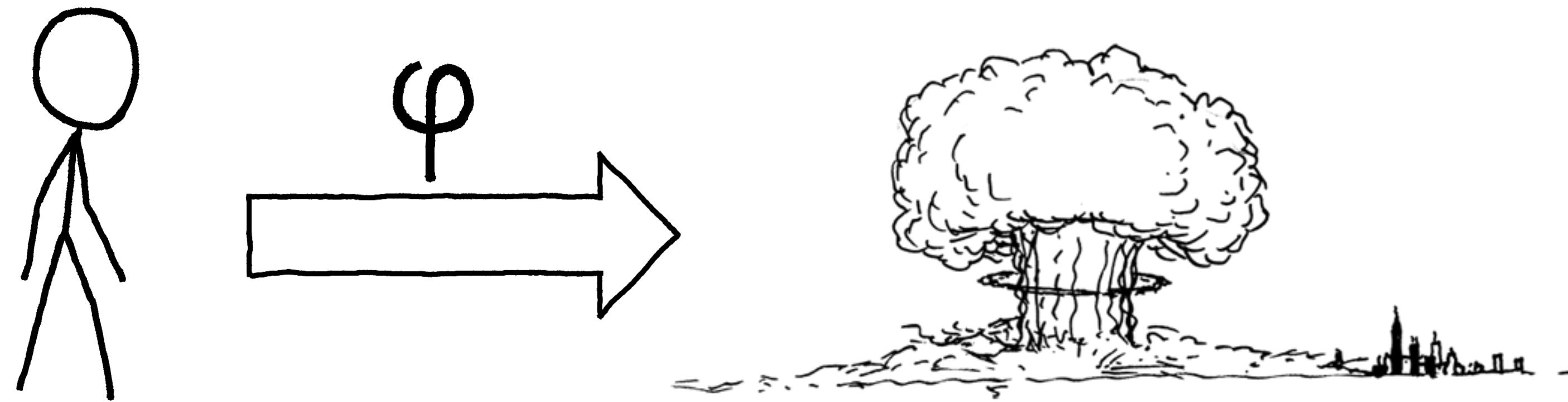
More on the story: <https://www.bbc.com/news/world-us-canada-52496514>

A Tweet to Arms



More on the story: <https://www.bbc.com/news/world-us-canada-52496514>

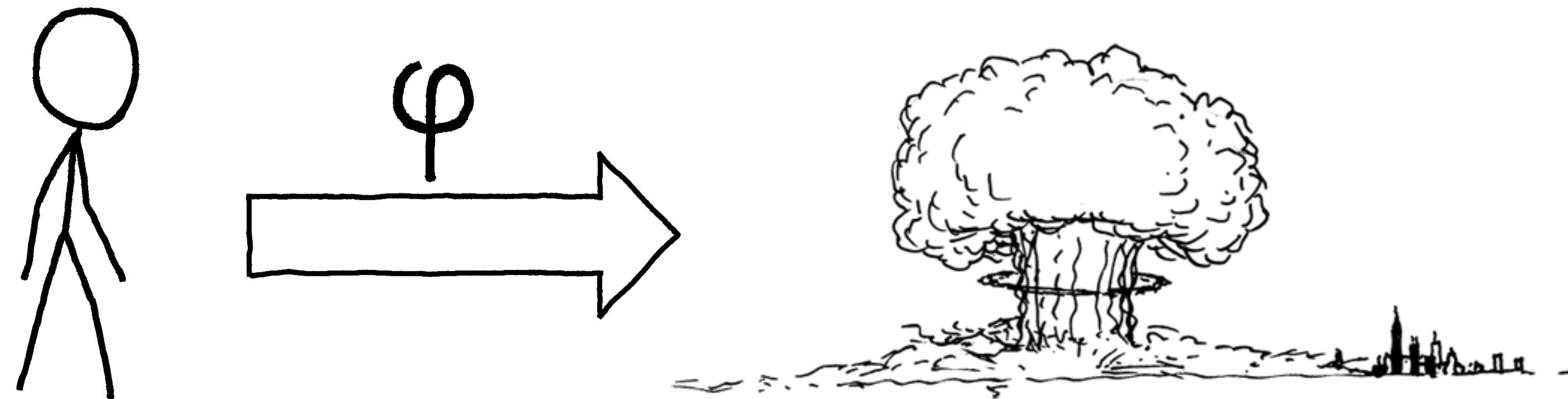
A Classical Description



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

Where to look for the answer to the question:
What is the right (or wrong) thing to do and why?

A Classical Description



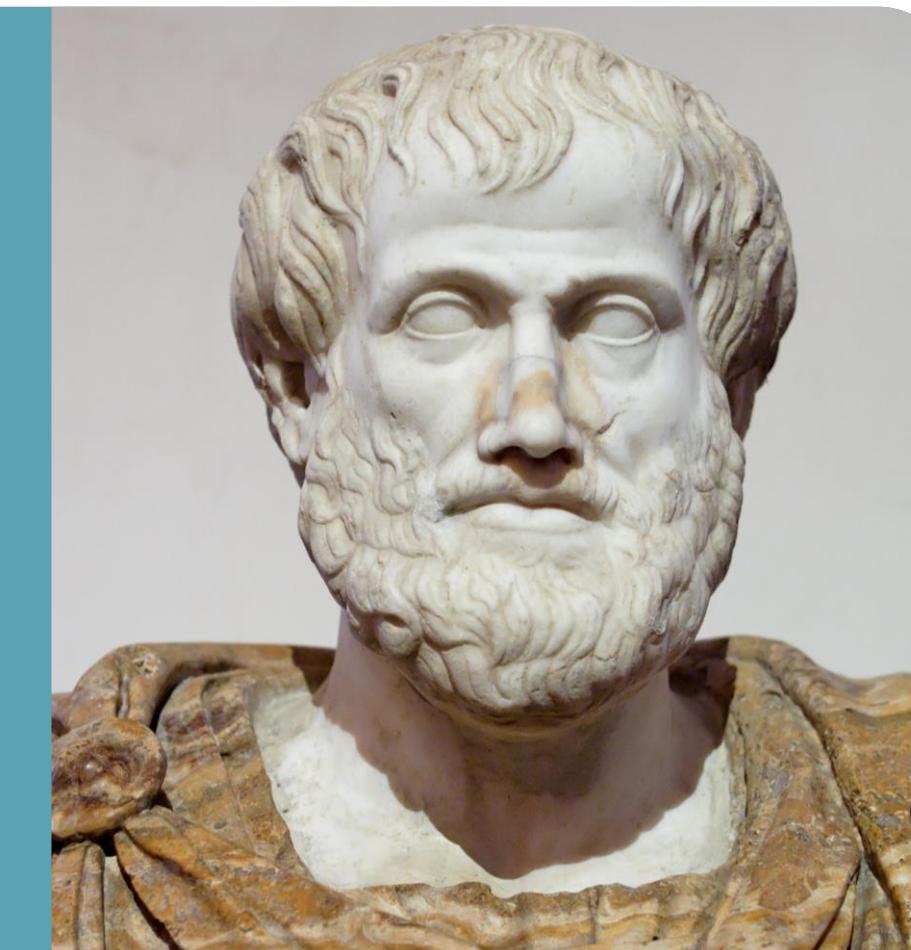
An **agent** performs an **act**. An **act** has **consequences**.

Virtue Theories

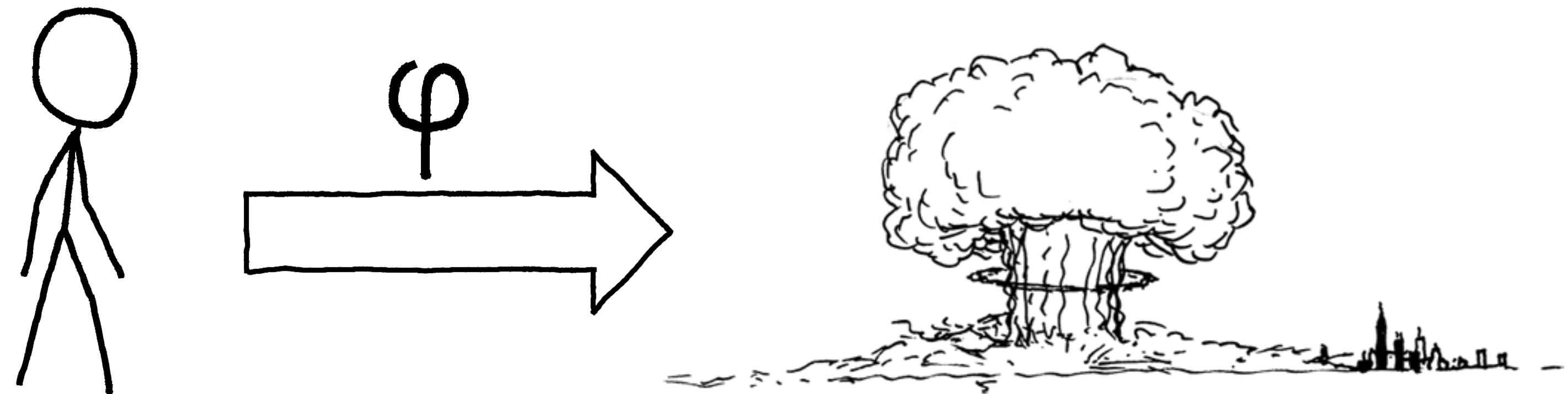
most famously: Aristotle

384–322 BC

Nicomachean Ethics



A Classical Description



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

Deontological Theories

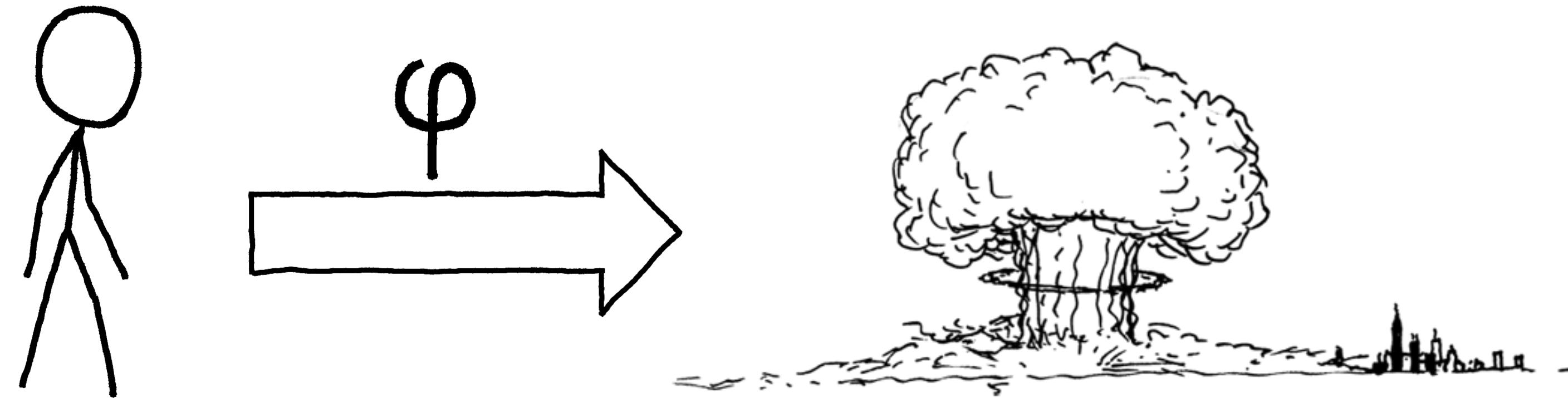
most famously: Immanuel Kant

1724–1804

Grundlegung zur Metaphysik der Sitten



A Classical Description



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

Consequentialist Theories

most famously: John Stuart Mill

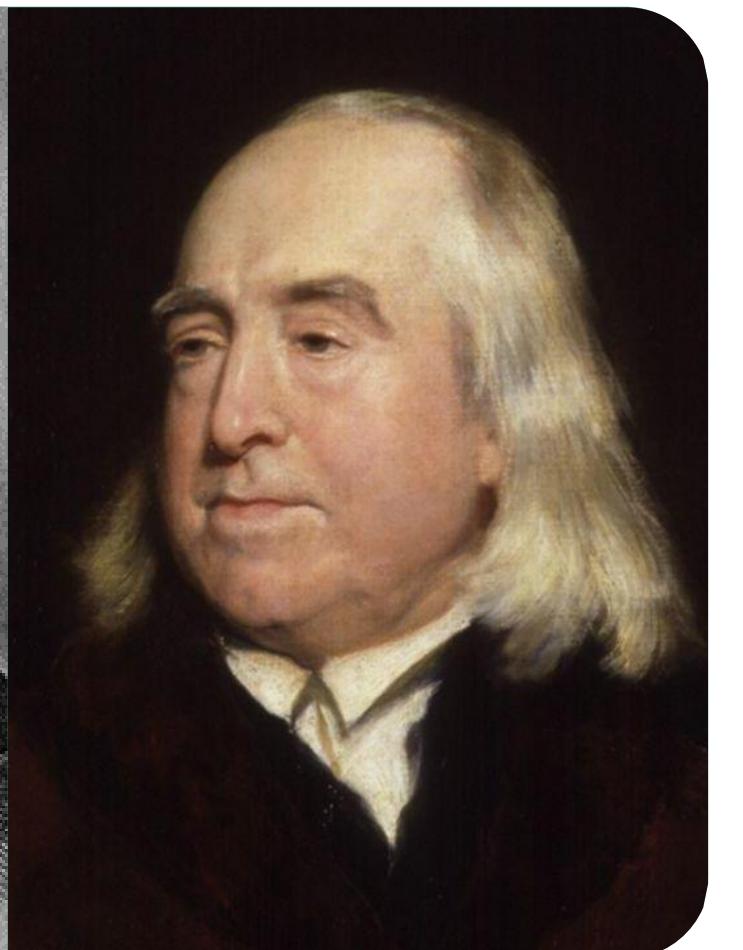
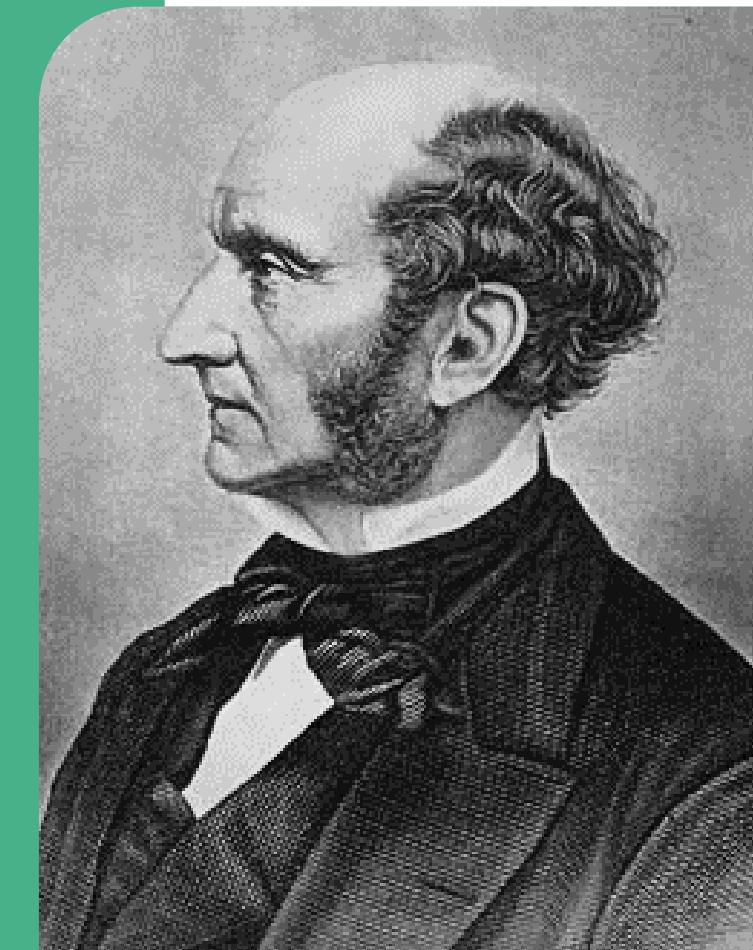
1806 – 1873

Utilitarianism

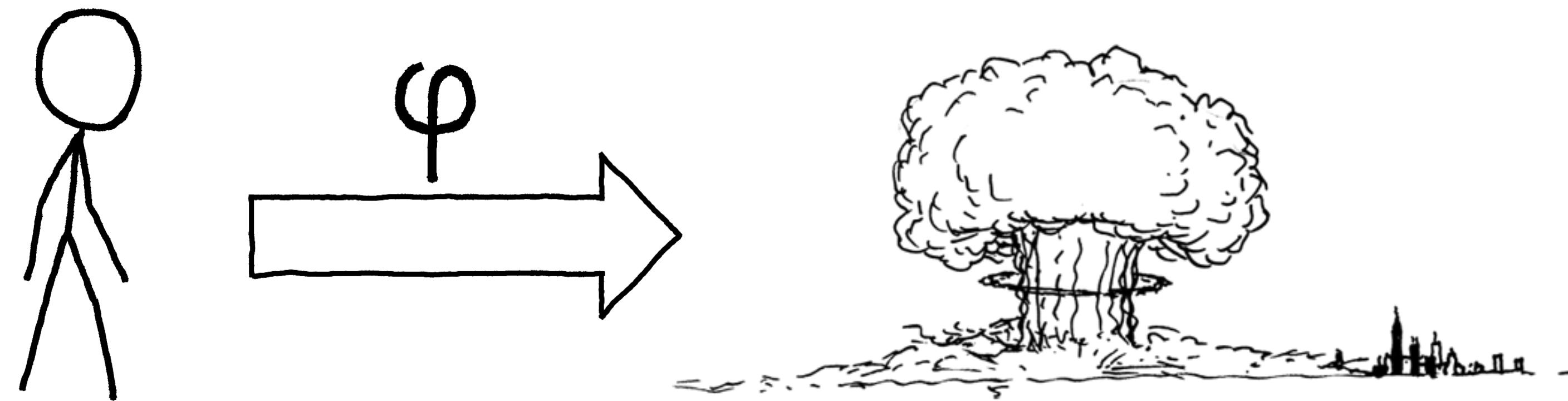
Jeremy Bentham

1748–1832

*An Introduction to the Principles
of Morals and Legislation*



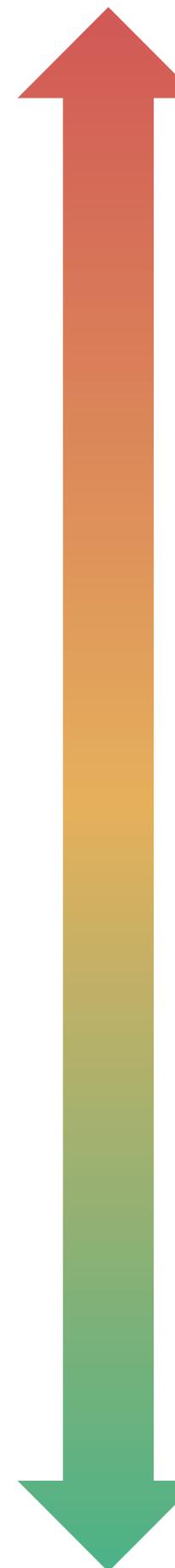
A Classical Description



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

What qualifies as an (moral) agent? And who matters morally?

nonsense



The **hammer** fell off the board and onto Niko's foot.
What a bad hammer! The hammer shouldn't have done this.

The **fridge** kept the beverages cold.
This is good, as the fridge had the obligation to do so.

My **goldfish/hamster/cat/dog** bit me.
This was morally impermissible!

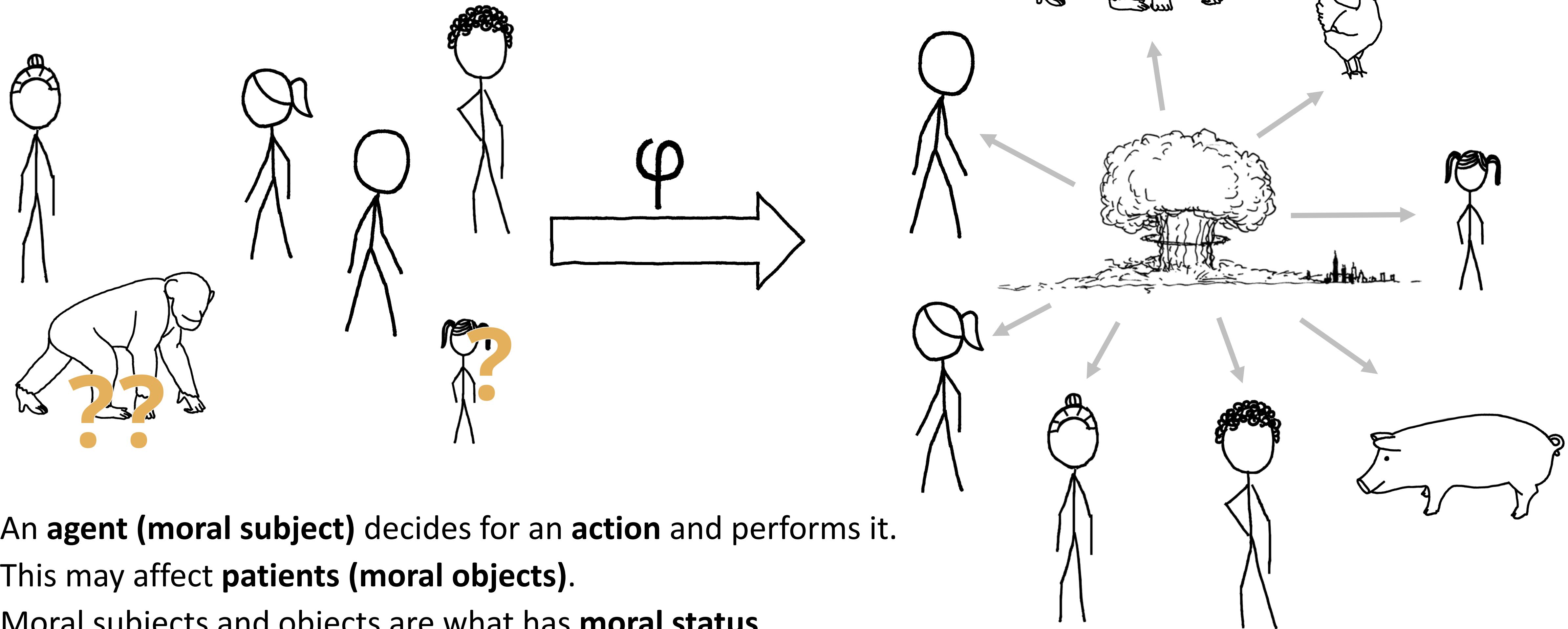
This **toddler** deliberately broke his plate when he saw that we are having spinach today.
He ought not to act like this!

These **8th graders** beat up Tim.
That was clearly wrong.

