Potsdam Institute for Climate Impact Research

GaNe
FutureLab on Game Theory & Networks of Interacting Agents

Measures of individual and groupwise ex-post and ex-ante responsibility in extensive-form games with unquantifiable uncertainty

work in progress by **Jobst Heitzig** & **Sarah Hiller**

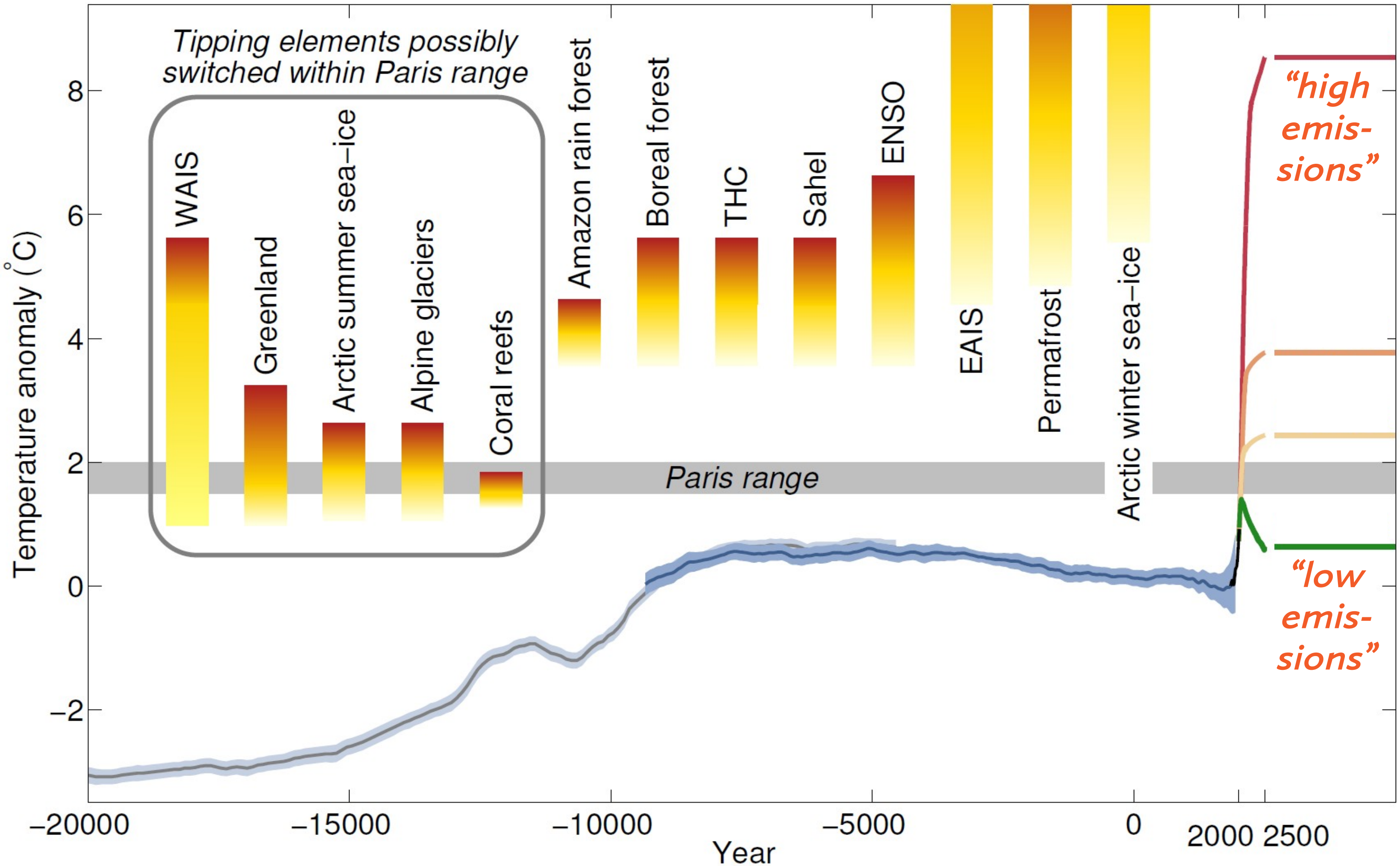
Formal Ethics 2019
Ghent, June 2019

Leibniz-Gemeinschaft

Freie Universität Berlin

Foretaste

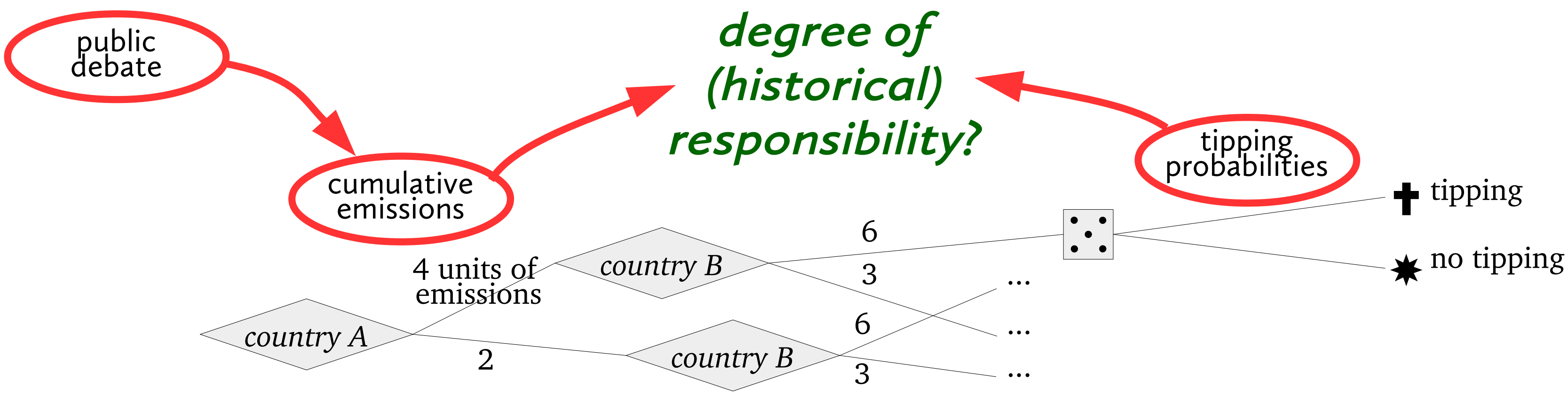
Climate tipping elements, Paris Agreement



Schellnhuber et al., Nature Climate Change 2016

Simplistic Example: Triggering a climate tipping

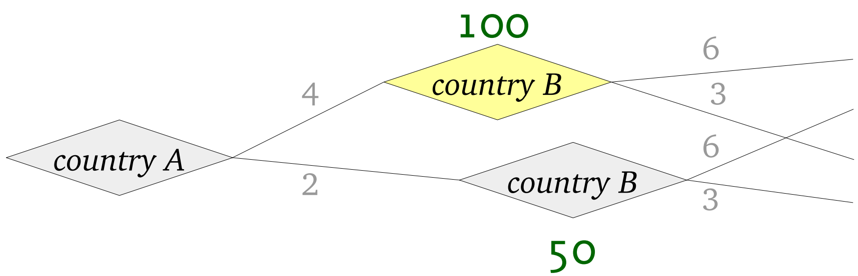
Country *A* chooses high or low greenhouse gas emissions,
then country *B* chooses high or low emissions,
then unwanted climate tipping either occurs or not



Use Probabilities; be Optimistic about Influence

Country *A* chooses high or low greenhouse gas emissions,
then country *B* chooses high or low emissions,
then unwanted climate tipping either occurs or not