Donald cheated on his wife.
How immoral of him!

standard way of
talking

THE SCOPE OF MORALITY: MORAL AGENTS AND PATIENTS





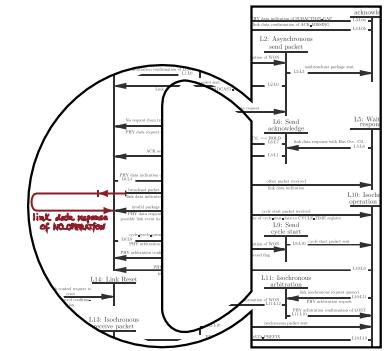


Ethics for Nerds

An Advanced Course in Computer Science
Summer Semester 2020

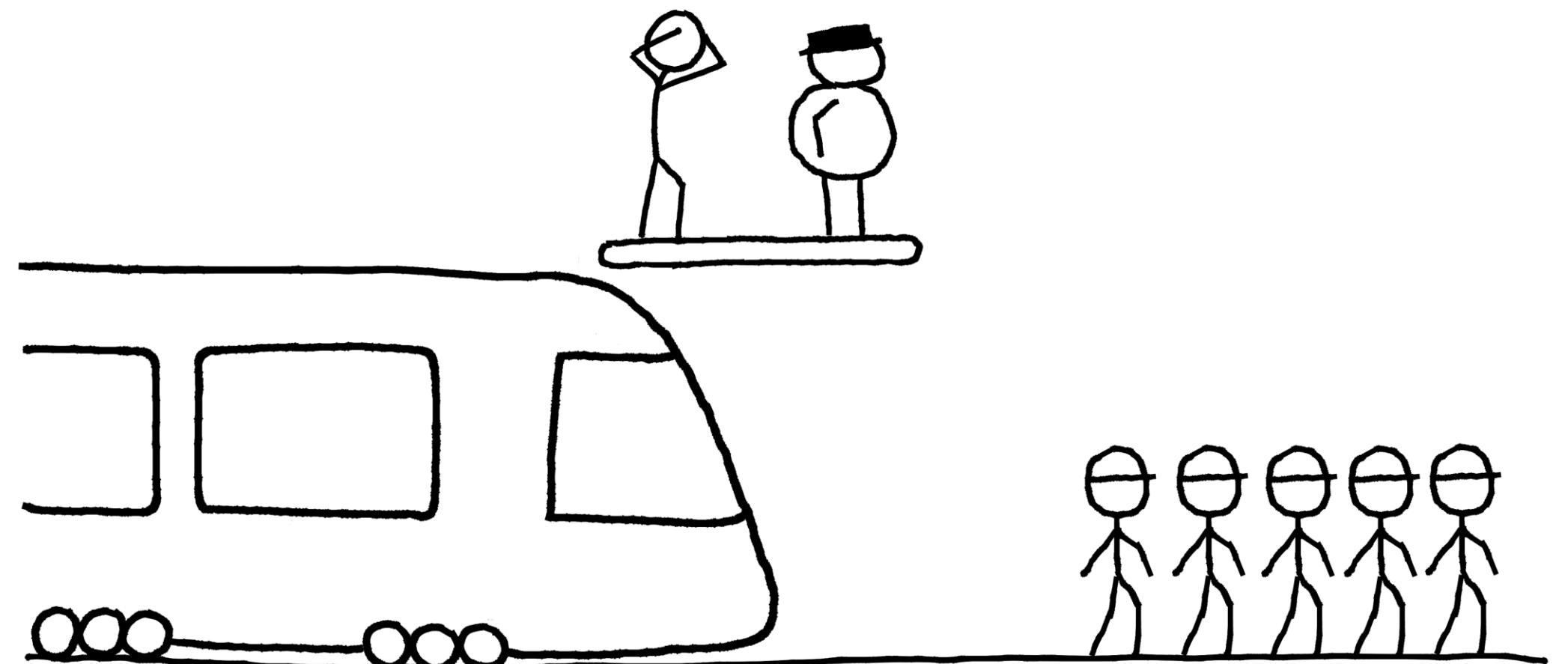
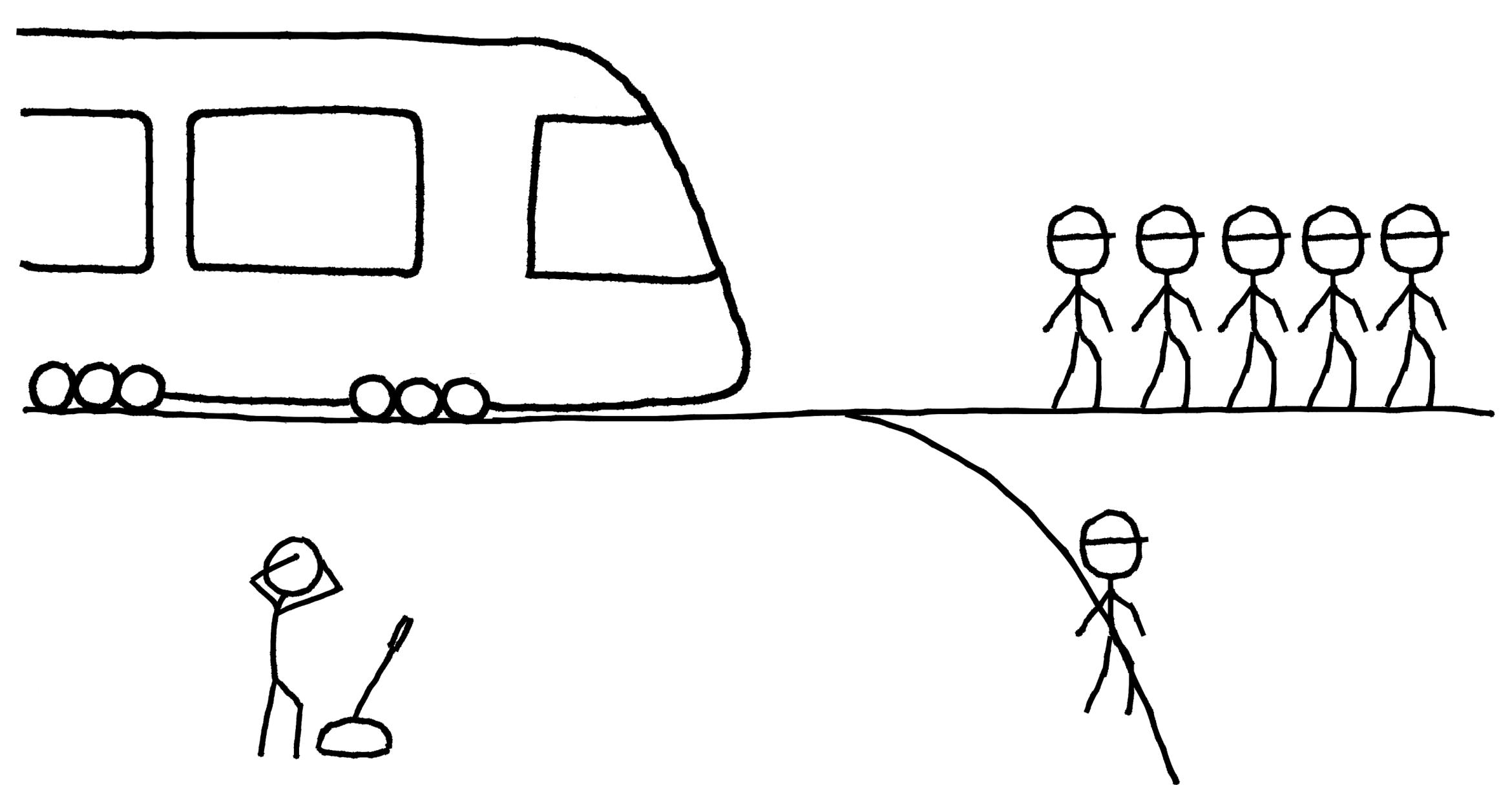
Ethics 3.1
Consequentialism

The Consequentialist Framework



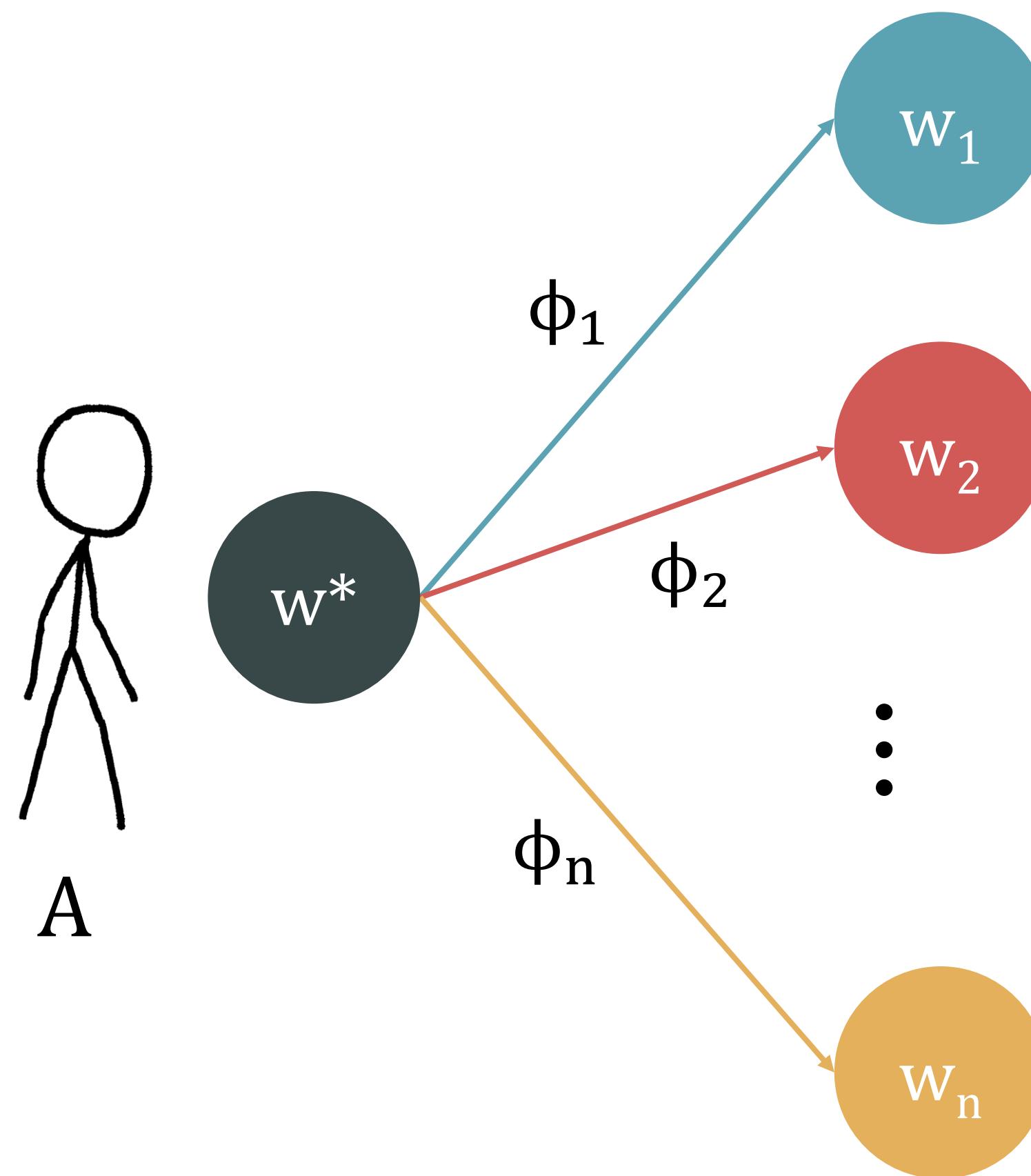
Prof. Holger Hermanns,
Kevin Baum, Sarah Sterz

INTUITION

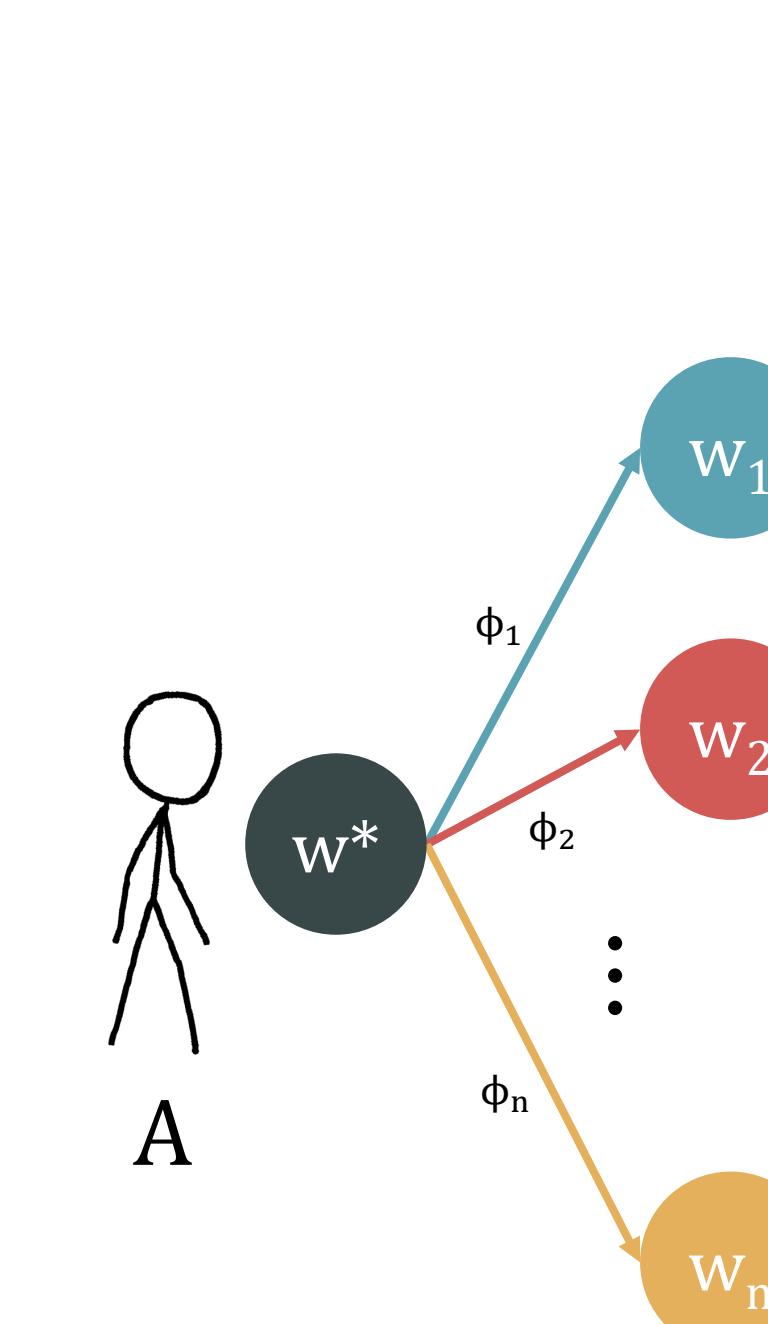


DECISION SITUATIONS

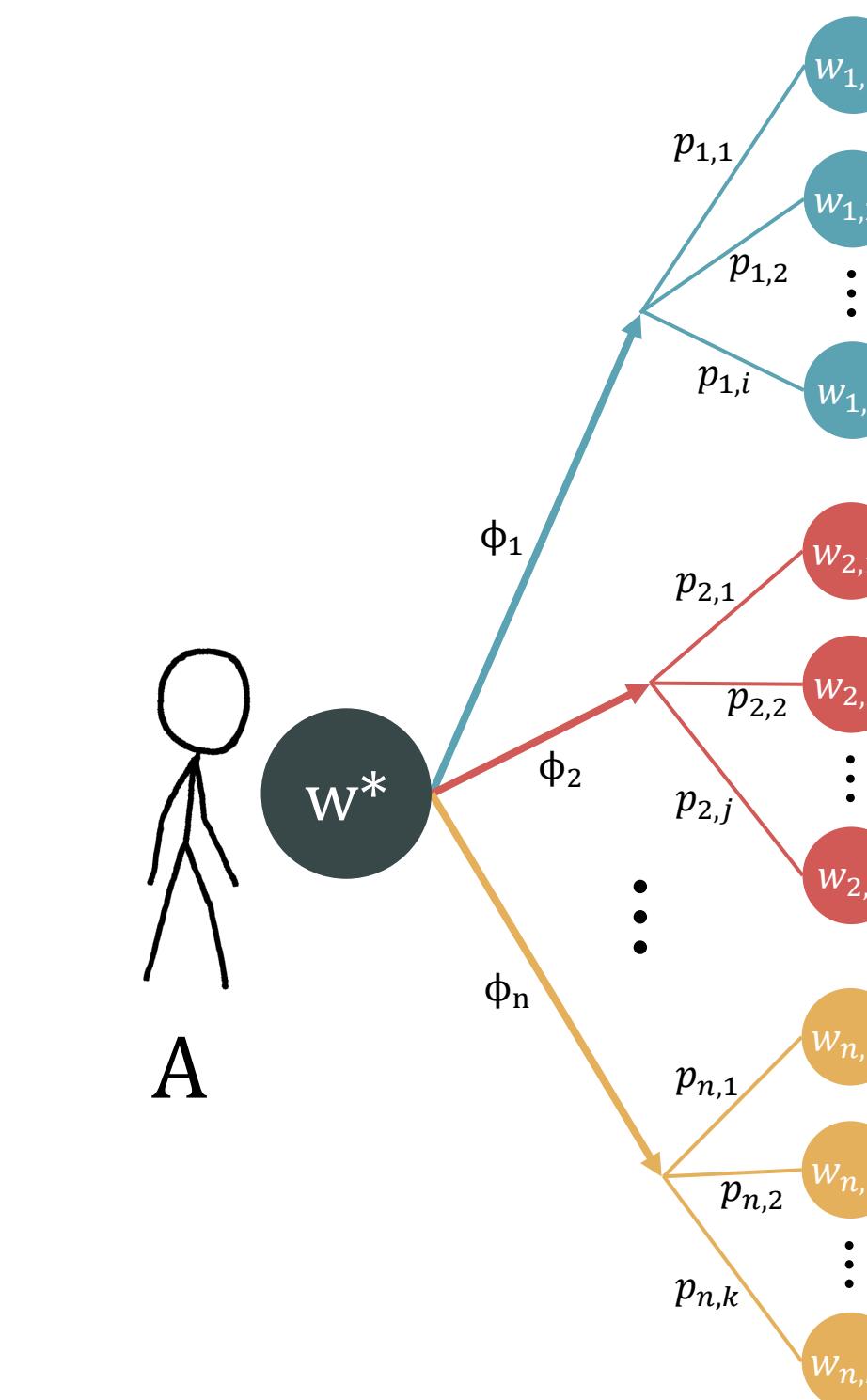
Usual model from an objective perspective:



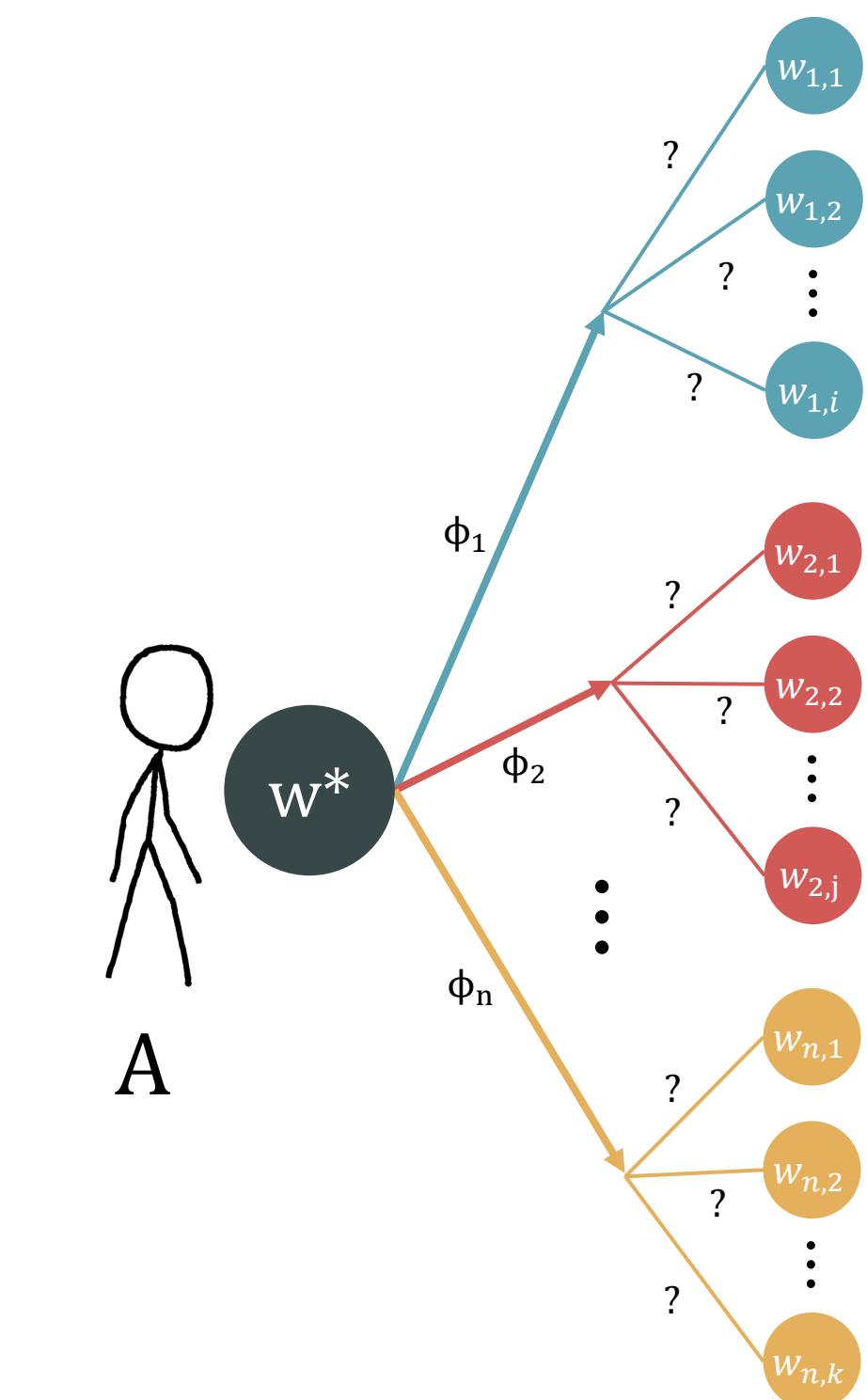
Usual models from a subjective perspective:



under certainty



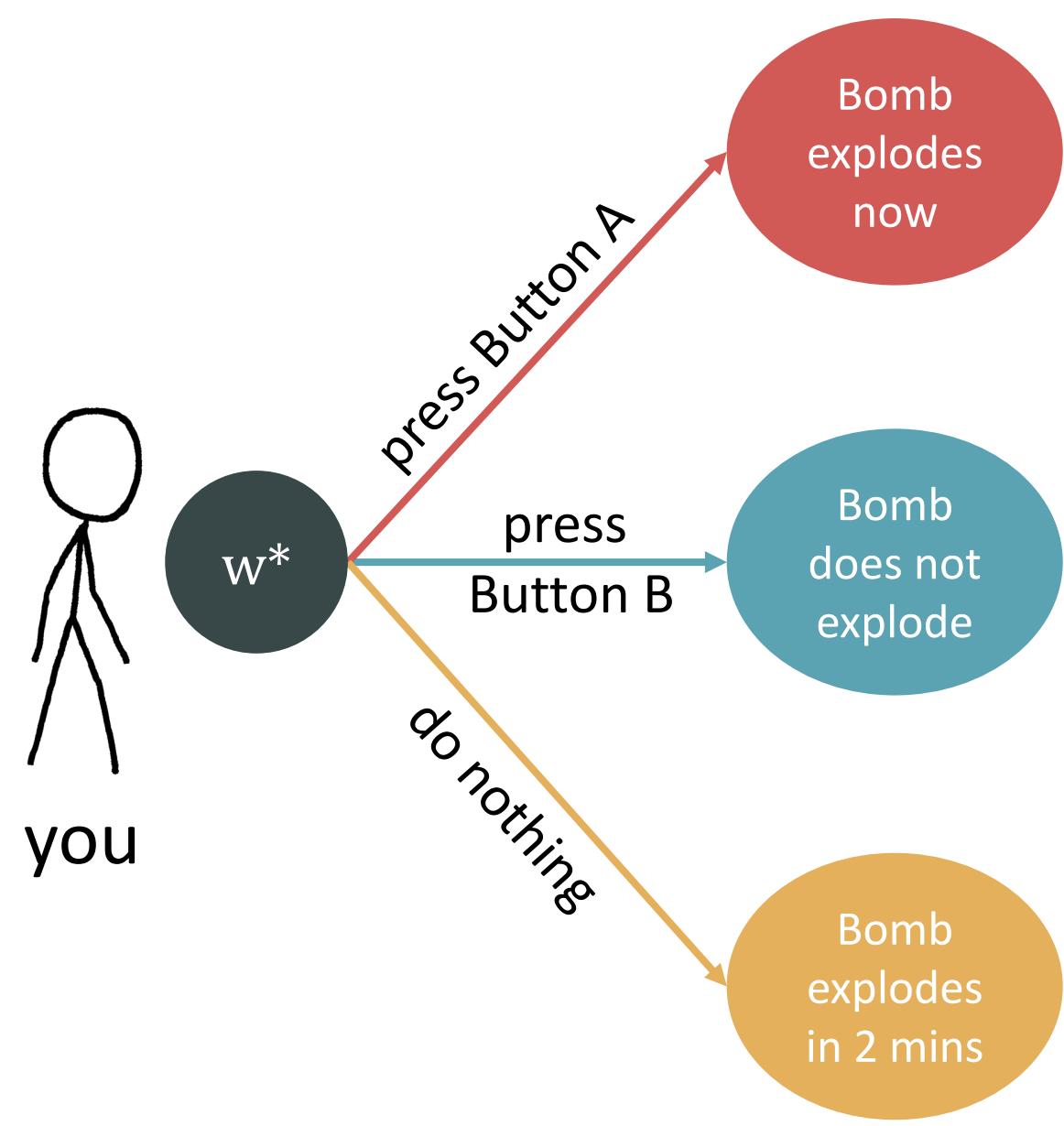
under risk



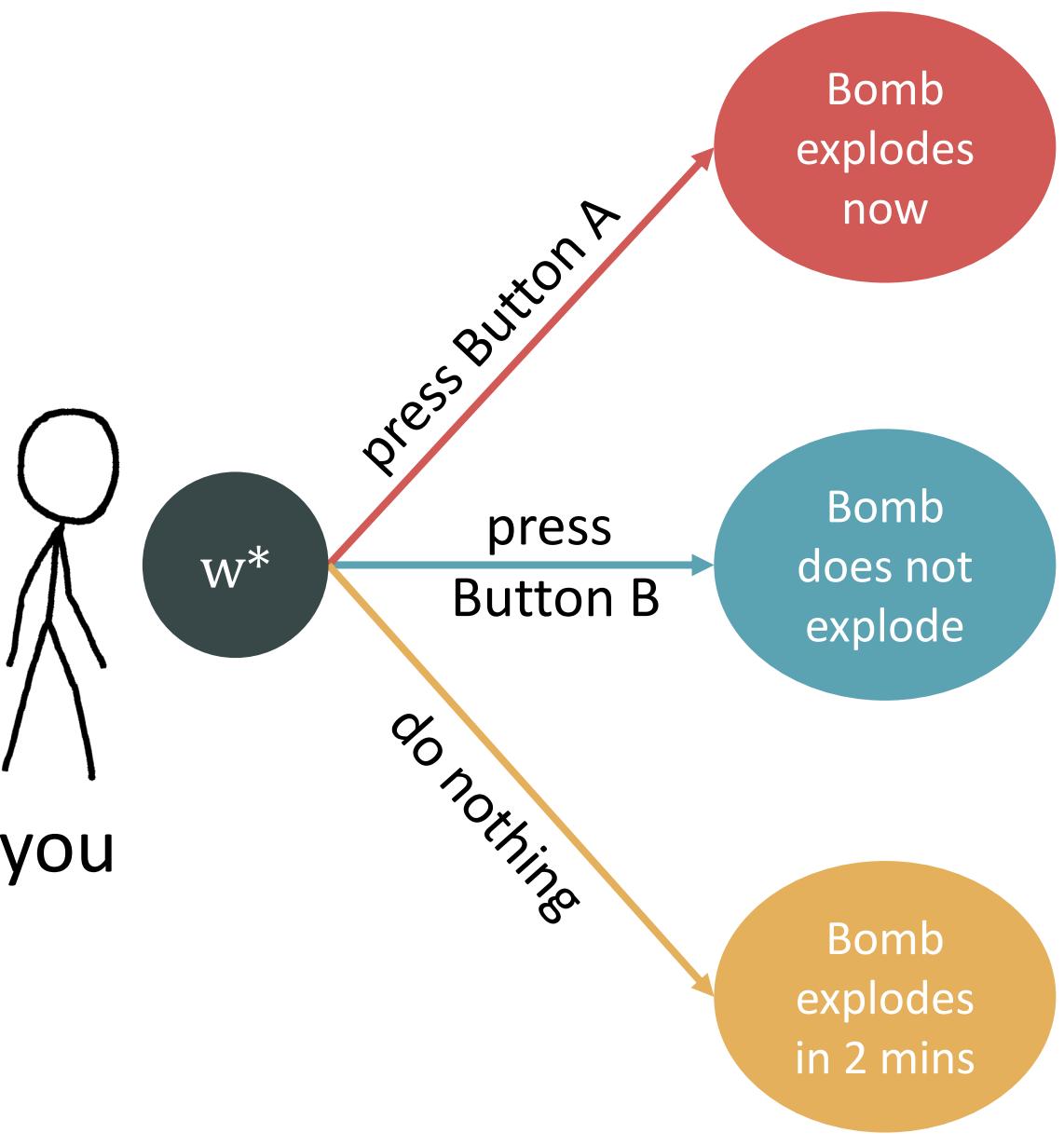
under uncertainty

DECISION SITUATIONS

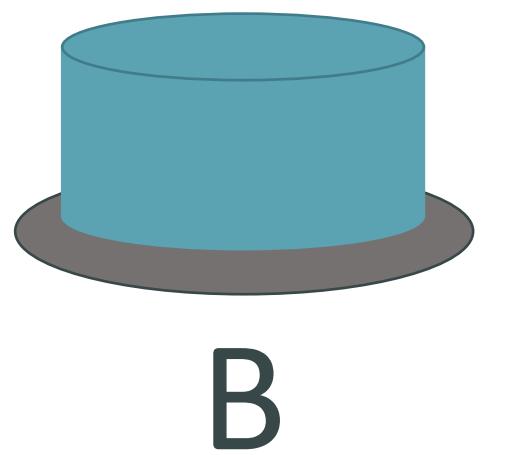
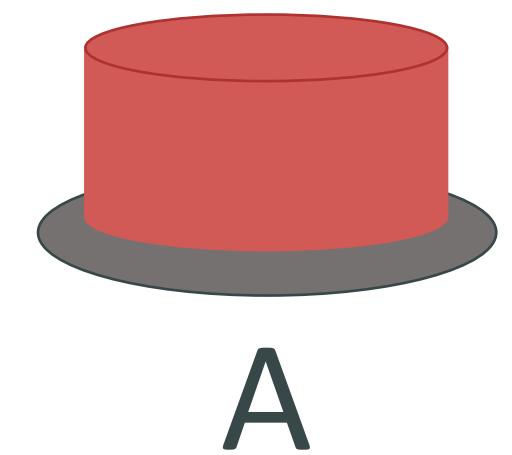
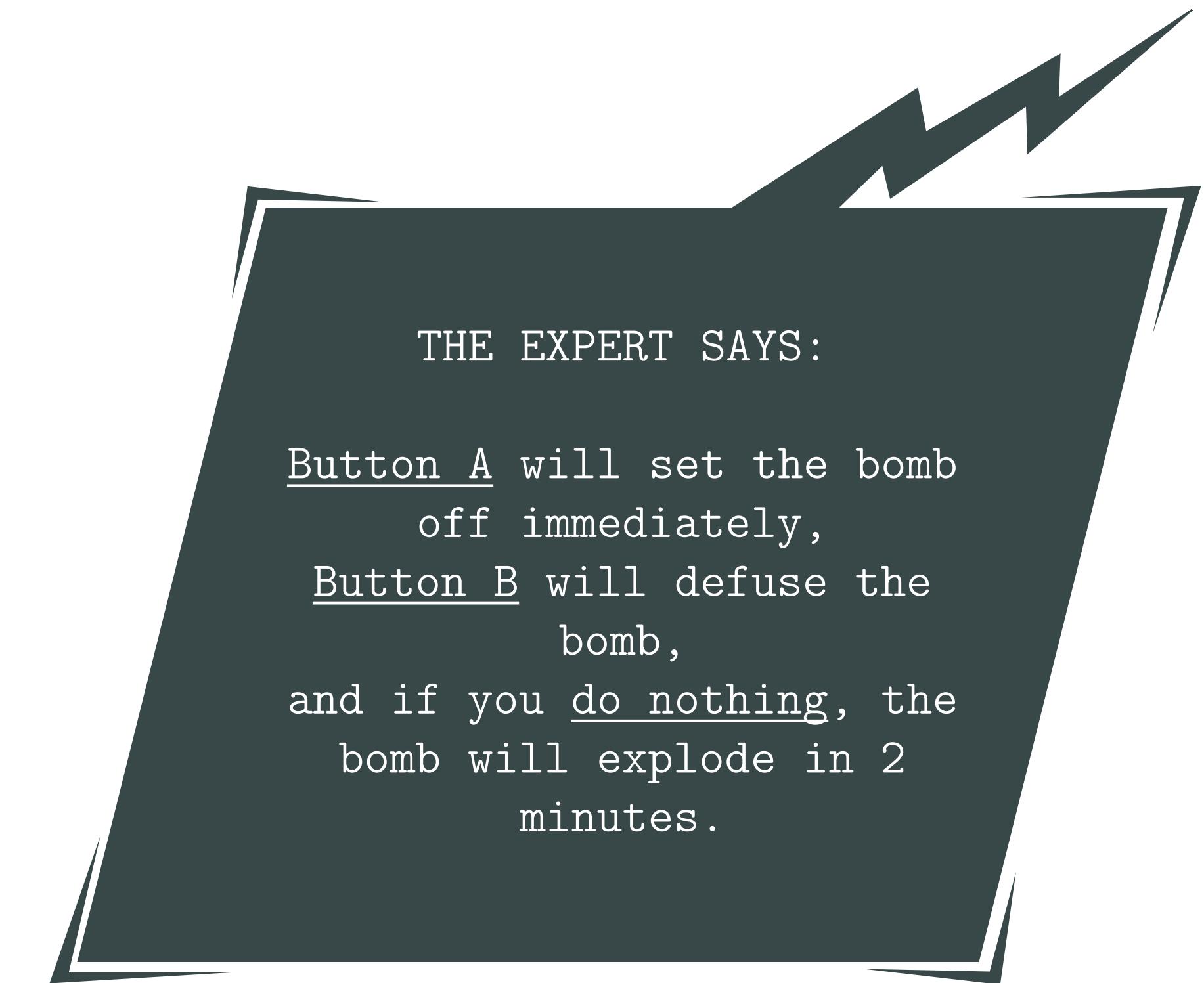
You have to defuse a bomb with two buttons.
You know for sure that the expert is right.



objective decision situation



subjective decision situation



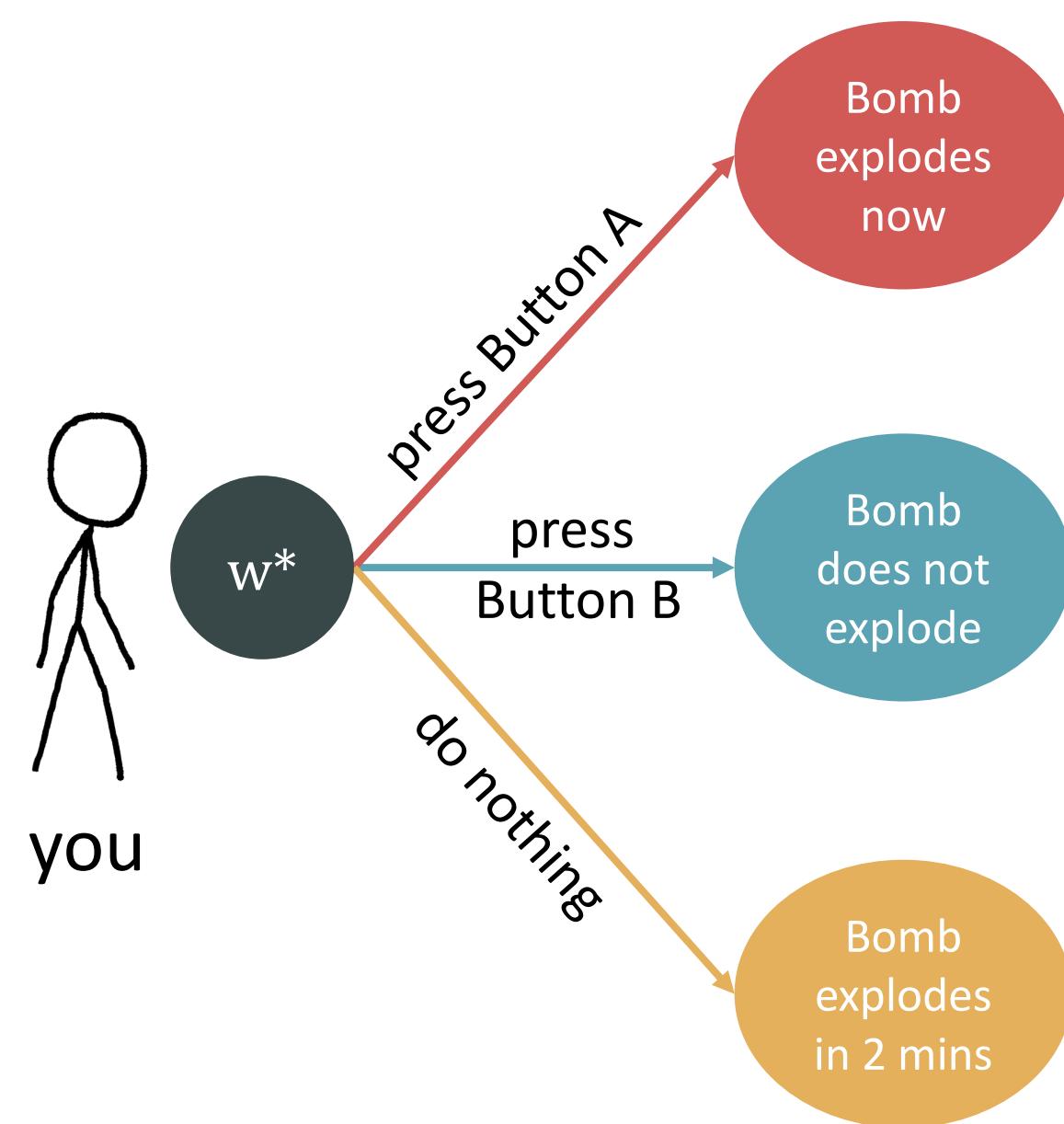
A

B

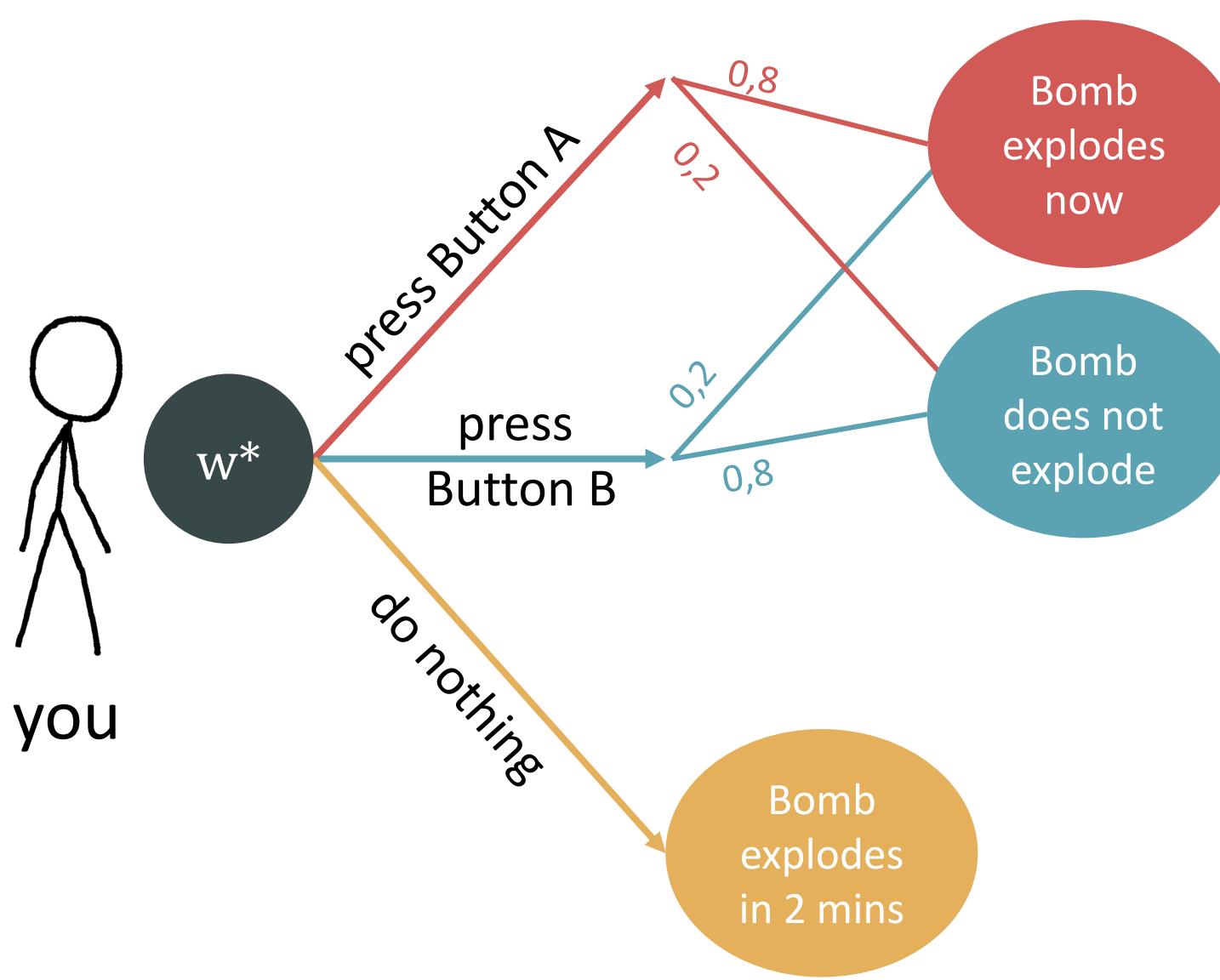
DECISION SITUATIONS

You have to defuse a bomb with two buttons.

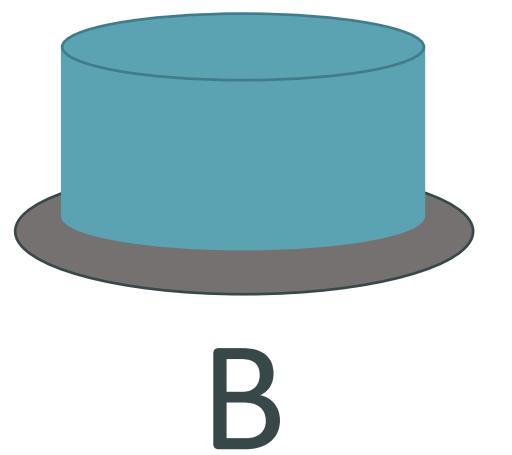
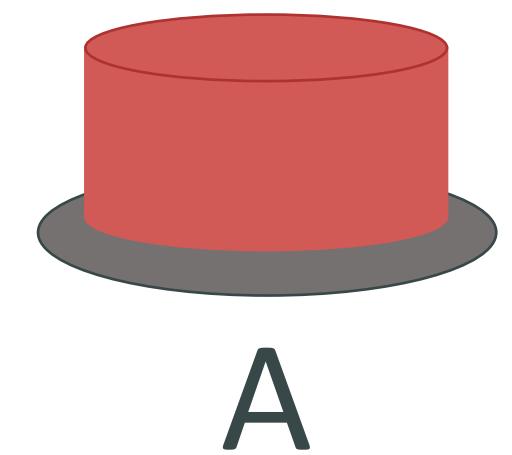
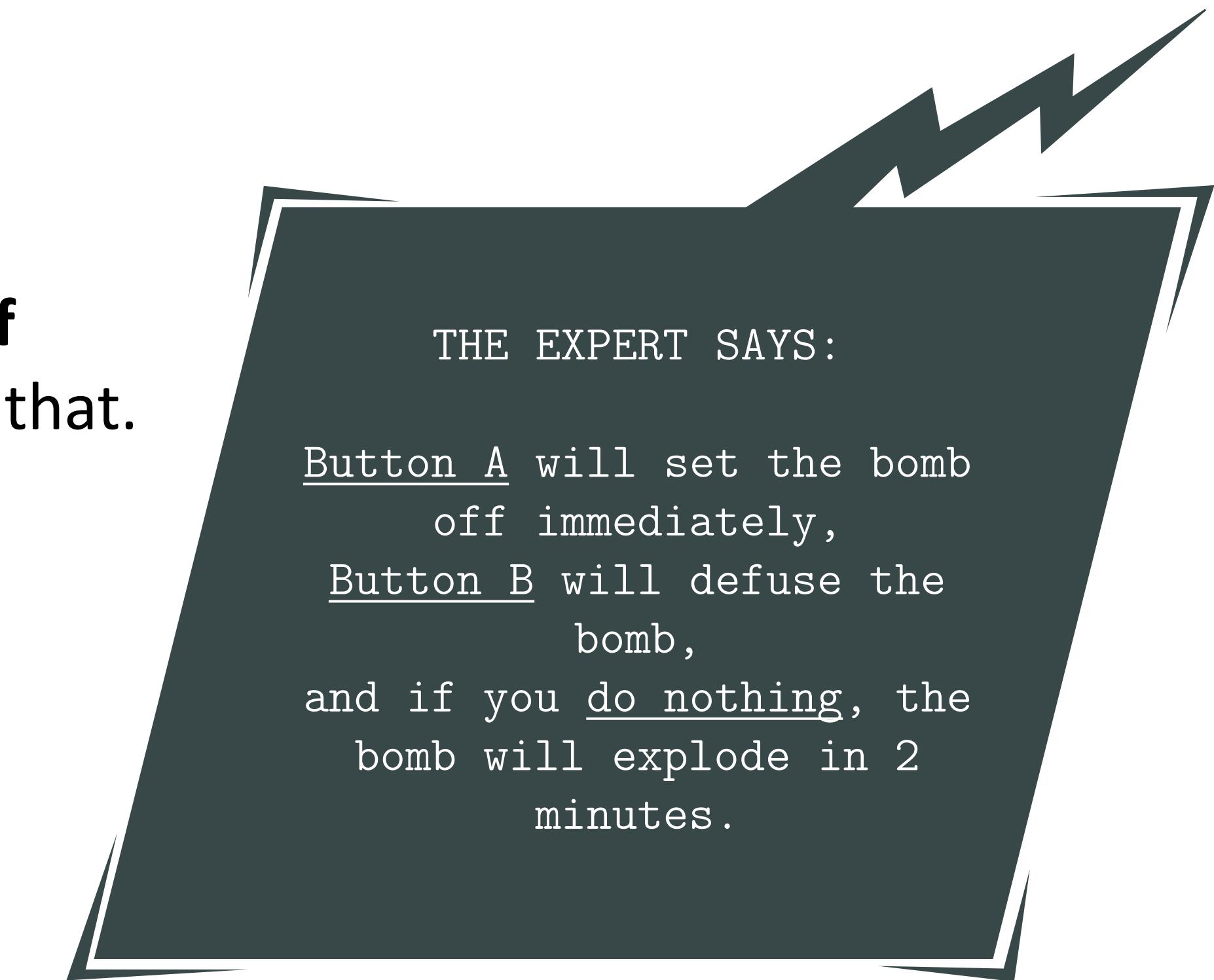
You know that the **expert gets wrong what the buttons do for 20% of comparable cases**. This time, the expert is right but you do not know that.



objective decision situation



subjective decision situation



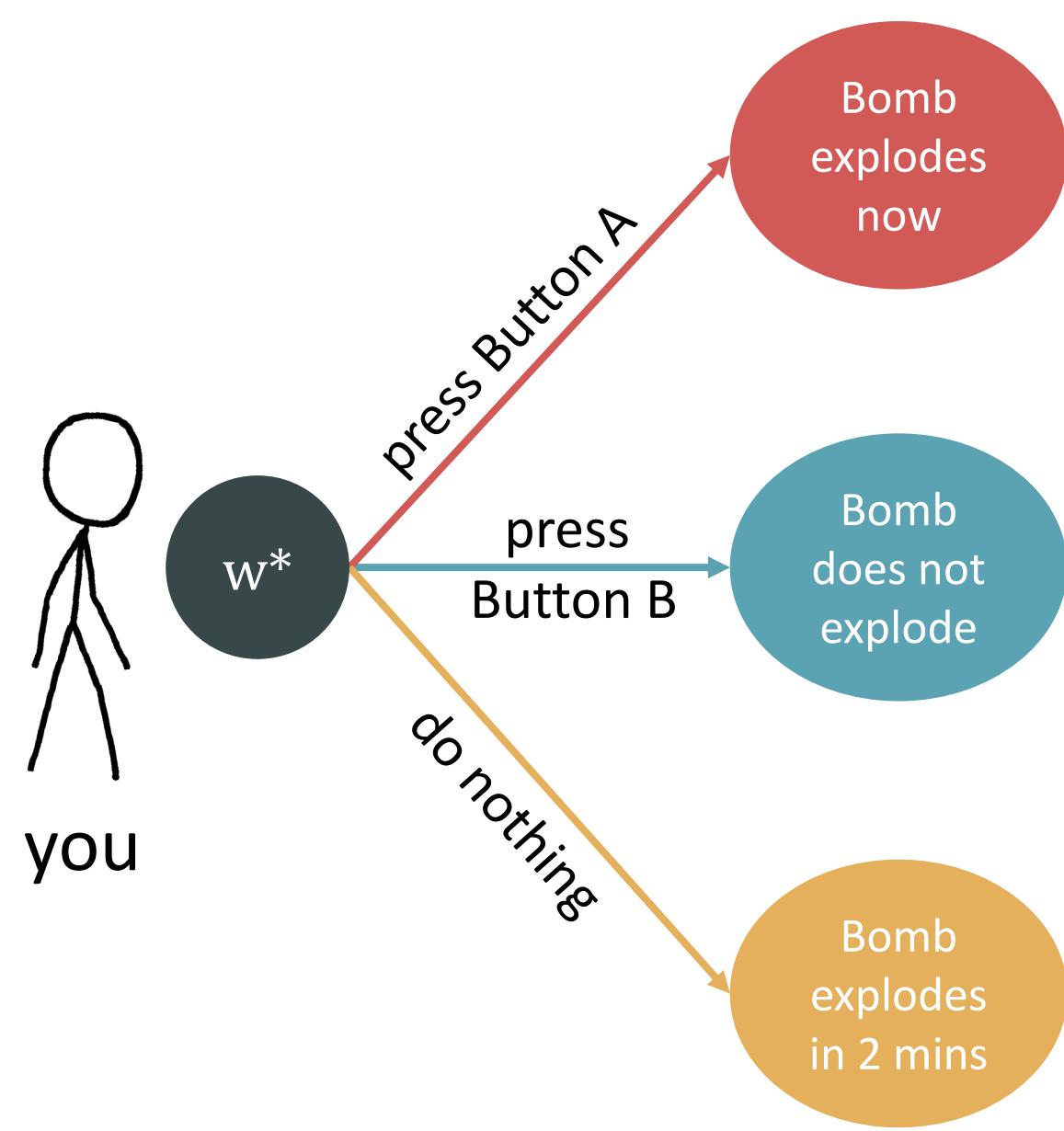
A

B

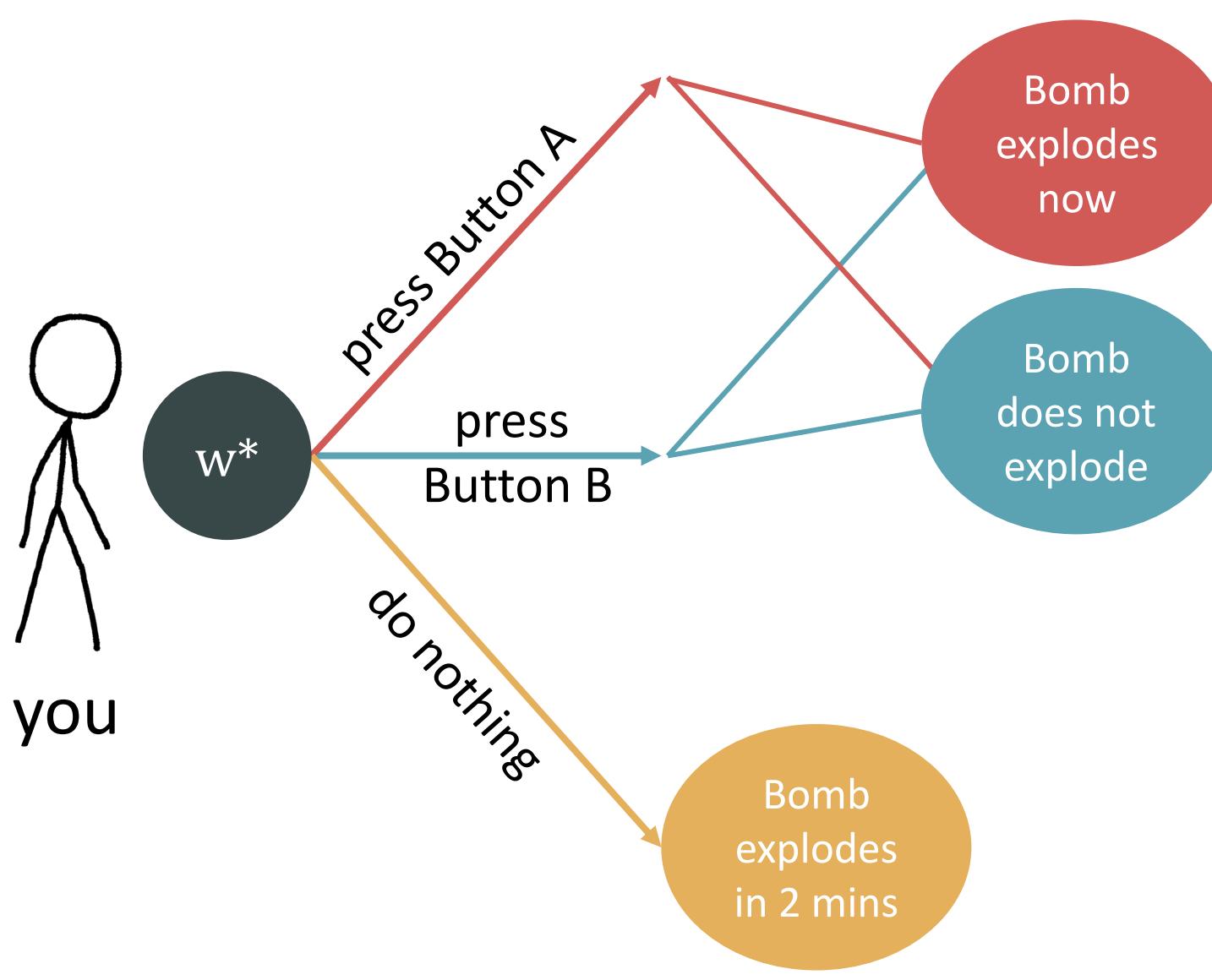
DECISION SITUATIONS

You have to defuse a bomb with two buttons.