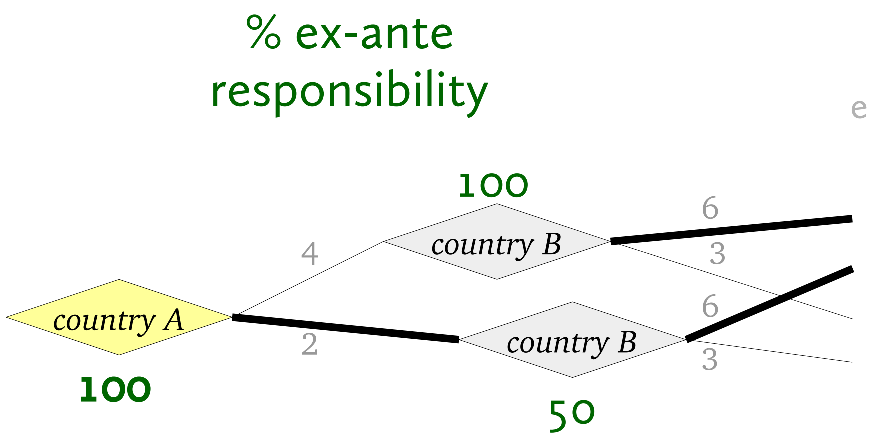
% ex-ante responsibility



cum. global emissions	% prob. of tipping	% ex-post responsibility B
10	100	100
8	50	50
7	0	0
5	0	0

Use Probabilities; be Optimistic about Influence

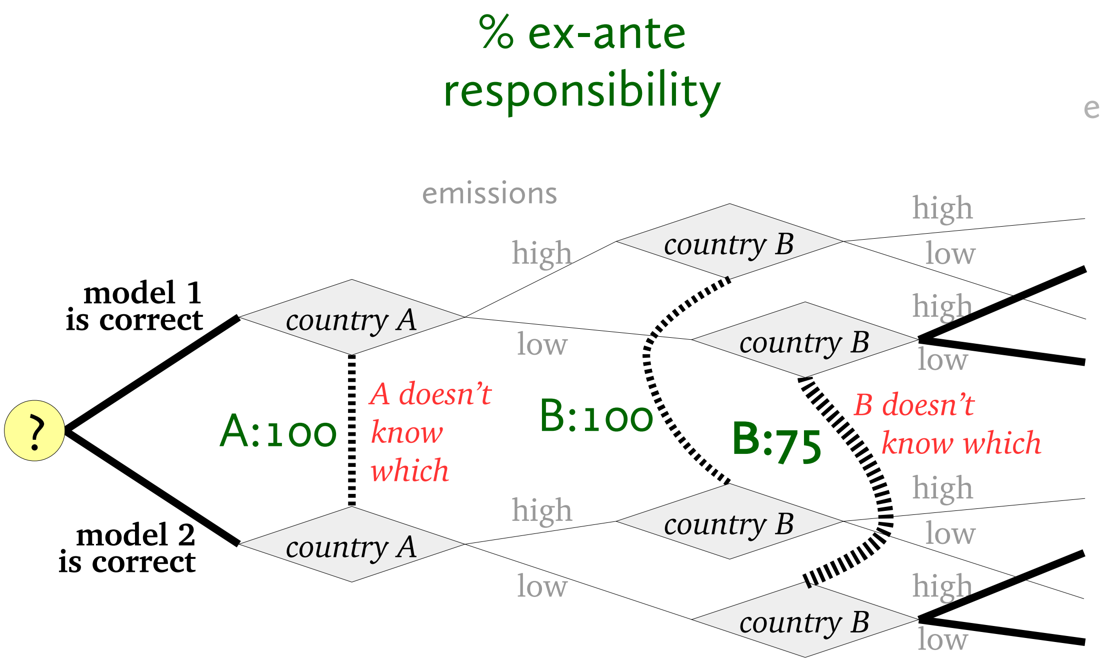
Country *A* chooses high or low greenhouse gas emissions,
then country *B* chooses high or low emissions,
then unwanted climate tipping either occurs or not



cum. global emissions	% prob. of tipping	% ex-post responsibility A	B
10	100	100	100
8	50	50	50
7	0	100	0
5	0	50 (?)	0