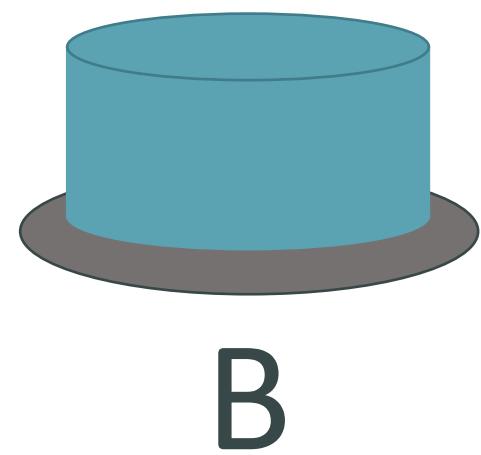
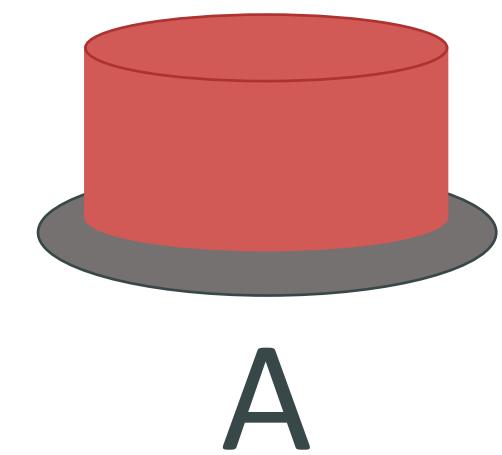
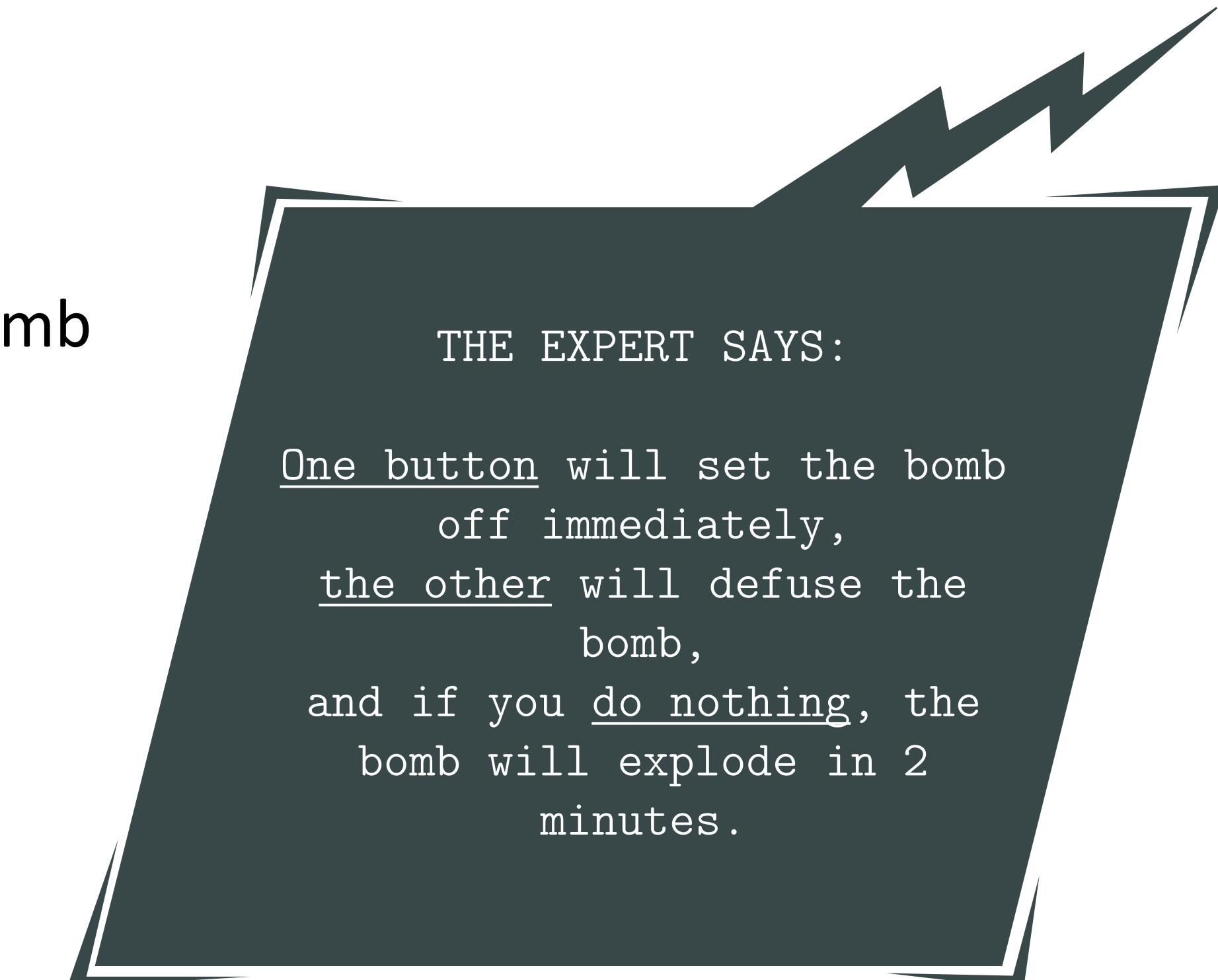
You know that the **expert is right**. Button A is the one that lets the bomb explode, and button B defuses the bomb, but you don't know that.



objective decision situation



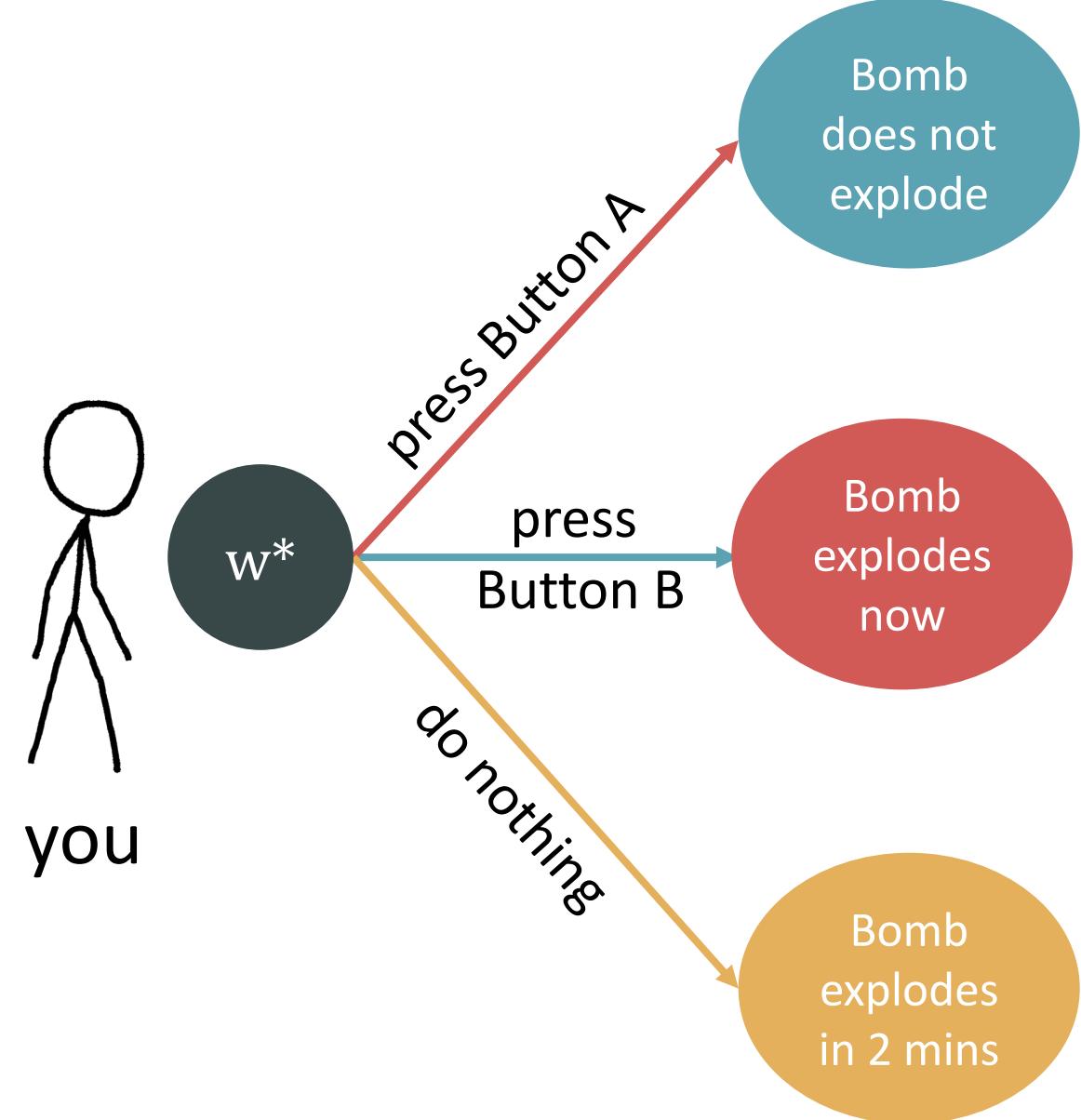
subjective decision situation



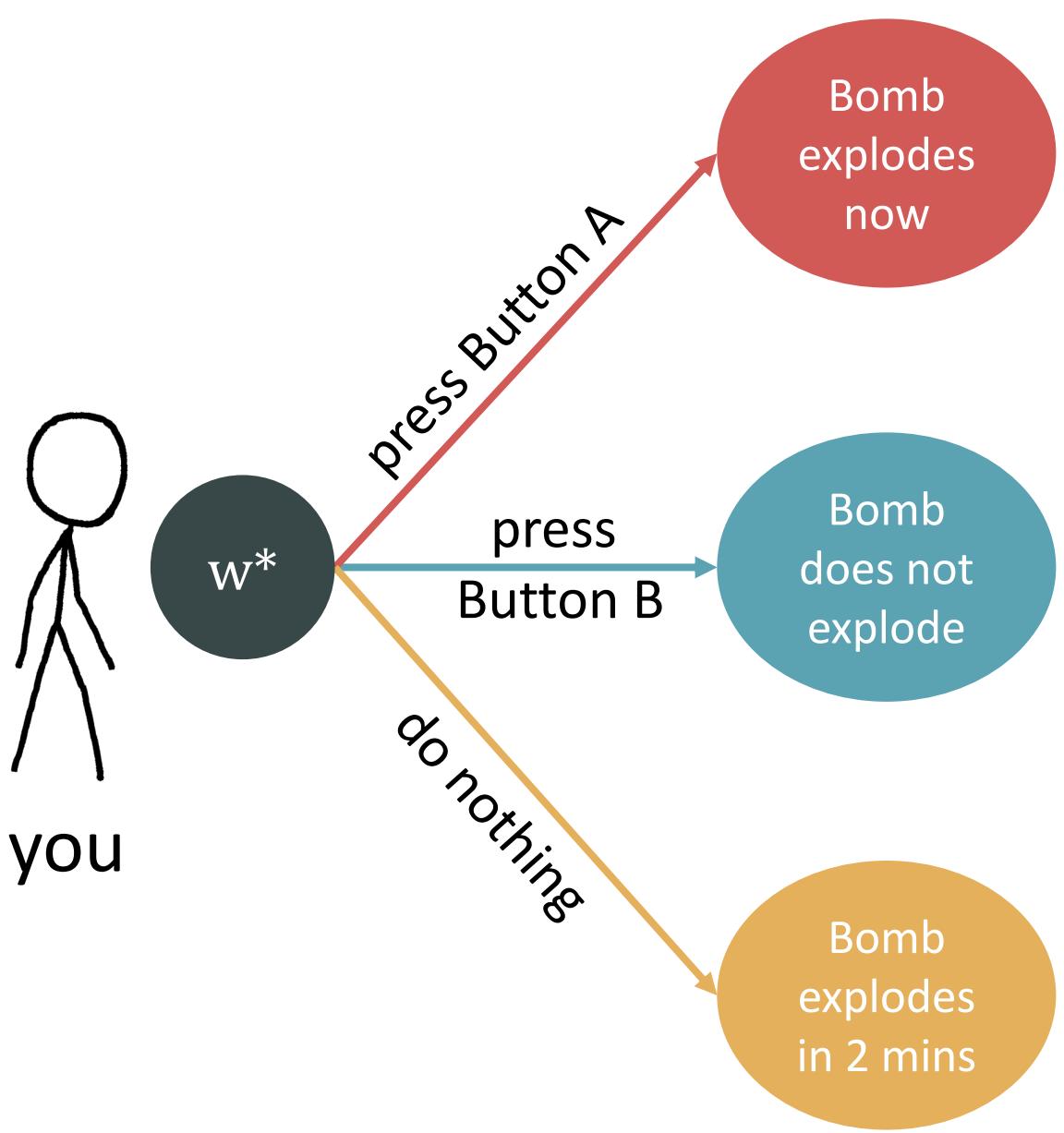
DECISION SITUATIONS

You have to defuse a bomb with two buttons.

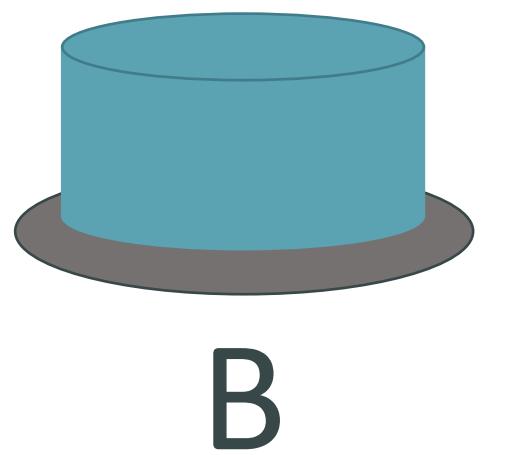
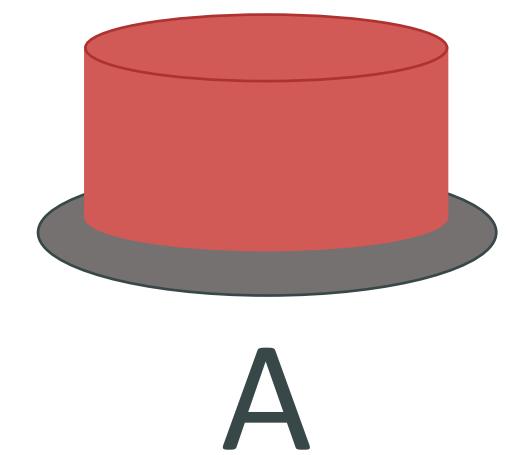
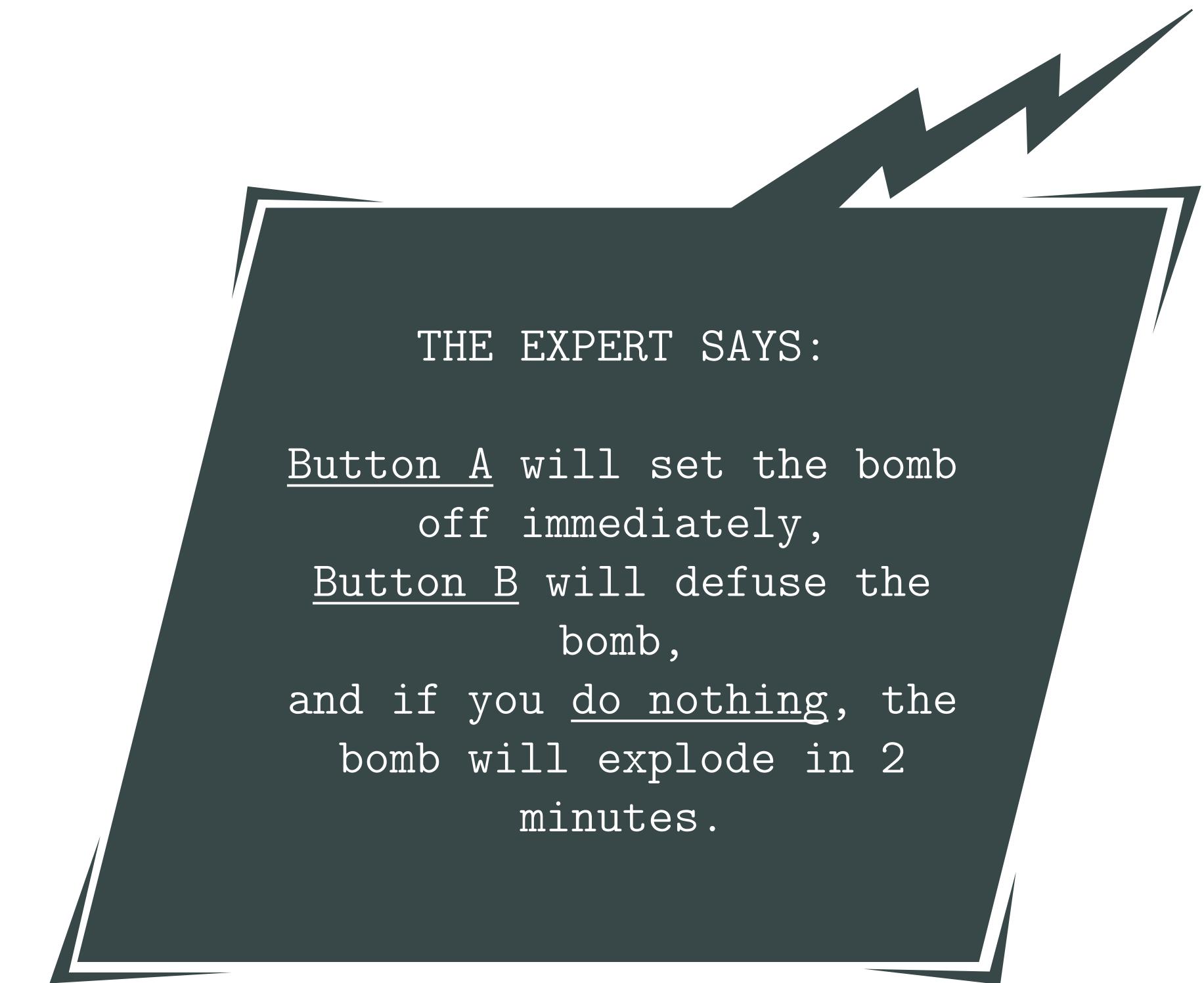
You believe that the expert for sure is right. But in fact he is wrong.



objective decision situation



subjective decision situation



CONSEQUENTIALISM

The moral/normative/deontic status of an action solely depends on its consequences (and the consequences of the alternative actions)

Standard ‘framework’:

Agent A ought to perform one of the right actions.

An action ϕ is a right action iff the **relevant qualities** of the **consequences** of ϕ fulfill a **specific condition**.

Which qualities?

Which consequences?

Which condition?

CONSEQUENTIALISM

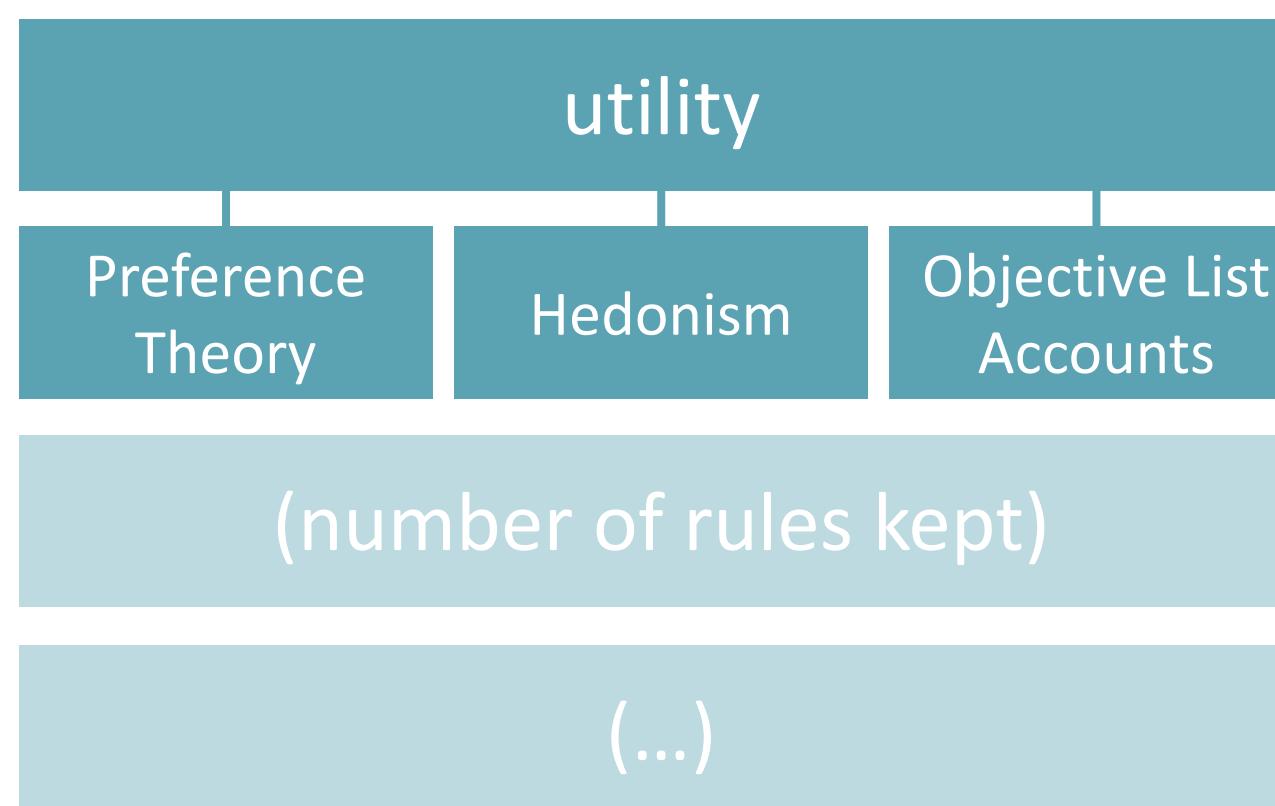
The moral/normative/deontic status of an action solely depends on its consequences (and the consequences of the alternative actions)

Standard ‘framework’:

Agent A ought to perform one of the right actions.

An action ϕ is a right action iff the **relevant qualities** of the **consequences** of ϕ fulfill a **specific condition**.

Which qualities?



Which qualities?

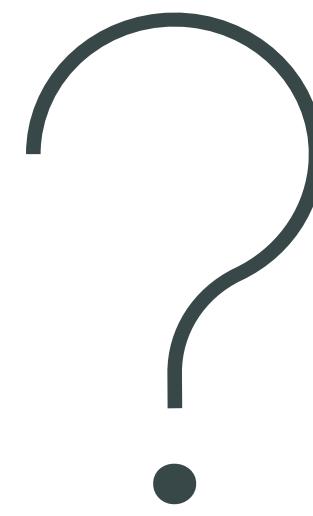
Depends on the axiology!

Axiology (working definition)

A theory about what has value, especially about what has intrinsic value.