Be pessimistic about Unknowns

Country *A* chooses high or low greenhouse gas emissions,
then country *B* chooses high or low emissions,
then unwanted climate tipping either occurs or not
+ uncertain probabilities



cum. global emissions	% prob. of tipping	% ex-post responsibility A	% ex-post responsibility B
10	100	100	100
8	50	50	**75**
7	0	100	0
5	0	50	0
12	100	100	100
10	**100**	50	**75**
9	75	100	0
7	**25**	50	0

The *Max-diff* -Formula for *Ex-Ante* Responsibility

Verbal definitions:

A ***scenario S for agent group G at node v*** is a choice of branch for each unquant. uncertainty node and all other agents' information states in the branch rooted at *v*.

A ***strategy s for agent group G at node v*** is a choice of action for each of *G*'s information states in the branch rooted at *v*.

The ***conditional value of G's strategy s at v given S*** is a certain strictly increasing function f (e.g., f(P) = P or f(P) = logit P) of the probability P, evaluated at *v*, of a good outcome, conditional on *S* and *s*.

G's conditional influence at v given S is the difference between the largest & smallest conditional values of all of *G*'s strategies in *S* at *v*.

G's degree of ex-ante responsibility at v is its maximum conditional influence at *v* over all possible scenarios at *v*.

Formally:

$$\mathrm{ear}(G,v) = \max \{ \max \{ f(\, P(\mathrm{good}|v,S,s)\,) : \text{strat. } s \text{ for } G \text{ at } v \} - \min \{ f(\, P(\mathrm{good}|v,S,s)\,) : \text{strat. } s \text{ for } G \text{ at } v \} : \text{scenario } S \text{ for } G \text{ at } v \}$$

Introduction

Random Citation:

*"Whether humans are responsible **for** the bulk of climate change is going to be left to the scientists, but it's all of our responsibility **to** leave this planet in better shape for the future generations than we found it."*

(Mike Huckabee, US Republican)

Sort of Questions We Care about Here:

- Am I ethically **responsible *for***
  climate change / Katrina / Bob's homelessness / etc.?
  And in what sense? And to what degree?
- Do I have ethical **responsibility *to***
  mitigate climate change / compensate victims / etc.?
  And to what degree? And to what length or amount?
- How can responsibility be **quantified** in view of **many agents** & different forms of **uncertainty**?

Strategy:
study examples → formulate theses → suggest formulae

Vague Initial Working Definition of "More or Less Responsible for/to…"

Ex-post (=backward-looking) responsibility:

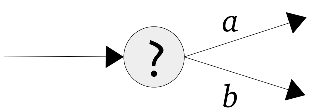
An agent i is the more ***responsible for*** *a (typically ethically undesired) factual outcome q the more i could have exerted influence to prevent q.*

Ex-ante (=forward-looking) responsibility:

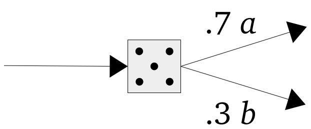
Given that outcome q is considered ethically undesirable, *agent i has the more* ***responsibility to*** *help prevent q the more potential influence i can exert to prevent q.*

Toolbox &
First Example

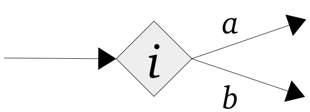
Tree-Shaped Models of Multiagent Decisions under Uncertainty



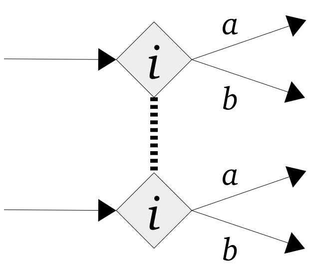
unquantified uncertainty node with branch labels
modeller believes these are all possible subbranches but has no idea whatsoever about their probabilities



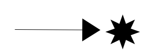
stochastic node with probabilities and branch labels
modeller believes these are all possible subbranches and that these are their probabilities
(different stochastic nodes are considered to be independent random events)



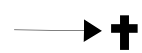
decision node with action labels for agent i
modeller believes agent i has exactly these options and will decide among them at free will
and that the probabilities of her choosing each option are unknown



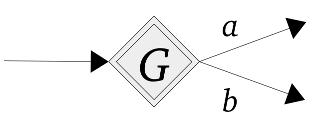
information state
with two decision
nodes for agent i
modeller believes agent i will not know
in which of the two nodes she is



good outcome node



bad outcome node

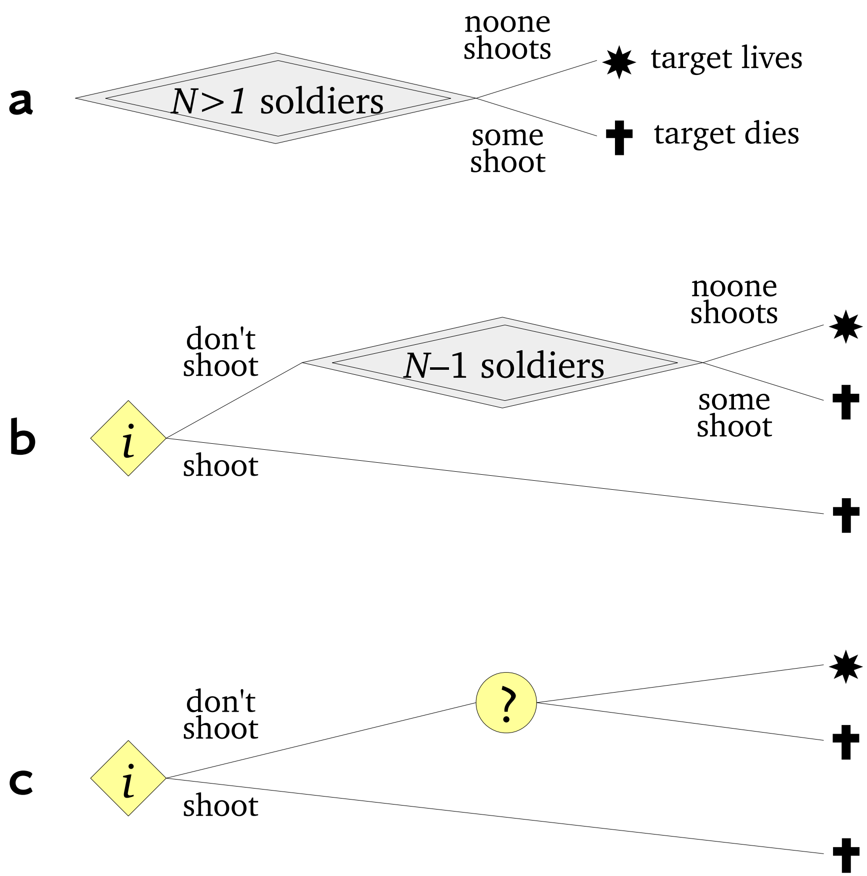


summarizing decision node with options for agent **group** G
modeller believes group G has exactly these options and modeller is not interested in more detail

(enriched version of extensive-form game trees)