The study of what has value, especially what has intrinsic value.

Something is intrinsically valuable when it is valuable ‘as such’, ‘in itself’, or ‘for its own sake’. Nothing else is needed in order to make it valuable. But what has intrinsic value?



<i>lemons</i>	<i>happiness</i>	<i>health</i>
<i>life</i>	<i>fulfilled wishes</i>	<i>buttons</i>
<i>prime numbers</i>	<i>knowledge</i>	<i>the earth</i>
<i>love</i>	<i>people</i>	<i>beauty</i>
	<i>umbrellas</i>	

Which qualities?

Depends on the axiology!

Hedonism

pleasure vs pain

Intuition: How good/bad is someone feeling?

Preference Theory

preference satisfaction vs preference frustration

Intuition: How well are someone's wishes, desires etc satisfied?

Objective List Accounts

objective goods

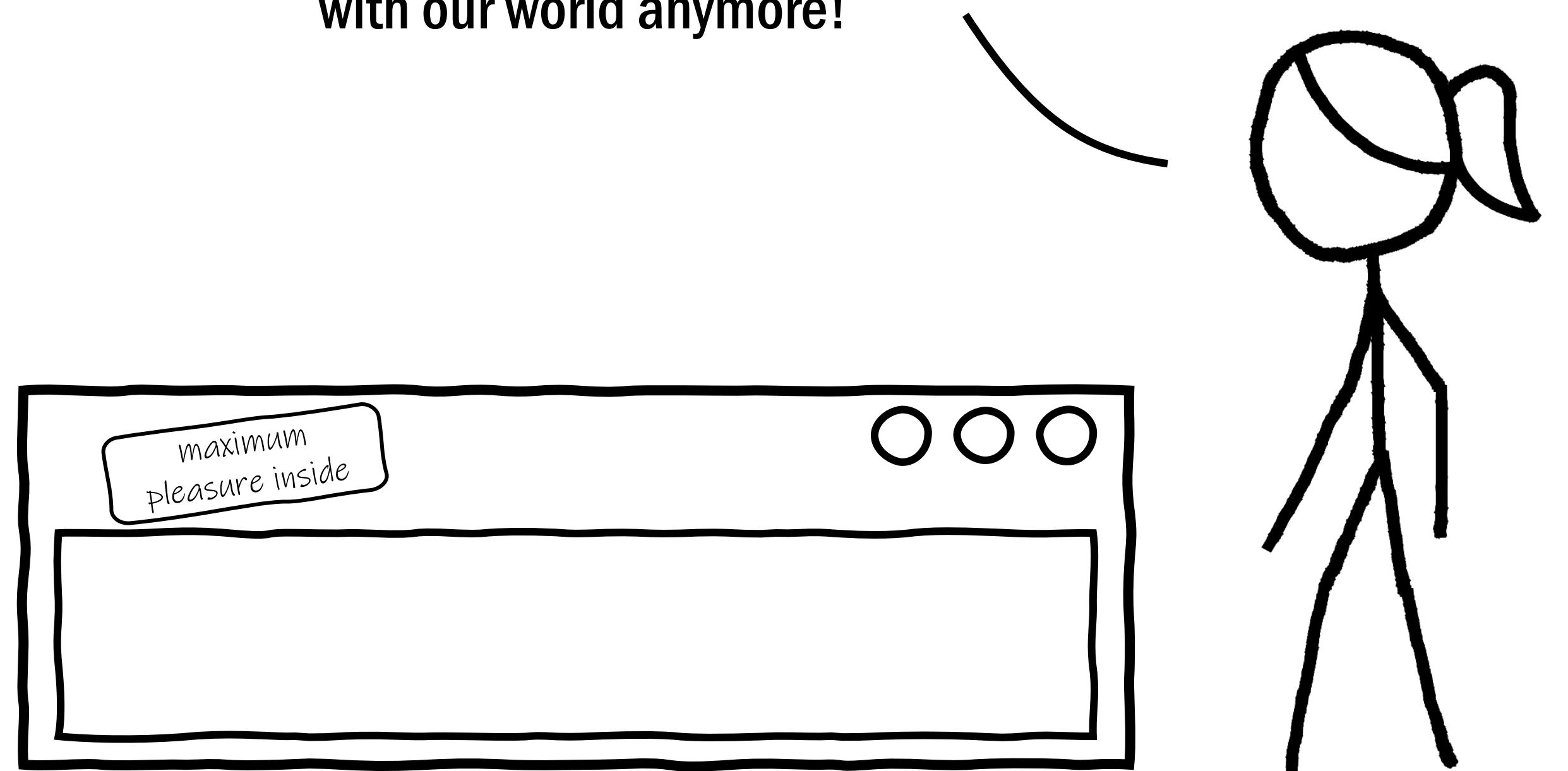
Intuition: How much objective goods from the list does someone have and to what degree?

You have access to a **pleasure machine**:

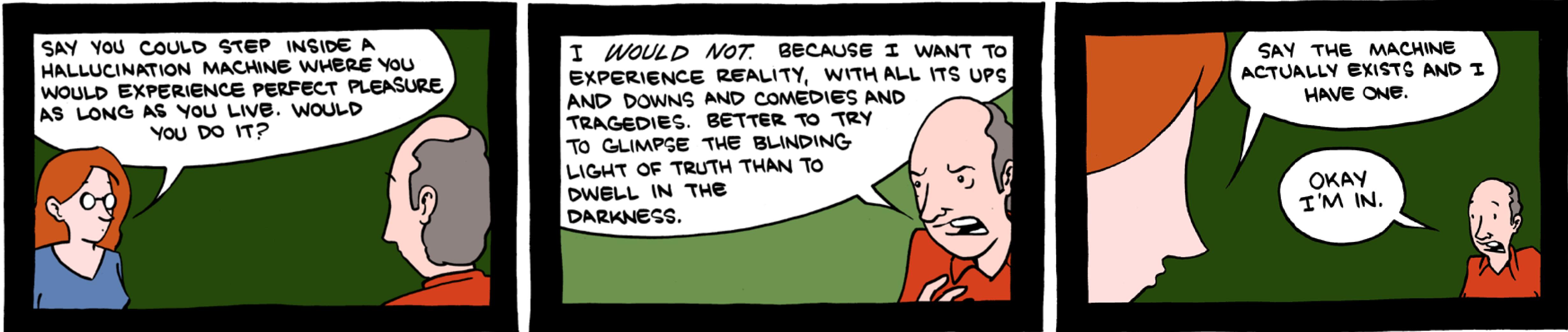
The machine is guaranteed to simulate a perfect, long life to any person in the machine. They will be truly happy in the machine for the rest of their life. The downside: they can never again be taken out of the machine.

- Is it good for someone to be inside the machine?
- Is it good for someone to be inside the machine against their will?

I do not want to be inside the pleasure machine ever! It is my worst nightmare to be in the machine where I cannot interact with our world anymore!



AXIOLOGIES



<https://www.smbc-comics.com/comic/2012-05-29>

The moral/normative/deontic status of an action solely depends on its consequences (and the consequences of the alternative actions)

Standard ‘framework’:

Agent A ought to perform one of the right actions.

An action ϕ is a right action iff the **relevant qualities** of the **consequences** of ϕ fulfill a **specific condition**.

Which consequences?

subjective/expected

objective/de facto

...

Which consequences?

Imagine a normal light switch which was rewired by a villain to detonate a bomb and kill all of us. Timo wants to turn on the lights and presses the switch.

subjective accounts

only (counterfactual) consequences, that have to be expected by the agent have to be taken into account

Timo could not have expected that the switch would detonate a bomb, so the detonation does not have to be taken into account when evaluating what Timo ought to do.

objective accounts

only the de facto (counterfactual) consequences have to be taken into account

Timo's flicking the switch set off the bomb, so the detonation does have to be taking into account when evaluating what Timo ought to do.

The moral/normative/deontic status of an action solely depends on its consequences (and the consequences of the alternative actions)

Standard ‘framework’:

Agent A ought to perform one of the right actions.

An action ϕ is a right action iff the **relevant qualities** of the **consequences** of ϕ fulfill a **specific condition**.

Which condition?

maximization

satisfaction

equalization

...

CONSEQUENTIALISM

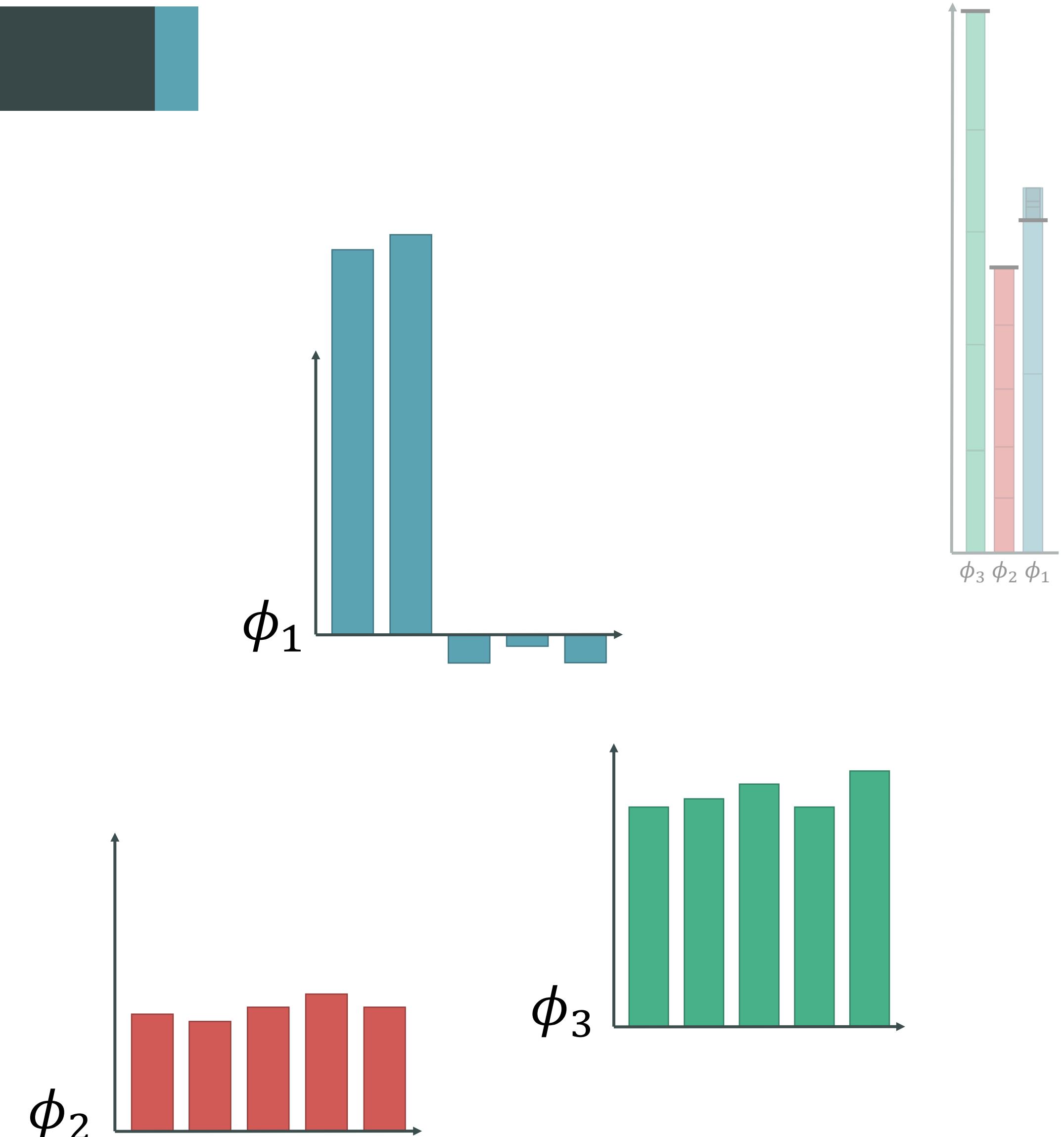
Which condition?

Idea 1 – Maximization:

the sum of the quantified relevant qualities of the consequences of ϕ are at least as high as the sum of the quantified relevant qualities of the consequences of all the alternatives of ϕ

Pro: judges ϕ_3 to be better than ϕ_2

Contra: judges ϕ_1 to be better than ϕ_2



CONSEQUENTIALISM

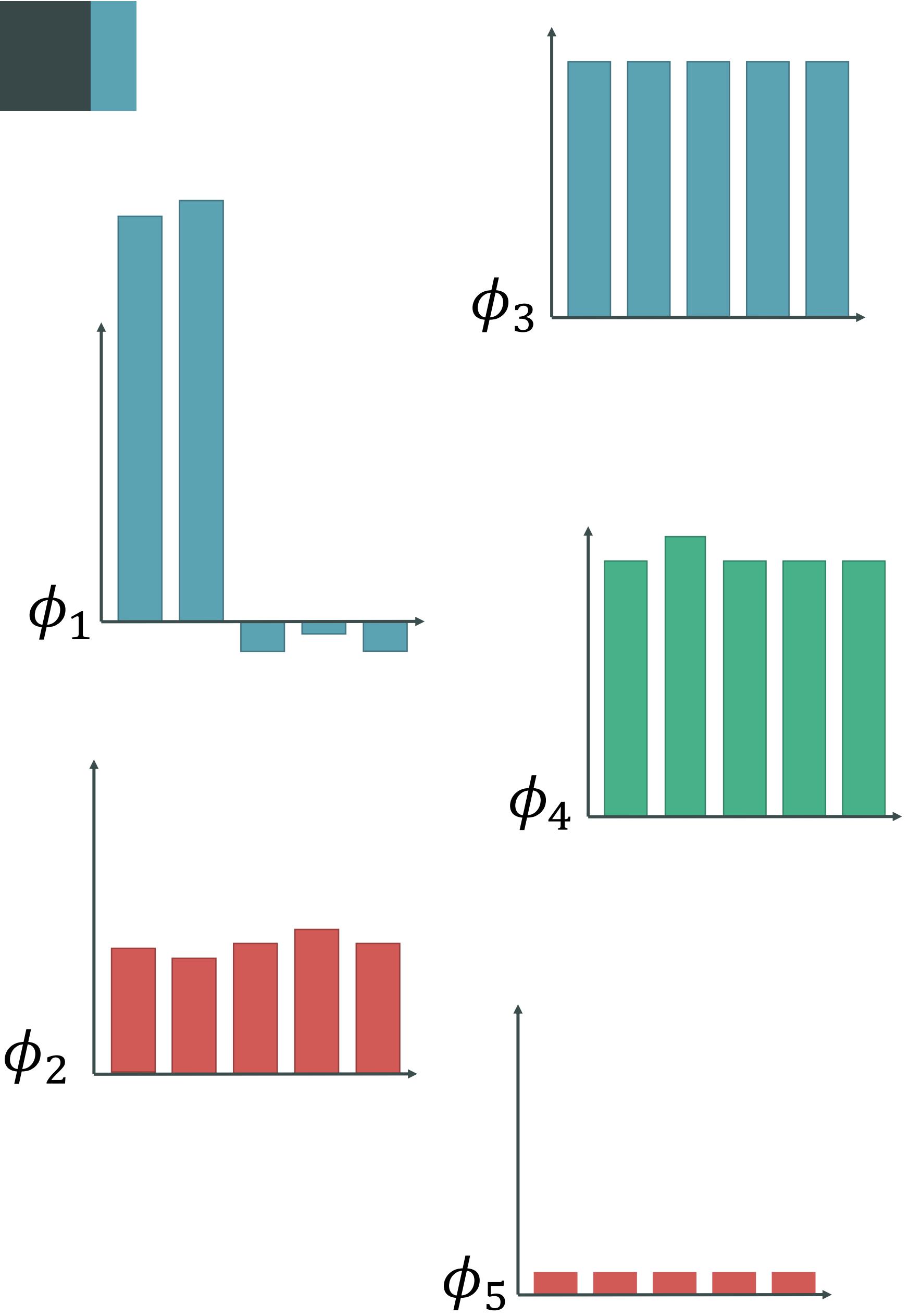
Which condition?

Idea 2 – Equalization:

the relevant qualities of the consequences of ϕ are most equally distributed among all relevant beings

Pro: judges ϕ_2 to be better than ϕ_1

Contra: judges ϕ_3 to be equally good as ϕ_5
and ϕ_4 to be worse than ϕ_5



Which condition?

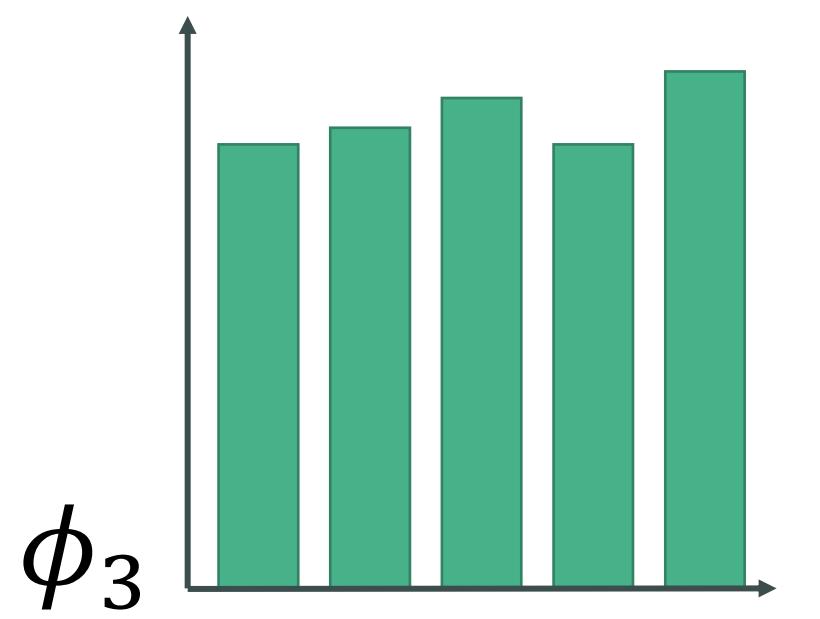
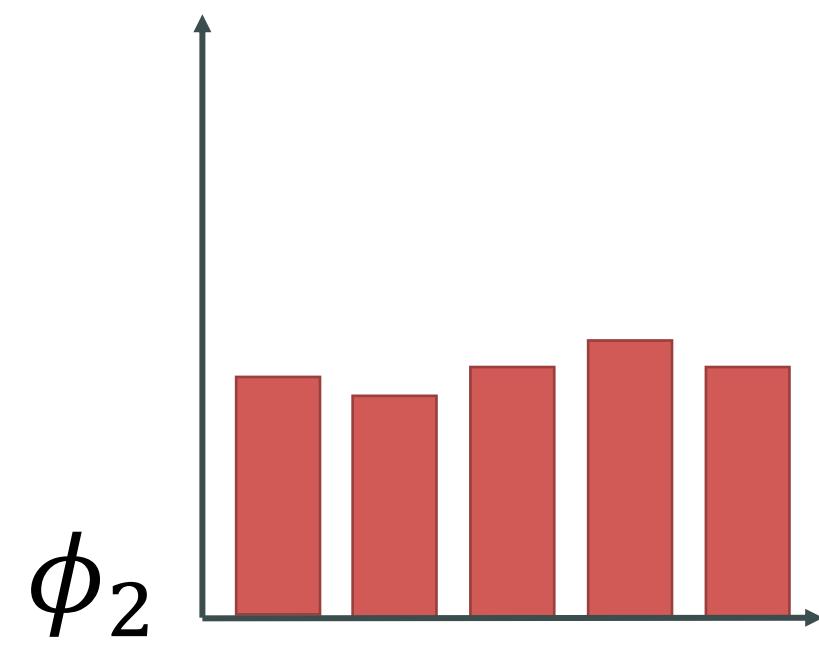
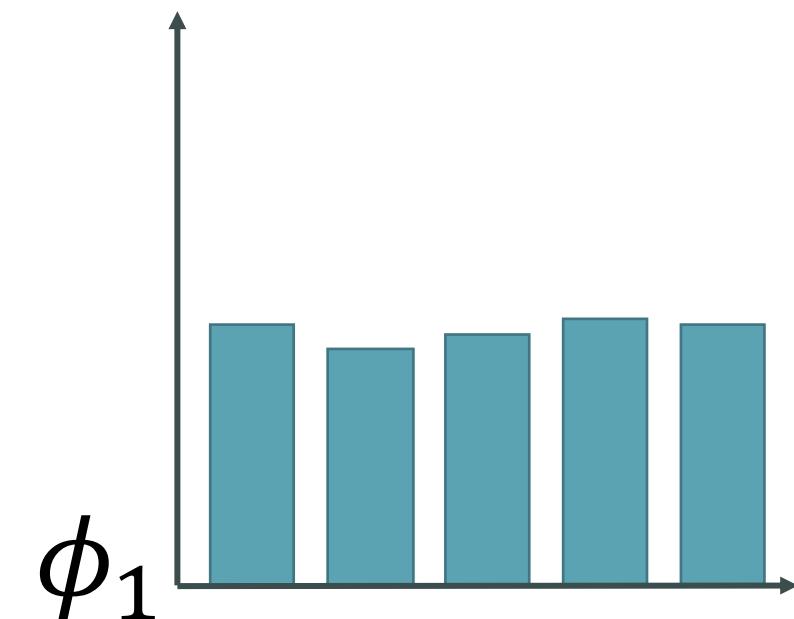
Idea 3 – Collective Satisfaction:

the sum of the quantified relevant qualities of the consequences of ϕ is above a certain threshold

Pro: judges both ϕ_1 and ϕ_2 to be right

Contra: judges both ϕ_2 and ϕ_3 to be right

Assume that all of the following are above the threshold:



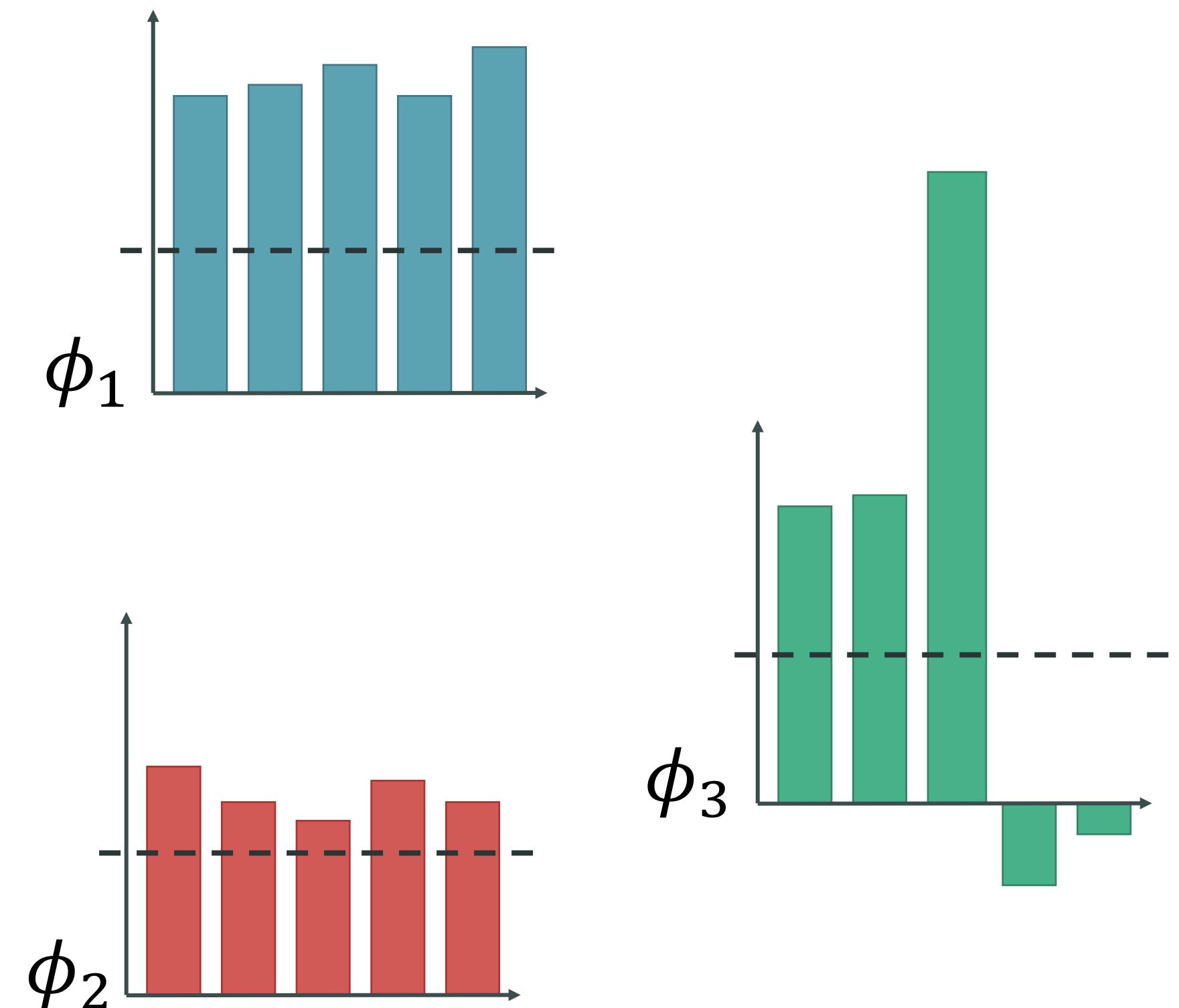
Which condition?