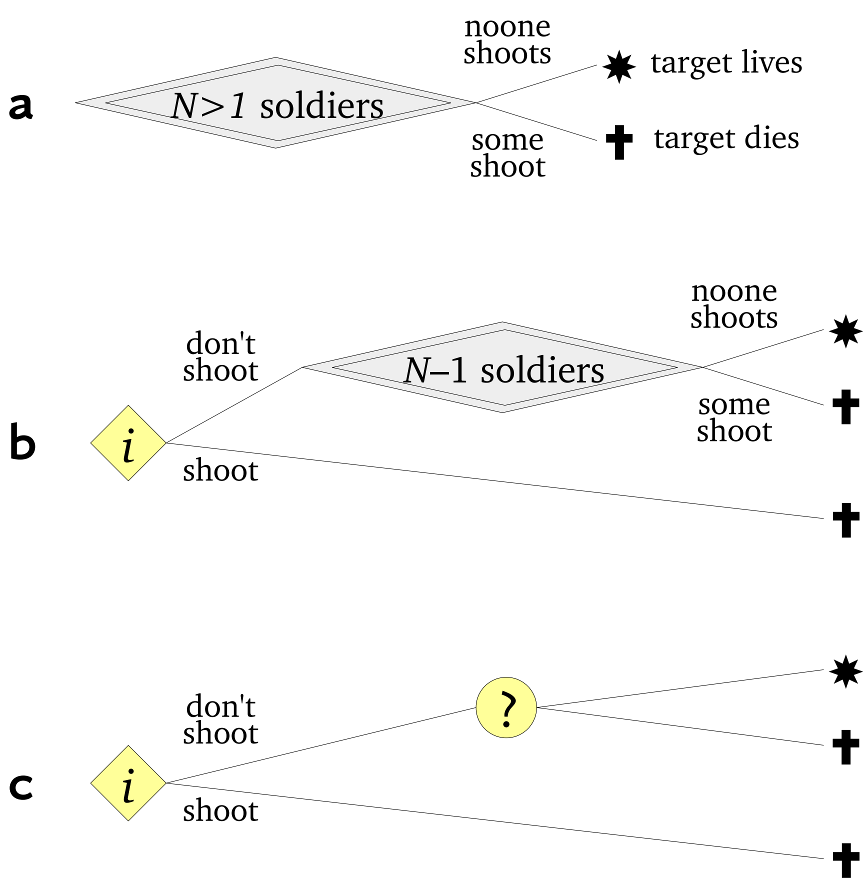
Extreme Example 1: Killing by Fire-Squad

Different models of the situation:



Extreme Example 1: Killing by Fire-Squad

Different models of the situation:



Thesis 1:
a,b,c are all *equivalent* w.r.t. the assessment of i's responsibilities! In particular, the number N is irrelevant (if > 1)

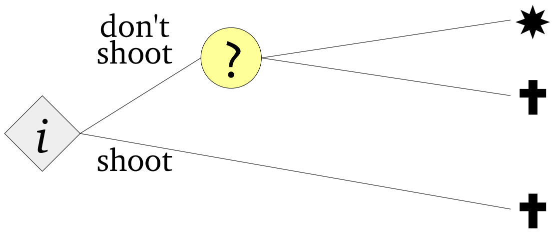
Thesis 2:
i has, ethically, *full "ex-**ante**" responsibility* for what the result will be.

Thesis 3:
If i shoots, she has also *full "ex-**post**" responsibility* for the result. If not, she has none.

Example 1 (contd.): "Ethical" vs. "Psychological" Assessment

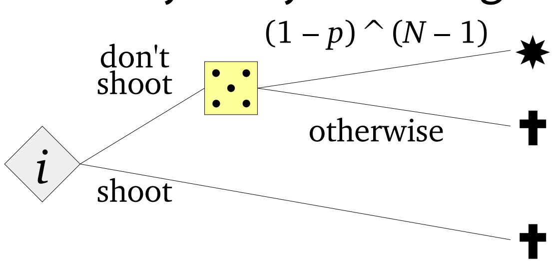
"Ethical observer's" model: i **cannot know** with what probability the others will shoot

c



i might *prefer* the "psychological" model: each other soldier shoots with **probability** p, thus it's very likely that target dies anyway

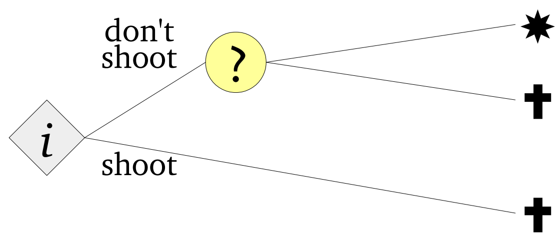
d



Example 1 (contd.): “Ethical” vs. “Psychological” Assessment

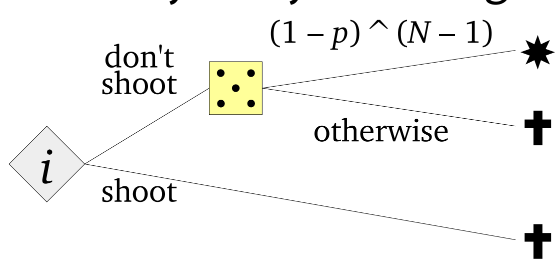
“Ethical observer's” model: i **cannot know** with what probability the others will shoot

c



i might *prefer* the “psychological” model: each other soldier shoots with **probability** p, thus it's very likely that target dies anyway

d



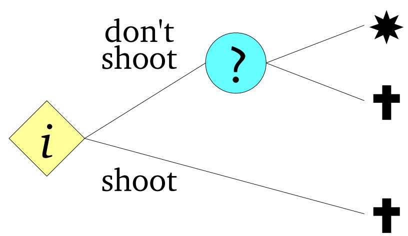
Thesis 4:
i cannot rightfully claim to know the probabilities that the others shoot, hence the *proper model* is **c** and not **d**.

Thesis 5:
In other situations, where **d** *would be* the proper model, i would only have *partial responsibility* of degree $(1-p)$ ^ $(N-1) \ll 1$ (resulting in exponentially fast “diffusion of responsibility”)

WHAT ROLE DOES TIMING PLAY?

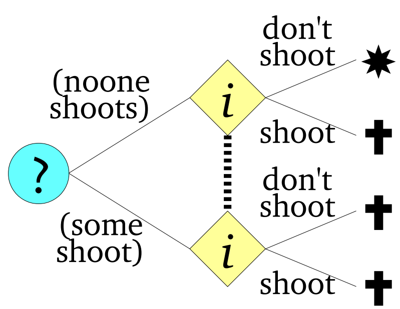
i decides "first":

c



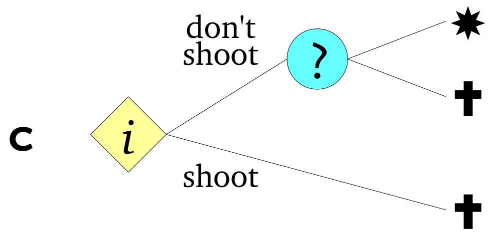
(some) others decide "first",
but *i* doesn't know how
(dashed line):

c'

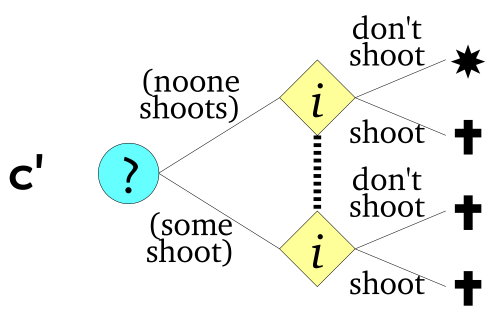


Example 1 (contd.):
What Role does Timing Play?

i decides "first":



(some) others decide "first",
but *i* doesn't know how
(dashed line):

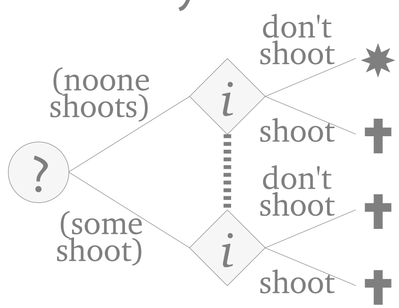


Thesis 6:
Such timing issues
have no effect on *what i can know*,
hence they are irrelevant here.
c,c' are in this sense equivalent.