Idea 4 – Individual Satisfaction:

the quantified relevant qualities of the consequences of ϕ are above a certain threshold for all relevant beings

Pro: judges ϕ_1 to be better than ϕ_3

Contra: judges both ϕ_1 and ϕ_2 to be right



CONSEQUENTIALISM

The moral/normative/deontic status of an action solely depends on its consequences (and the consequences of the alternative actions)

Standard ‘framework’:

Agent A ought to perform one of the right actions.

An action ϕ is a right action iff the **relevant qualities** of the **consequences** of ϕ fulfill a **specific condition**.

Which qualities? (Axiology)

utility	
Preference Theory	Hedonism
(number of rules kept)	
(...)	

Which consequences?

subjective/expected
objective/de facto
...

Which condition?

maximization
satisfaction
equalization
...

Example I (*Classical Utilitarianism*)

relevant qualities: pleasure and pain (Hedonism)

kind of consequences: de facto, counterfactual consequences (objective)

condition: maximization

→ (*Classical Utilitarianism*): An action is right iff no alternative action would, de facto, lead to a better world, i.e., there is no alternative with a higher total value of pleasure minus pain.

$$U(\phi) = \sum_{a \in A(w(\phi))} \text{pleasure}(a) - \text{pain}(a)$$

Possible world that is consequence of the action

all relevant individuals in world w

Pleasure and pain of individual a

Example II (*Expected Utility Preference Utilitarianism*)

relevant qualities: preference satisfaction and preference frustration (preference theory)

kind of consequences: counterfactual consequences that are to be expected (subjective)

condition: maximization

→ *Expected Utility Preference Utilitarianism*: An action is right iff no alternative action has a higher expected utility – i.e. an act ϕ is right iff there is no alternative act ϕ' with $EU(\phi') > EU(\phi)$, where

$$EU(\phi) = \sum_{w \in W(\phi)} p_\phi(w) * u(w)$$

Possible worlds as consequences of the action

Utility of consequence (is itself a sum)

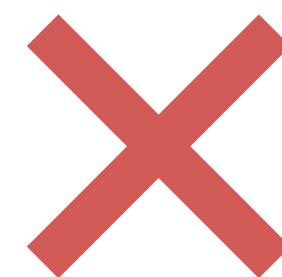
Probability for consequence given action

Those look a lot like you need concrete values for $pleasure(a)$, $pain(a)$, and $u(w)$

$$U(\phi) = \sum_{w \in W(\phi)} \sum_{a \in A(w)} pleasure(a) - pain(a)$$

$$EU(\phi) = \sum_{w \in W(\phi)} p_\phi(w) * u(w)$$

but in fact you should never guess values!



DON'T: Calculations

- “She has a pleasure of 5.”
- “His preference satisfaction is between -8 and -12.”
- “The utility of the obtaining possible world is over 9000.”
- “He has a pleasure of 20 and she has a pleasure of 120.”



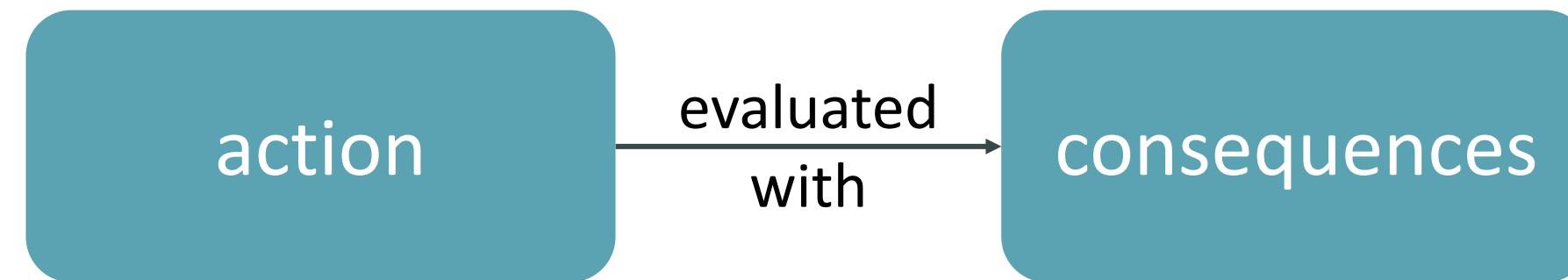
DO: Estimates

- “Her pleasure is very low.”
- “He has some mild preference frustration.”
- “The utility of the obtaining possible world is greater than the utility of another world.”
- “Her pleasure is much greater than his.”

A MISNOMER



Agent A ought to perform one of the right actions.
An action ϕ is a right action iff the **relevant qualities** of the **consequences** of ϕ fulfill a **specific condition**.



Agent A ought to perform one of the right actions.
An action ϕ is a right action iff it adheres to a right set of rules.
A set of rules is right iff the general adherence to these rules has consequences, where **relevant qualities** of these **consequences** fulfill a **specific condition**.



A MISCONCEPTION

Every action can be justified with consequentialism!

Anything goes with consequentialism!

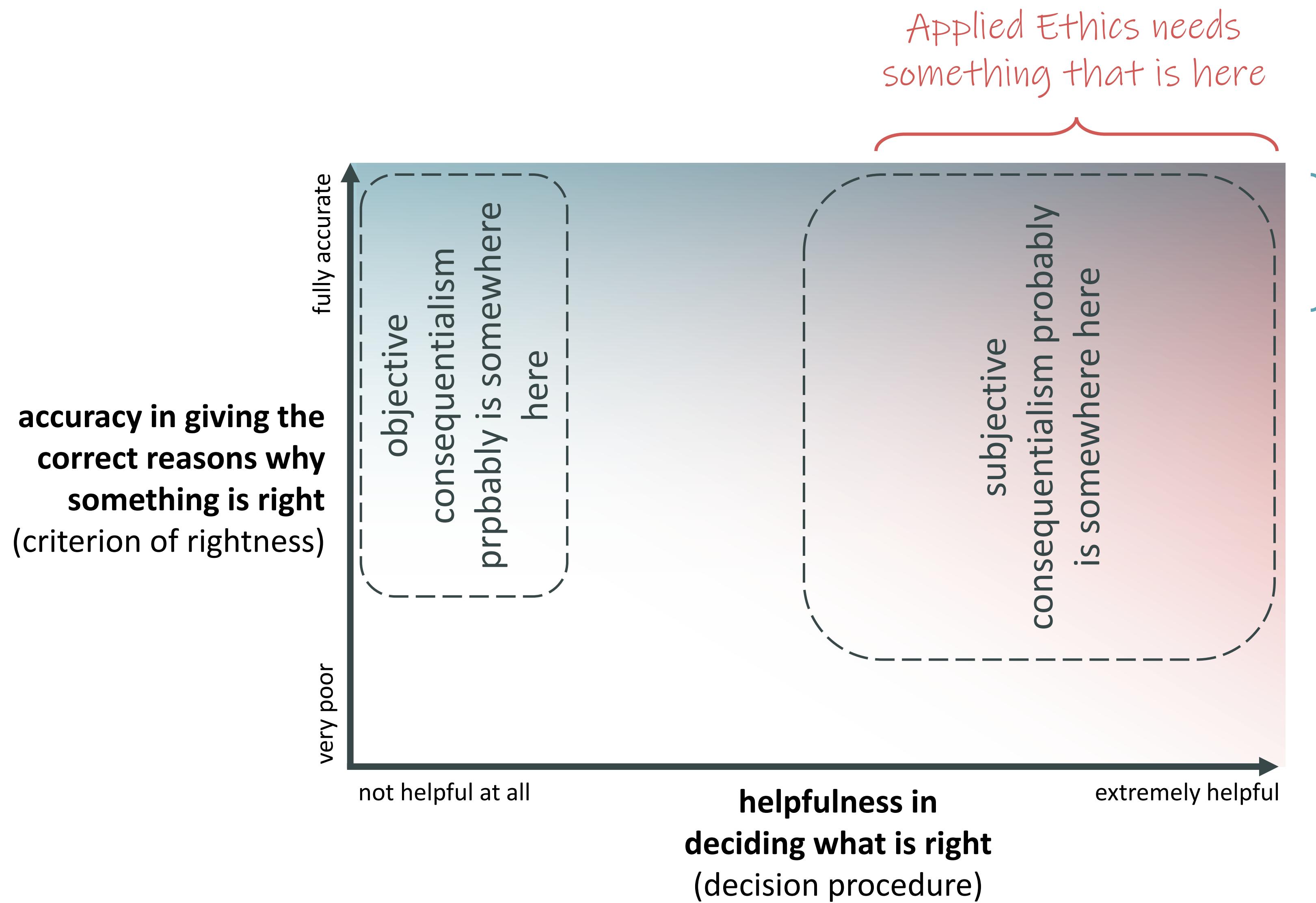
common, but unwarranted criticism

“The end justifies the means.”
does not mean the same as
“Every means is always justified.”

proper axiologies are non-arbitrary

just because certain forms of consequentialism
are hard or even impossible to apply, it does not
mean that they are bad theories

CRITERIA VS DECISION PROCEDURES





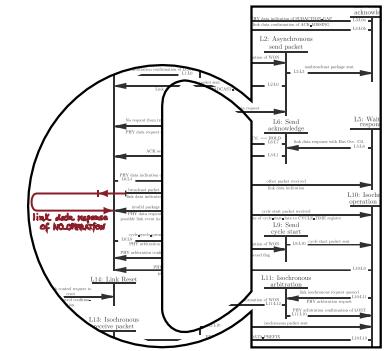


Ethics for Nerds

An Advanced Course in Computer Science
Summer Semester 2020

Ethics 3.2
Consequentialism

Problems for Consequentialism I



Prof. Holger Hermanns,
Kevin Baum, Sarah Sterz

DIFFERENT KINDS OF ATTACKS ON MORAL THEORIES

A theory T can be...

inadequate from a theory-external perspective

A theory T can be...

inadequate from a theory-internal perspective

A theory T can be...

inadequate from a theory-external perspective

- “ T come to the intuitively false assessments!” ($\sim T$ is extensionally inapt)
- “ T has counterintuitive implications!” ($\sim T$ is absurd in its consequences)
- “ T cannot make sense of phenomenon F , but every adequate theory has to be able to do so!” ($\sim T$ is theoretically unsound)
- “There is another theory T' which is a better theory than T .” ($\sim T$ should be replaced)
- ...

A theory T can be...

inadequate from a theory-internal perspective

- “ T comes to contradictory assessments!” ($\sim T$ is inconsistent)
- “ T is failing to achieve its own goals.” ($\sim T$ is efficacious)
- “ T is failing to live up to its own standards!” ($\sim T$ is unacceptable for proponents of T -like theories)
- “There is another theory T' which is a better variant of T .” ($\sim T$ can be improved but remains T -like.)
- ...

PROBLEMS OF CONSEQUENTIALISM

inadequate from a theory-external perspective

Too Demanding

*No Place for
Incommensurable Values*

Undermines Integrity

Nearest Dearest

inadequate from a theory-internal perspective

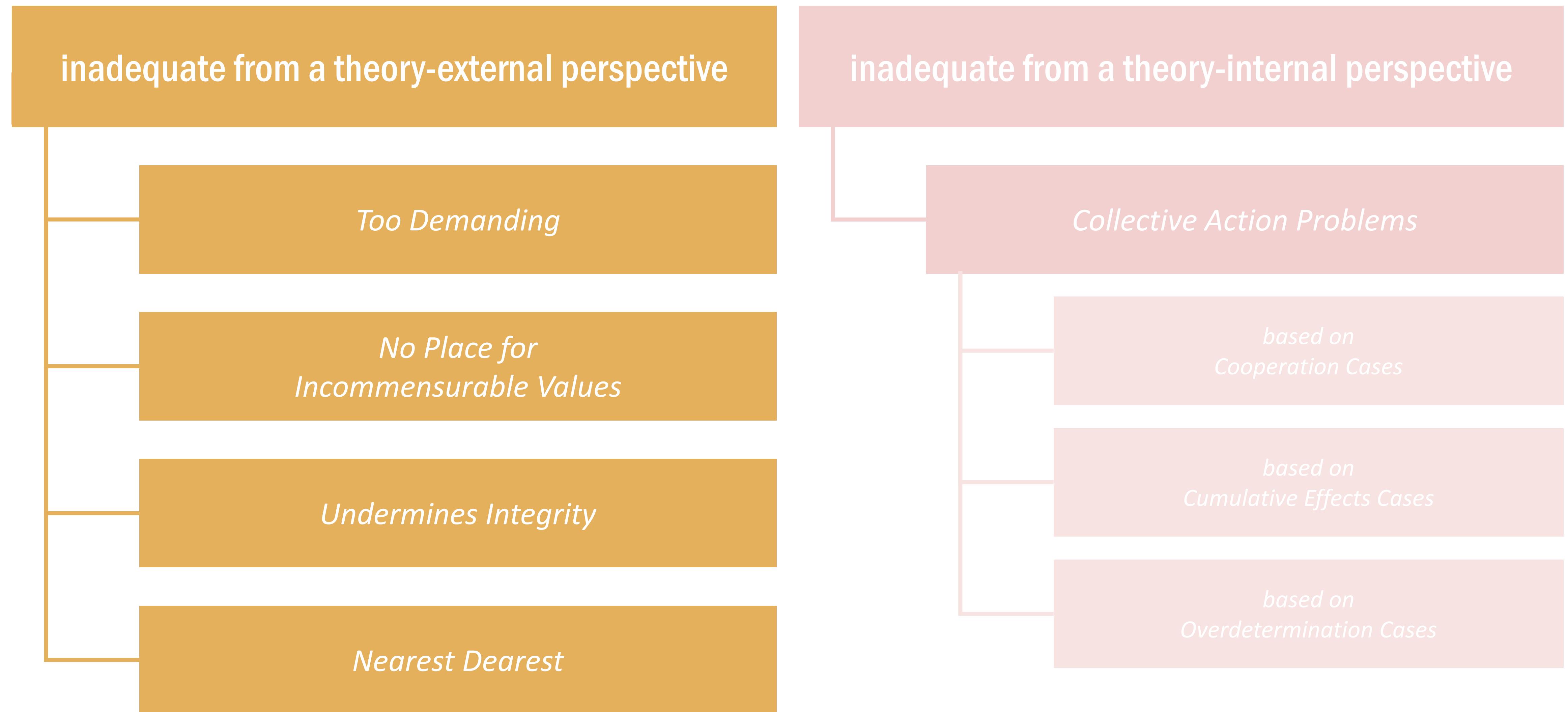
Collective Action Problems

*based on
Cooperation Cases*

*based on
Cumulative Effects Cases*

*based on
Overdetermination Cases*

PROBLEMS OF CONSEQUENTIALISM



PROBLEMS OF CONSEQUENTIALISM

inadequate from a theory-external perspective

Too Demanding

*No Place for
Incommensurable Values*

Undermines Integrity

Nearest Dearest

inadequate from a theory-internal perspective

Collective Action Problems

*based on
Cooperation Cases*

*based on
Cumulative Effects Cases*

*based on
Overdetermination Cases*

General Overdemandingness

Consequentialism can demand to make huge sacrifices

Examples:

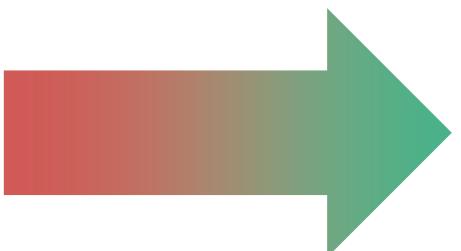
- catching a grenade to save a group of people
- jumping off the bridge if you are the fat man

P1: Standard forms of Utilitarianism imply that the fat man is obliged to sacrifice himself.

P2: If P1, then standard forms of Utilitarianism are overly demanding.

P3: If a moral theory T is overly demanding, then T is not an adequate moral theory.

C: Therefore, standard forms of Utilitarianism are not adequate moral theories.



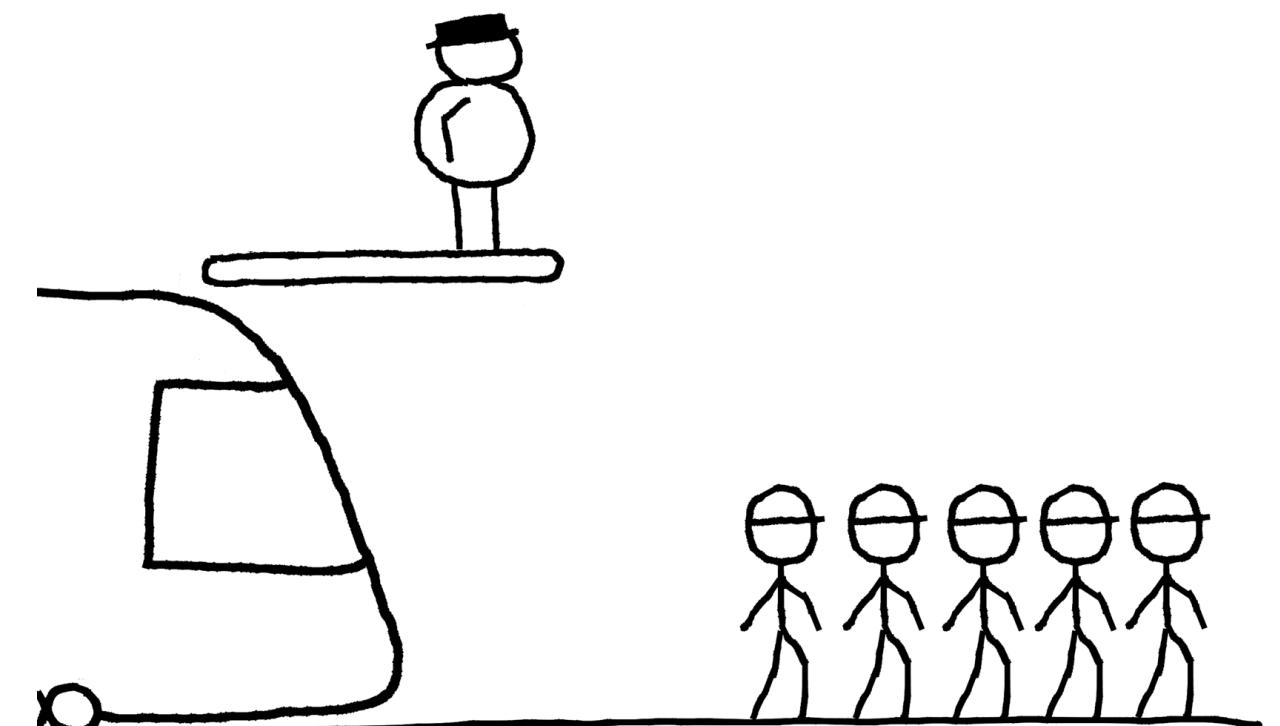
Possible Reactions

- bite the bullet (deny P2)
- add special weight to one's own consequences (fix P1)
- make room for supererogation in your theory and find a proper threshold (fix P1)

Supererogation (working definition)

There are non-obligatory, but permissible actions that are better than what you are obligated to.

...



PROBLEMS OF CONSEQUENTIALISM

inadequate from a theory-external perspective

Too Demanding

*No Place for
Incommensurable Values*

Undermines Integrity

Nearest Dearest

inadequate from a theory-internal perspective

Collective Action Problems

*based on
Cooperation Cases*

*based on
Cumulative Effects Cases*

*based on
Overdetermination Cases*

PROBLEMS OF CONSEQUENTIALISM

Too Demanding

No Incommensurable Values

Colosseum Case

Ceteris paribus, for any whatever great pain p and whatever small amusement a , there is a number n such that, if n people were to experience a this outweighs p – which makes it permissible, even obligatory to inflict p .

Example: gladiators fight to death for amusement of a sufficiently large audience:

- audience's joy > suffering of the gladiators
- it's right (according to consequentialism) to make them fight to death

counterintuitive!

PROBLEMS OF CONSEQUENTIALISM

<https://www.bbc.com/future/article/20170105-why-we-may-be-living-in-the-future-of-the-running-man>

FUTURE What is BBC Future? Follow the Food Future Planet 100 Year Life Best of... Japan 2020 Latest More ≡

FUTURE NOW | PREDICTION

Why we may be living in the future of The Running Man



By Frank Swain 5th January 2017

The vision of 2017 depicted in Arnold Schwarzenegger's 30-year-old dystopian action movie captures how our world is changing today.

In a world beset by a collapsing economy, the US media conspires with the government to keep the population in check with a combination of heavy-handed policing and a steady stream of vapid reality TV shows. Meanwhile, one of the most powerful men in the world is the host of a reality TV show.

Sound familiar? That was 2017 conjured by campy action thriller **The Running Man** when it was released 30 years ago.

PROBLEMS OF CONSEQUENTIALISM

Colosseum Case

Ceteris paribus, for any whatever great pain p and whatever small amusement a , there is a number n such that, if n people were to experience a this outweighs p – which makes it permissible, even obligatory to inflict p .

Example: gladiators fight to death for amusement of a sufficiently large audience:

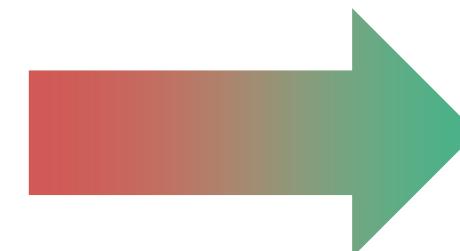
- audience's joy > suffering of the gladiators
- it's right (according to consequentialism) to make them fight to death

counterintuitive!

P1: Standard forms of Utilitarianism imply that it is allowed to let gladiators fight to death for the amusement of a large enough crowd.

P2: If P1, then standard forms of Utilitarianism are not adequate moral theories.

C: Therefore, standard forms of Utilitarianism are not adequate moral theories.



Possible Reactions

- biting the bullet (deny P2)
- denying the (actual) existence of such cases (deny P2): “for any relevant p and a , n would have to be larger than any realistic number of people, so I do not care.”
- adding different sorts of utility with lexicographic order (fix P1): mild amusement $<_{lex}$ severe pain/(painful) death
- ...

<https://www.bbc.com/future/article/20170105-why-we-may-be-living-in-the-future-of-the-running-man>

Why we may be living in the future of The Running Man

By Frank Swain 5th January 2017

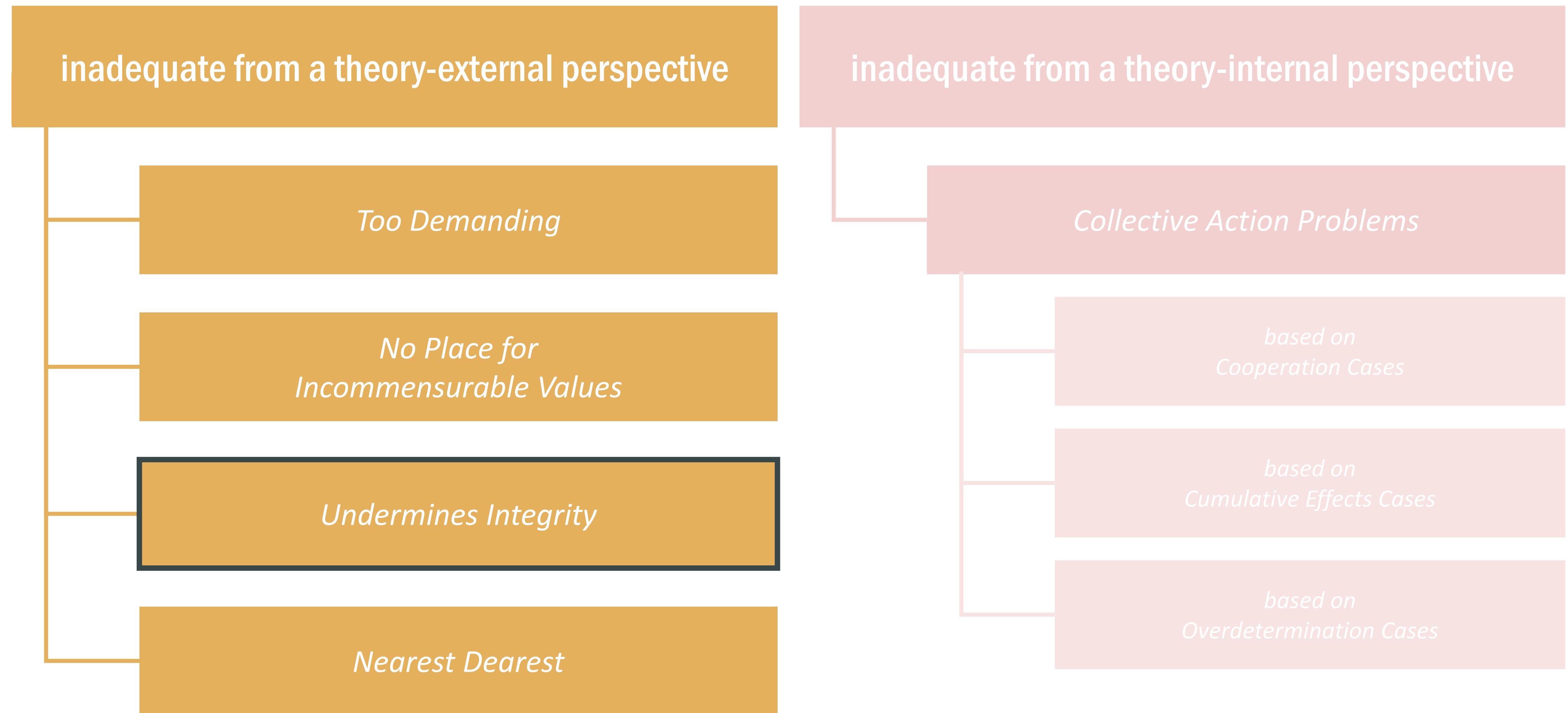
The vision of 2017 depicted in Arnold Schwarzenegger's 30-year-old dystopian action movie captures how our world is changing today.

In a world beset by a collapsing economy, the US media conspires with the government to keep the population in check with a combination of heavy-handed policing and a steady stream of vapid reality TV shows. Meanwhile, one of the most powerful men in the world is the host of a reality TV show.

Sound familiar? That was 2017 conjured by campy action thriller **The Running Man** when it was released 30 years ago.



PROBLEMS OF CONSEQUENTIALISM



The Chemist-Case

Consequentialism can demand you to violate central and identity-conferring standards of yours

Example:

George is a pacifist, but under suitable and non-exceptional conditions, it becomes obligatory for him to work for an armaments group and help to build new, cruel weapons.

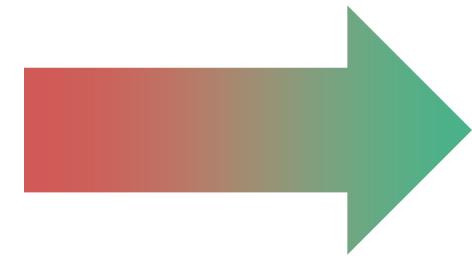
(Williams, Bernard, 1973. 'Integrity,' in J.J.C. Smart and Bernard Williams, Utilitarianism: For and Against New York: Cambridge, pp 97.)

P1: Standard forms of Utilitarianism imply that George ought to act against his central and identity-conferring standards.

P2: If P1, then standard forms of Utilitarianism do not allow to uphold central and identity-conferring standards.

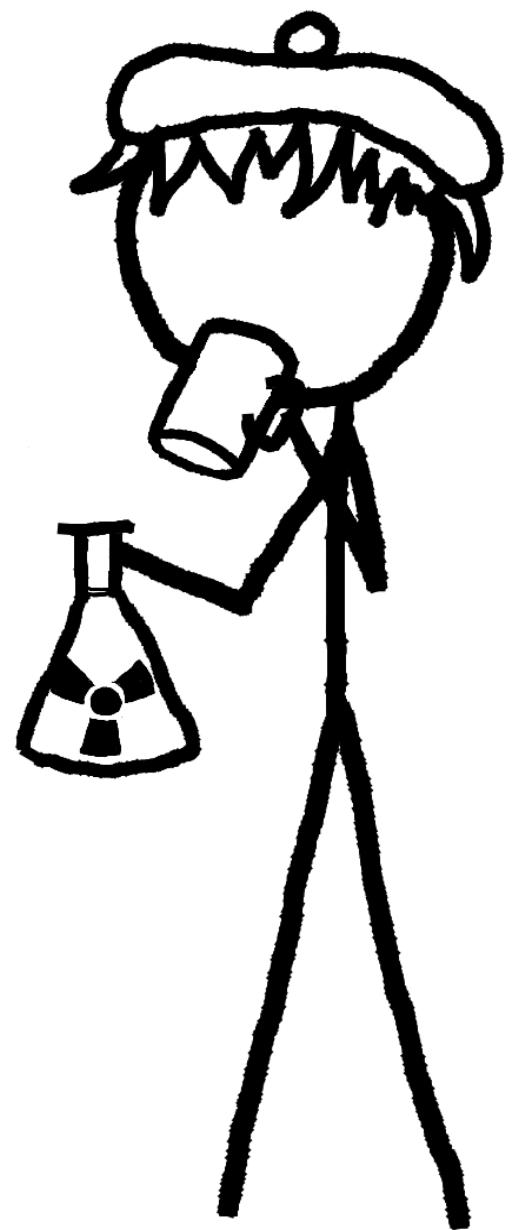
P3: If a moral theory T does not allow to uphold central and identity-conferring standards, then T is not an adequate moral theories.

C: Therefore, standard forms of Utilitarianism are not adequate moral theories.

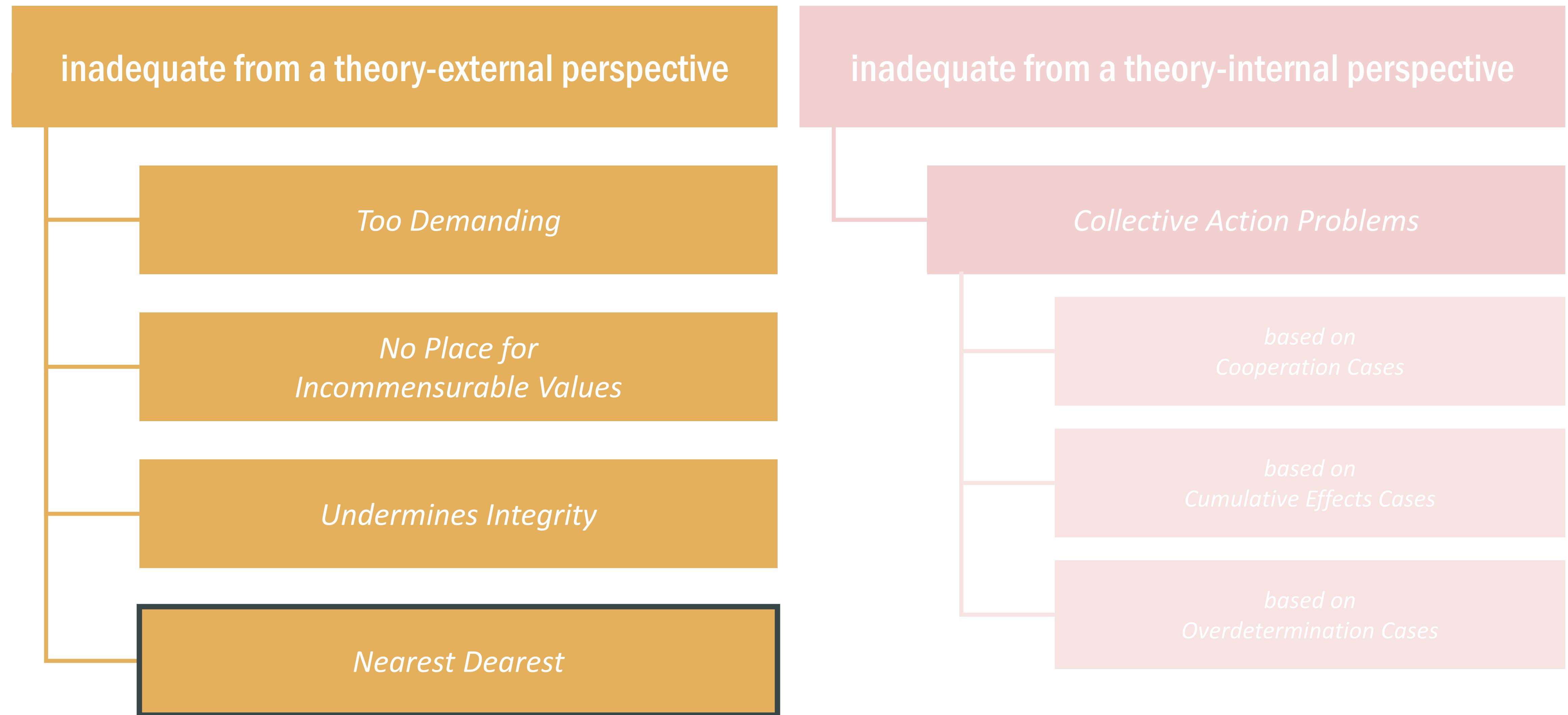


Possible Reactions

- biting the bullet (deny P3)
 - why integrity is of moral value is far from obvious
 - if integrity is important for people, it is implicitly integrated into our axiology
- There might be serious long-term effects unconsidered (deny P1)
- ...



PROBLEMS OF CONSEQUENTIALISM



Nearest-Dearest Problem

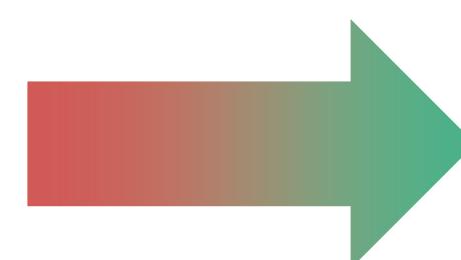
Consequentialism does not allow you to benefit the ones close to you (family, friends, ...) instead of strangers, if the overall utility was higher if you benefited the strangers

P1: Standard forms of Utilitarianism imply that Awes ought to rescue two children he does not know instead of his own daughter.

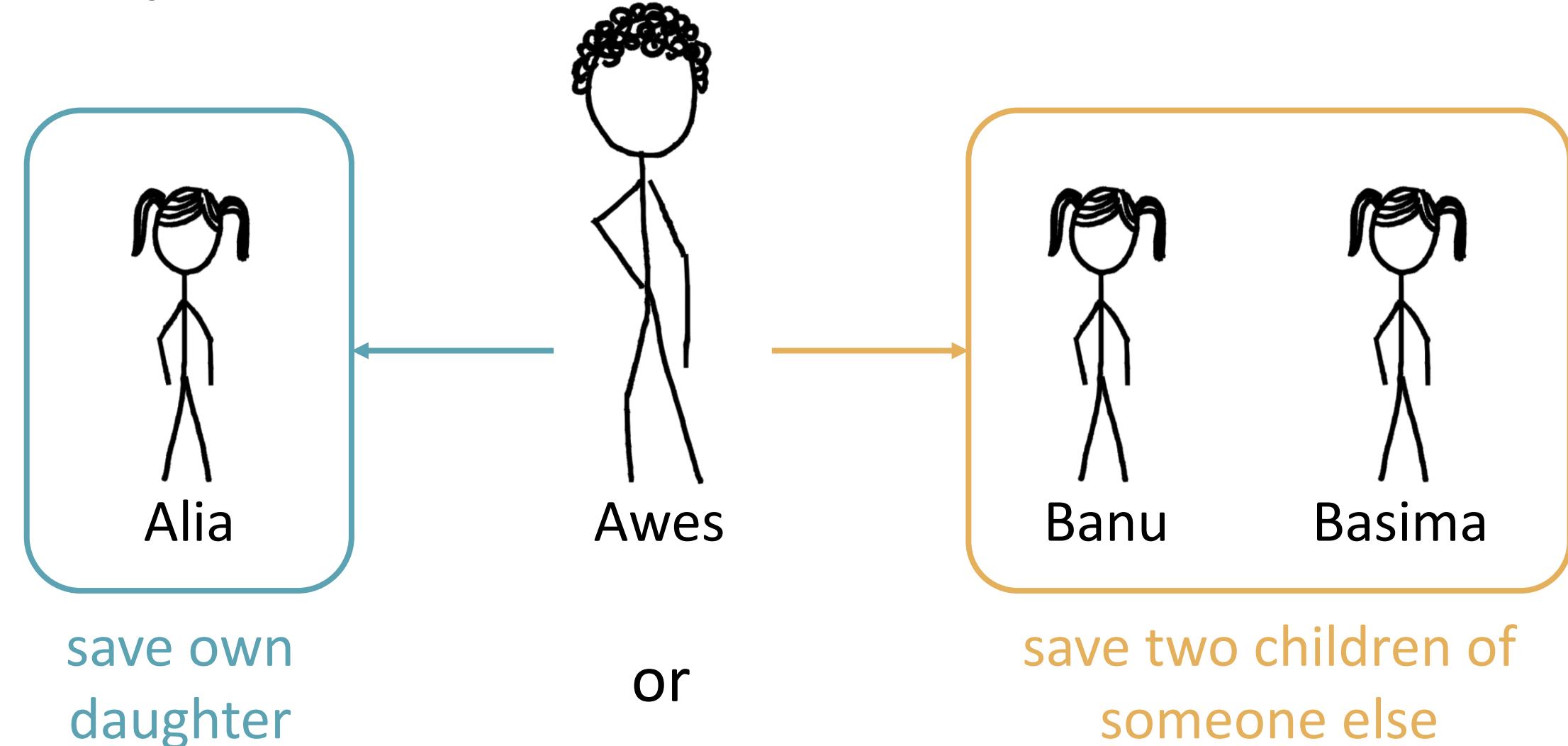
P2: If P1, then standard forms of Utilitarianism do not allow to prioritize our nearest and dearest.

P3: If a moral theory T does not allow to prioritize our nearest and dearest, then T is not an adequate moral theories.

C: Therefore, standard forms of Utilitarianism are not adequate moral theories.



Example:



Possible Reactions

- biting the bullet (deny P3)
- deny P1 (for details see Jackson, Frank, 1991. 'Decision-theoretic consequentialism and the nearest and dearest objection' in *Ethics*, Number 101, Volume 3, 461-482.)
- give a higher weight to the ones close to you (fix P1)
- ...



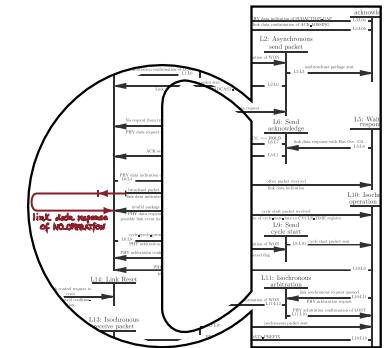


Ethics for Nerds

An Advanced Course in Computer Science
Summer Semester 2020

Ethics 3.3
Consequentialism

Problems for Consequentialism II



Prof. Holger Hermanns,
Kevin Baum, Sarah Sterz

PROBLEMS OF CONSEQUENTIALISM

inadequate from a theory-external perspective

Too Demanding

*No Place for
Incommensurable Values*

Undermines Integrity

Nearest Dearest

inadequate from a theory-internal perspective

Collective Action Problems

*based on
Cooperation Cases*

*based on
Overdetermination Cases*

*based on
Cumulative Effects Cases*

PROBLEMS OF CONSEQUENTIALISM

inadequate from a theory-external perspective

Too Demanding

*No Place for
Incommensurable Values*

Undermines Integrity

Nearest Dearest

inadequate from a theory-internal perspective

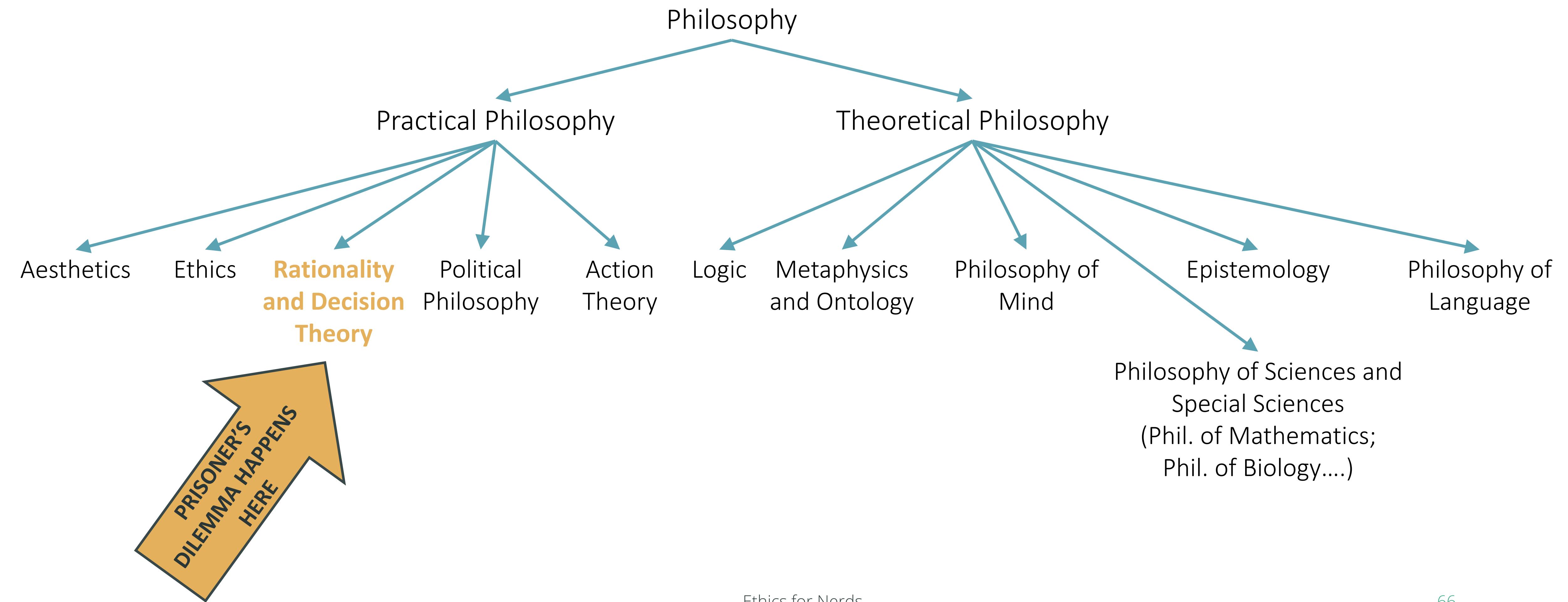
Collective Action Problems

*based on
Cooperation Cases*

*based on
Overdetermination Cases*

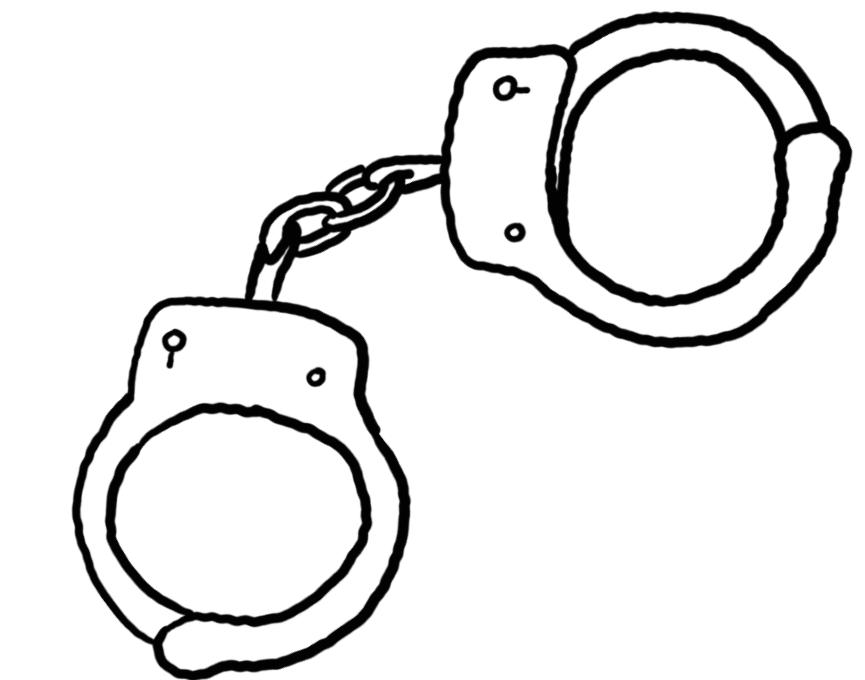
*based on
Cumulative Effects Cases*

RECAP: FIELDS OF PHILOSOPHY



COLLECTIVE ACTION AND RATIONALITY: THE PRISONER'S DILEMMA

		Col	
		remain silent	make a deal
		remain silent	make a deal
Row	remain silent	5 years	free
	make a deal	25 years	20 years

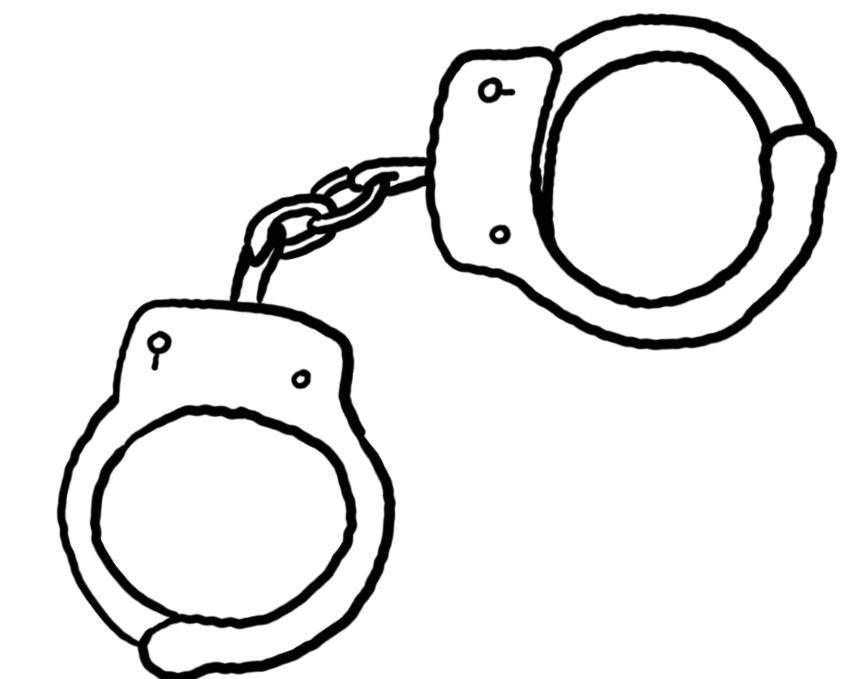


Principle of Strict Strategic (Rational) Dominance:

If there is one strategy for player A (say, performing ϕ) that is better for A independent of whatever strategy player B chooses, then it is (rationally) right for A to play that strategy (i.e., to ϕ).

COLLECTIVE ACTION AND RATIONALITY: THE PRISONER'S DILEMMA

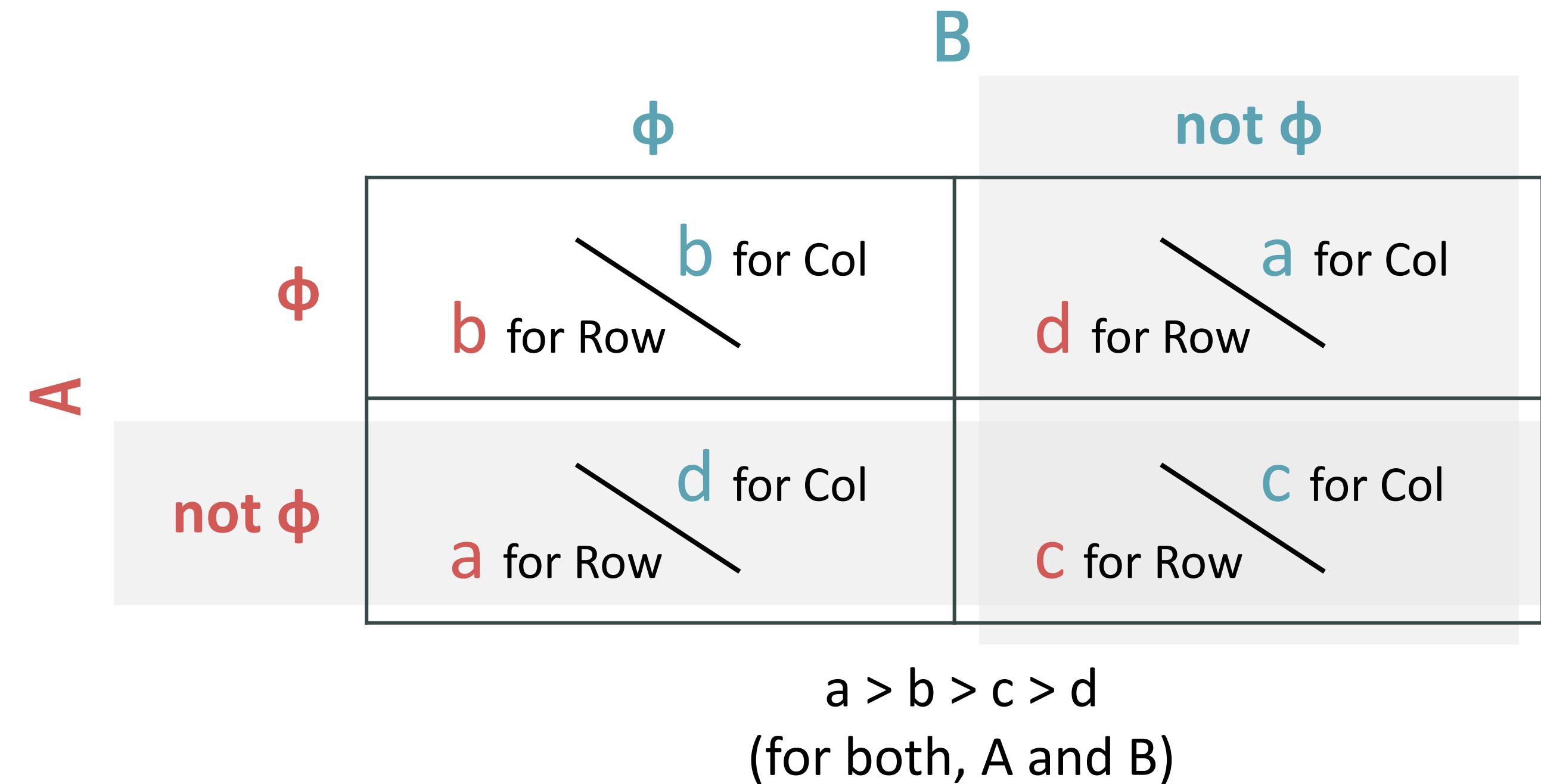
		Col	
		remain silent	make a deal
		remain silent	make a deal
Row	remain silent	5 years 5 years	free 25 years
	make a deal	25 years free	20 years 20 years



- Making a deal is the dominant strategy (for both).
- It is rational for Row to make a deal (the same holds for Col)

Problem: If both remain silent, both would be better off. That seems odd at least!

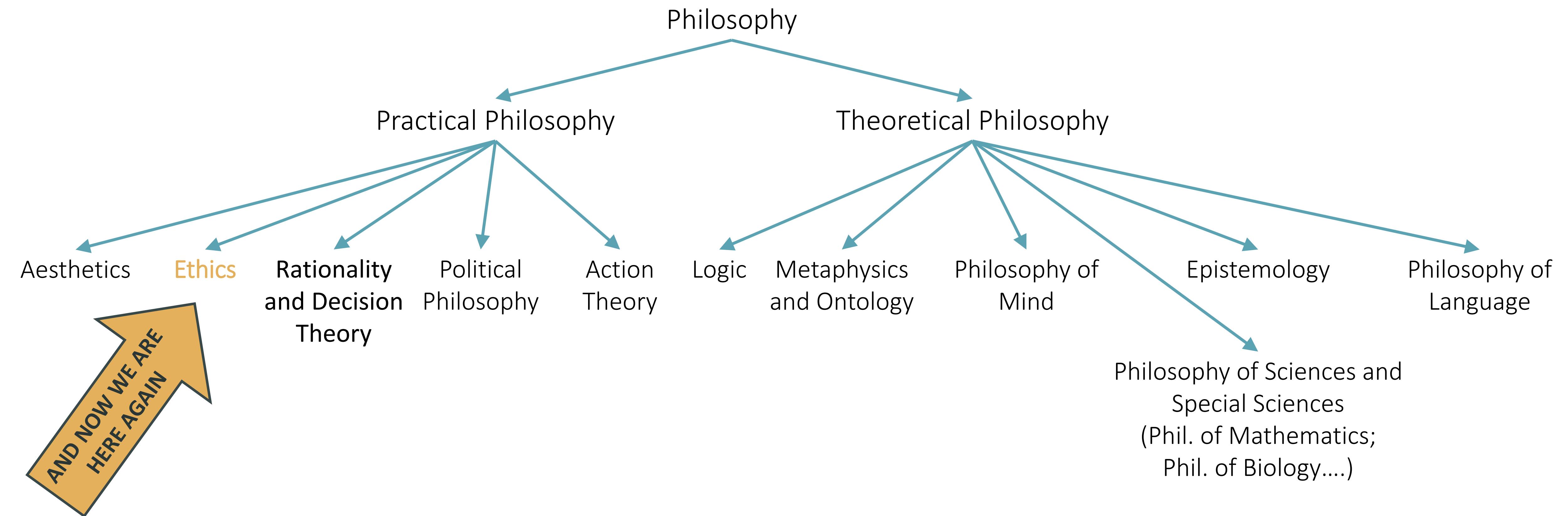
COLLECTIVE ACTION AND RATIONALITY: THE PRISONER'S DILEMMA



- Not ϕ -ing is the dominant strategy.
- It is rational for A to not ϕ (the same holds for B).

Problem: If both ϕ , both would be better off. That seems odd at least!

RECAP: FIELDS OF PHILOSOPHY



A TALE OF TWO FACTORIES

Collective Action Problems
based on Cooperation Cases

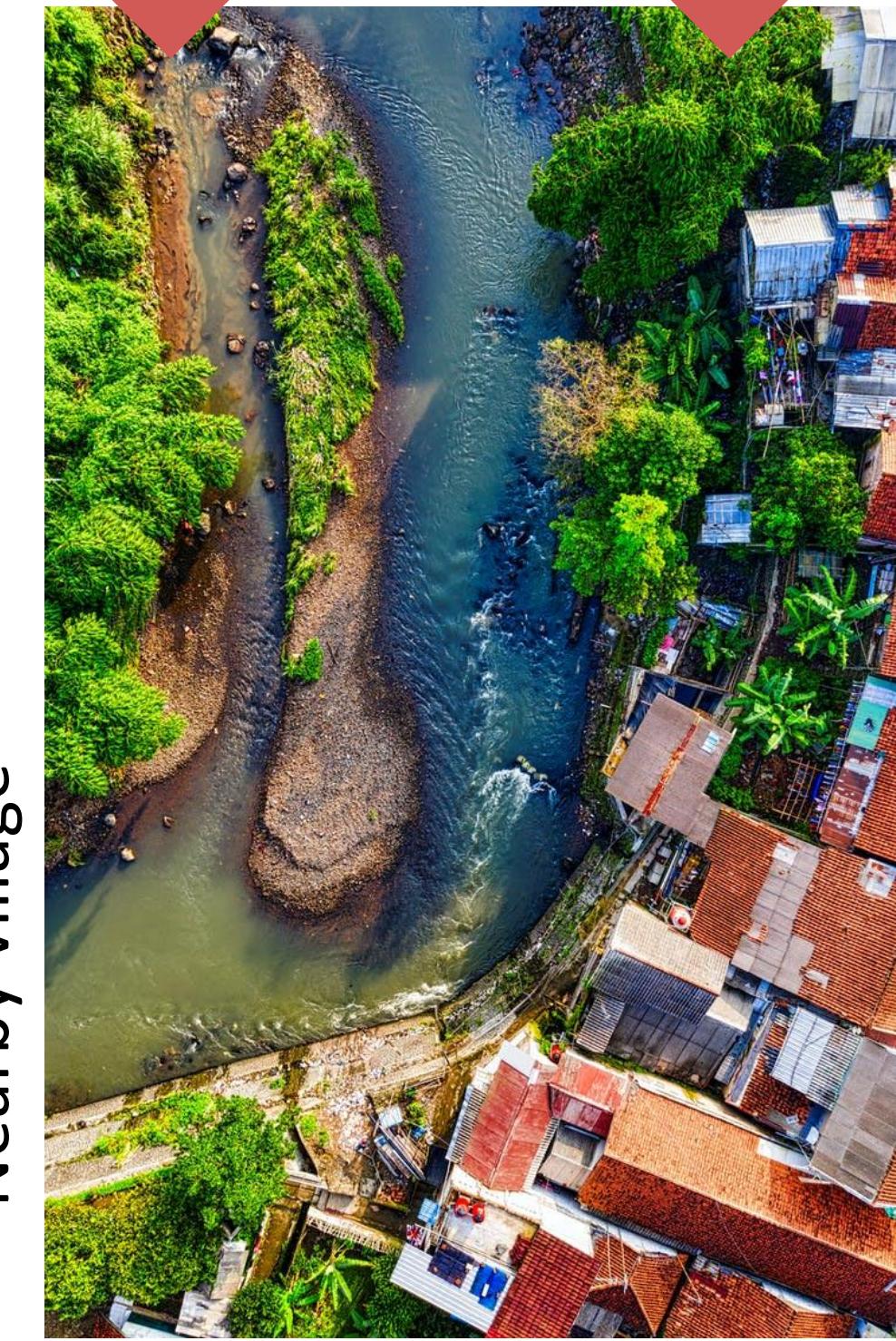


Factory A

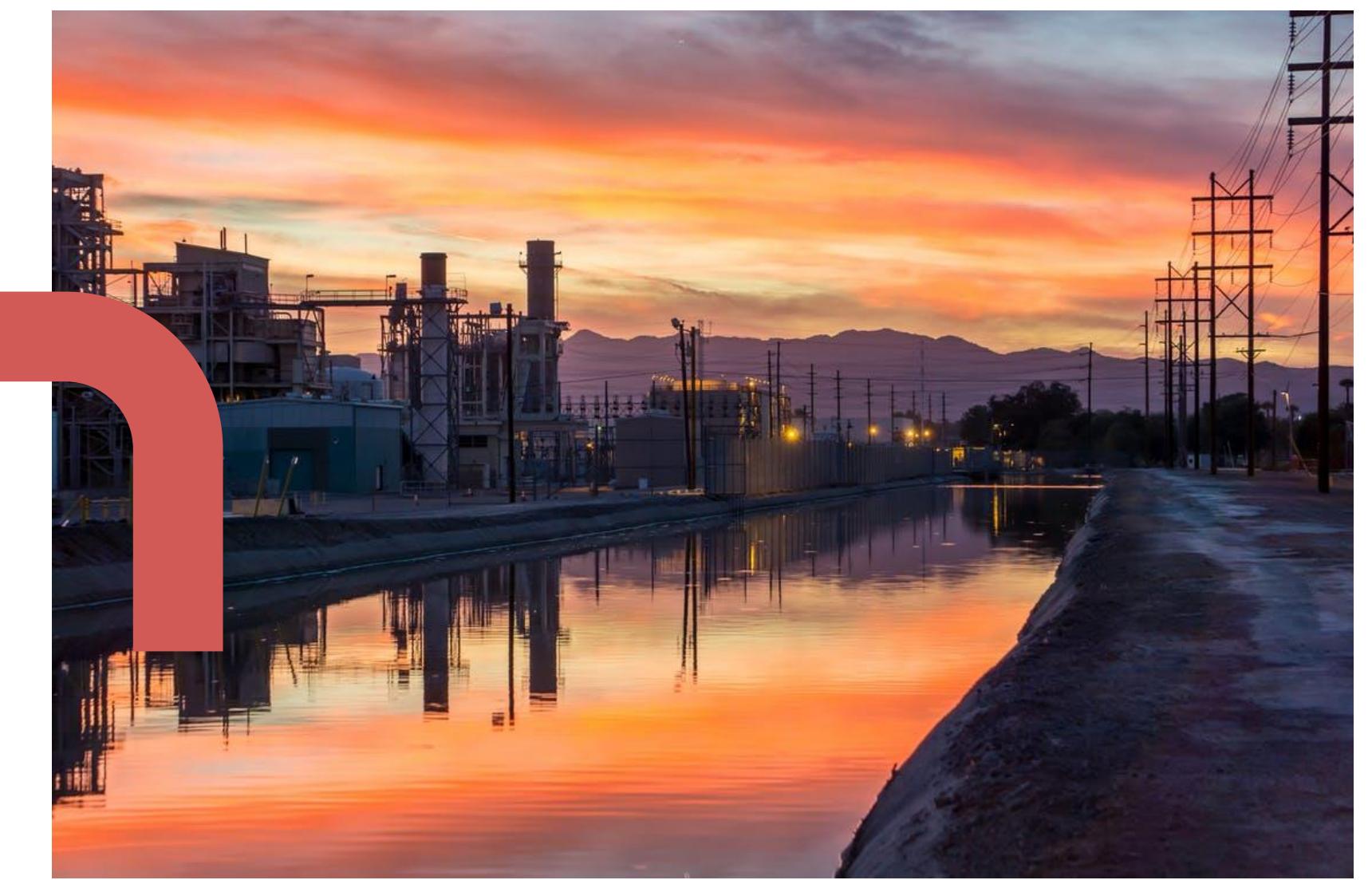
- pollution kills all the fish in a river
- one polluting company suffices for that
- only jobs in the area are at company A, company B and being a fisherman



Nearby village



poisonous
waste



Factory B

- polluting is a competitive advantage
- if one company pollutes, but the other company does not pollute, the non-polluting company will go bankrupt
- if both do not pollute or both do pollute, then both companies can flourish

Collective Actions (concerned with morals again)

Coordination/Cooperation Variant

		competitor B	
		produce cleanly	pollute
competitor A	produce cleanly	A's workers ✓ B's workers ✓ fishermen ✓	A's workers X B's workers ✓ fishermen X
	pollute	A's workers ✓ B's workers X fishermen X	A's workers ✓ B's workers ✓ fishermen X

Collective Actions (concerned with morals again)

Coordination/Cooperation Variant

		B	
A	ϕ	ϕ	$\text{not } \phi$
ϕ		+++	- - -
$\text{not } \phi$		- - -	-

Principle of Strict Strategic (Moral) Dominance:

If there is one strategy for player A (say, performing ϕ) is better independent of whatever strategy player B chooses, then it is morally right for A to play that strategy (i.e., to ϕ).

does not apply here...

Collective Actions (concerned with morals again)

Coordination/Cooperation Variant

		B	
	φ	φ	not φ
A	φ	+++	- - -
	not φ	- - -	-

Assume that both know/that it is a fact: “Whatever I will do, the other one will not ϕ .”

Collective Actions (concerned with morals again)

Coordination/Cooperation Variant

		B	
		ϕ	not ϕ
		A	B
ϕ	ϕ	+++	- - -
	not ϕ	- - -	-

Assume that both know/that it is a fact: “Whatever I will do, the other one will not ϕ .”

→ They are led by consequentialism to the second worst result.

Principle of Moral Harmony:

A moral theory, T , is correct if and only if the agents who satisfy T , whoever and however numerous they may be, are guaranteed to produce the morally best world that that they could together bring about

(Douglas Portmore. *Maximalism and Moral Harmony* in Philosophy and Phenomenological Research, Volume 96, Issue2, 2016, p. 13)

Collective Actions (concerned with morals again)

Coordination/Cooperation Variant

		B	
		ϕ	not ϕ
		A	B
ϕ	ϕ	+++	- - -
	not ϕ	- - -	-

Assume that both know/that it is a fact: “Whatever I will do, the other one will not ϕ .”

- They are led by consequentialism to the second worst result.
- If consequentialism embraces the Principle of Moral Harmony, then consequentialism fails in its own lights!

Collective Actions (concerned with morals again)

Coordination/Cooperation Variant

I. THE TWO FACTORIES

Ann and Ben are owners of two factories which are located opposite each other on a river. Both agents opt for a production process which releases waste chemicals into the river and thereby kill all the fish in the river and destroy the livelihood of a fishing community downstream. The waste from one factory alone would suffice to kill all the fish, and adding the waste from the other factory does no additional damage whatsoever (say, the river flows into the sea where the waste is diluted below any harmful concentration). If Ann or Ben were to unilaterally produce cleanly, this would make their production uncompetitive compared to the other factory, put them out of business, and destroy the livelihood of their employees. However, if they both were to produce cleanly, then this problem would not arise, and both factories would remain in business and the fishing community would flourish. Ann and Ben each employ 100 workers, the fishing community counts 100 people, and all that matters morally in this case are the livelihoods of the workers and fishermen. Further, the only available actions are either to pollute or to produce cleanly. In particular, Ann and Ben cannot come together and suggest and discuss a common strategy.⁶ So far, the case is represented by figure 1.⁷

What If I Cannot Make a Difference (and Know It)*

from *Ethics* 123, pp. 971–998 , 2015

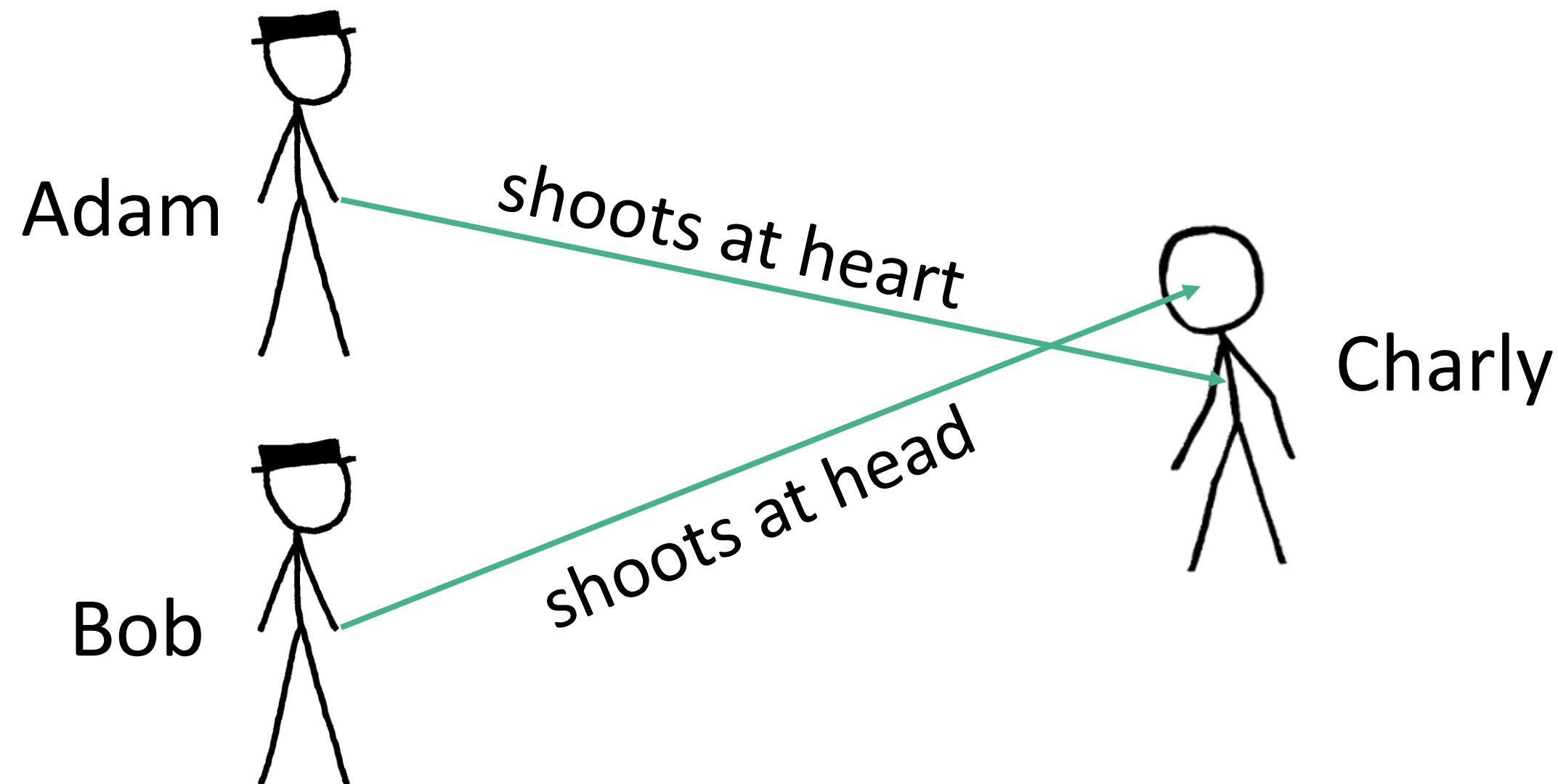
Felix Pinkert

		Ben	
		φ_c	φ_p
		all workers and fishermen have their livelihood (300)	only Ben's workers have their livelihood (100)
Ann	produce cleanly (φ_c)	all workers and fishermen have their livelihood (300)	only Ben's workers have their livelihood (100)
	pollute (φ_p)	only Ann's workers have their livelihood (100)	only Ann and Ben's workers have their livelihood (200)

Collective Actions

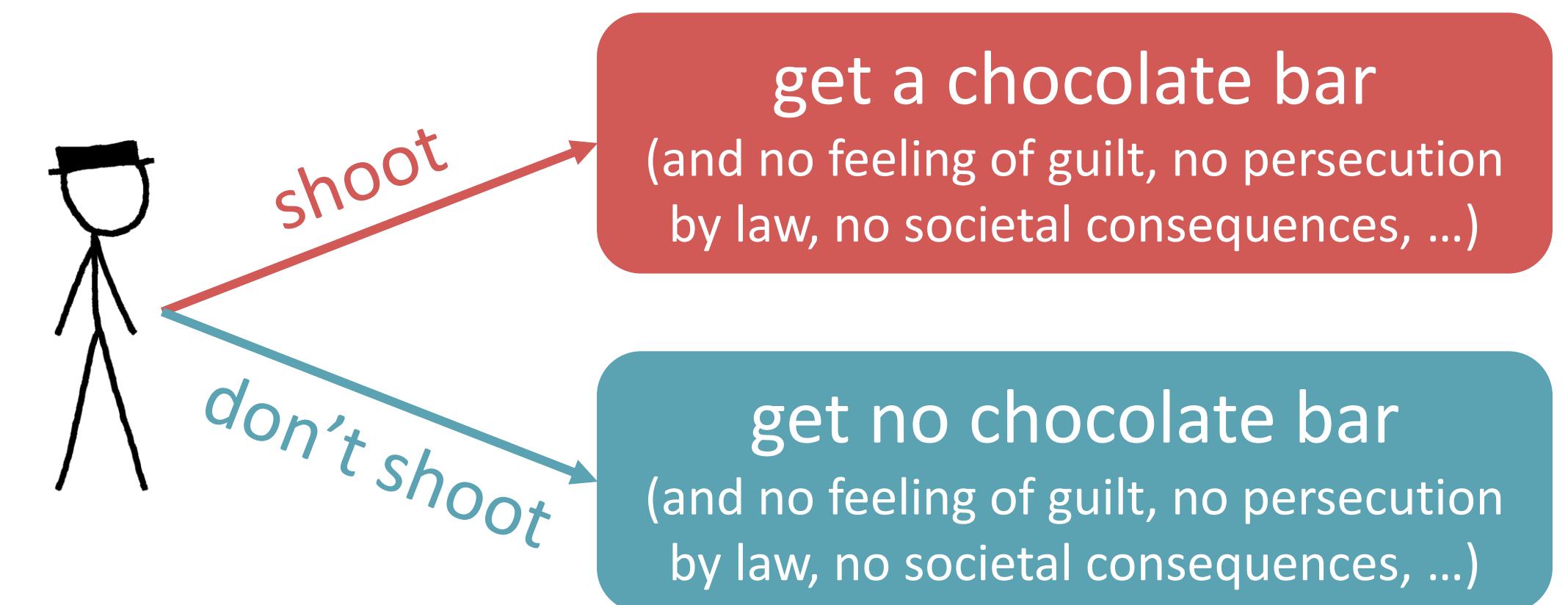
Overdetermination Variant

Adam and Bob are about to shoot Charly, each with one deadly shot:



- both are sure the other one will shoot
- and they are justified in this belief

Both Adam and Bob have the same choice to make:

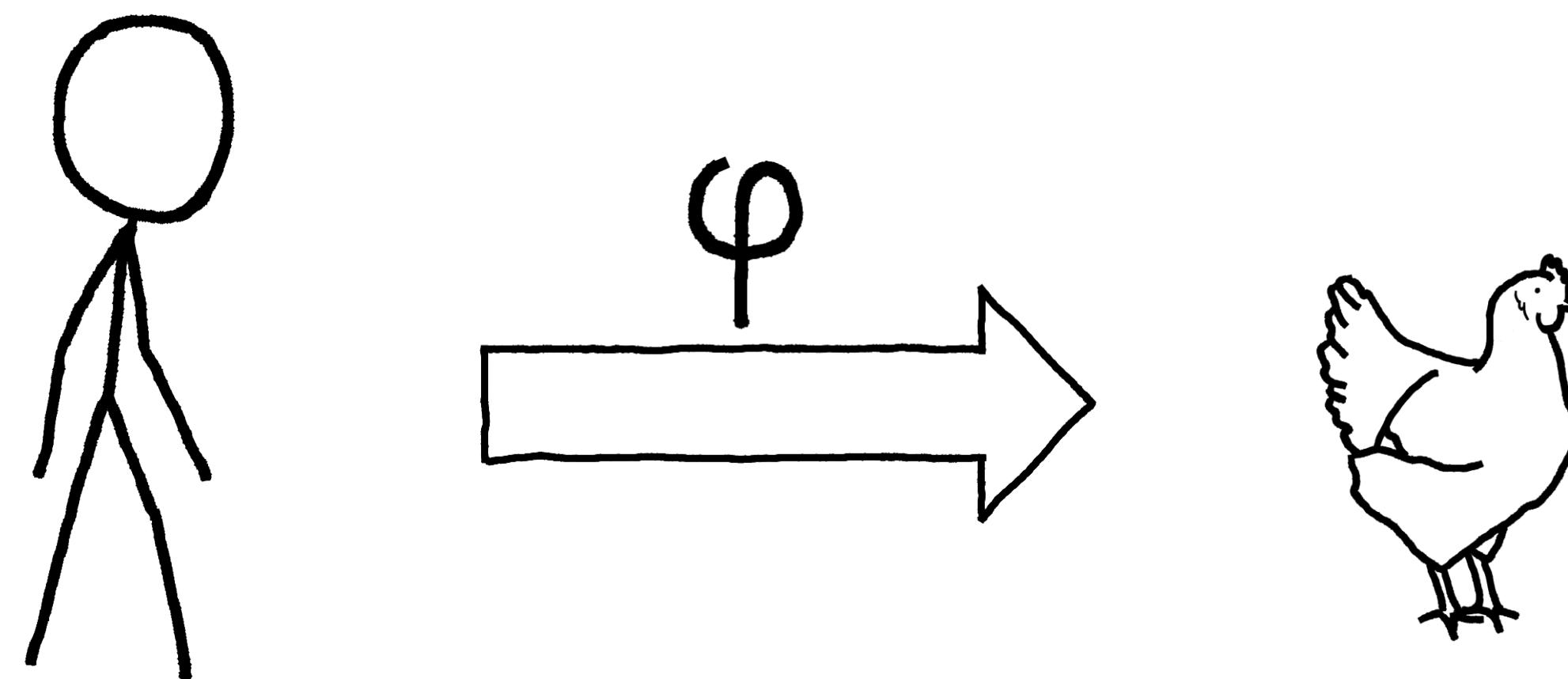


→ Naive forms of consequentialism will tell each of them to shoot, because Charly will die either way

Collective Actions

Overdetermination Variant

Should you become a vegan/vegetarian?
(Chicken Counter Case)



Collective Actions

Cumulative Variant

Any individual action doesn't make things worse,
but in total, the sum of actions have very bad results.

But because each individual action does not make things worse, standard forms of Consequentialism are unable to judge them wrong.



Examples: global warming, smog, pollution...

Possible Reaction

- Deny the (conceptual) possibility of cases of the three different types.
- Give up the Principle of Moral Harmony (or show that it was never/should have never been embraced by consequentialism).
- Deny that consequentialism leads to the sub-optimal outcomes (or, better, show that it led to the optimal outcomes, otherwise you might still run into problems, i.e., normative gaps).
- Find a consequentialist theory that does not run into the troubles.
- ...