Example 1 (contd.): What if the Gun Might not be Loaded?

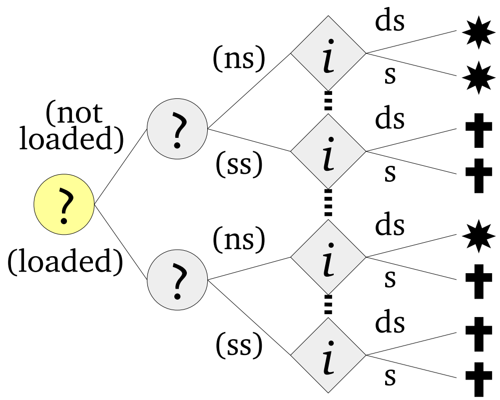
Gun certainly loaded:

c′



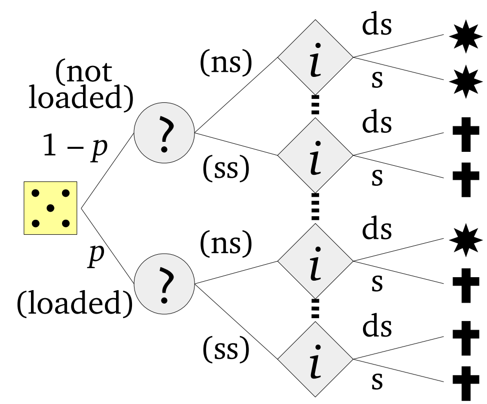
Loaded with unknown probability:

e



Loaded with known prob. p:

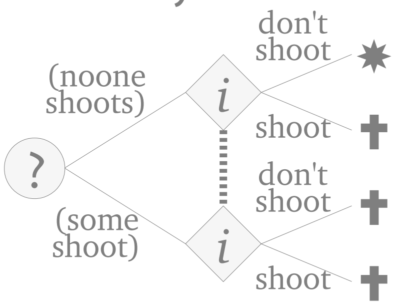
f



Example 1 (contd.):
What if the Gun Might not be Loaded?

Gun certainly loaded:

c'

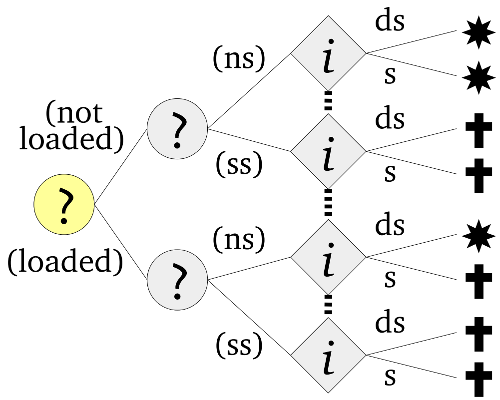


Thesis 7:
In **e**, *i must consider the possibility* that the gun might be loaded and has thus *full ex-ante responsibility.*

Thesis 8:
In **f**, *i knows to have limited influence*
→ *partial ex-ante responsibility* of *degree p*

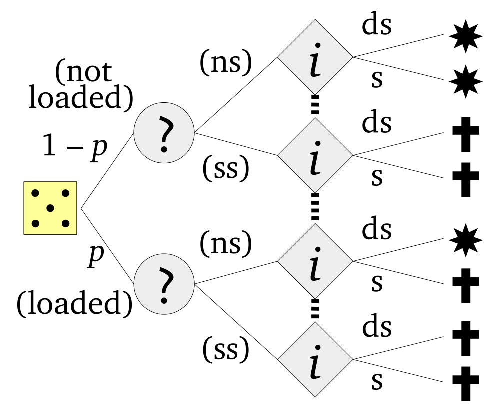
Loaded with
unknown probability:

e



Loaded with
known prob. p:

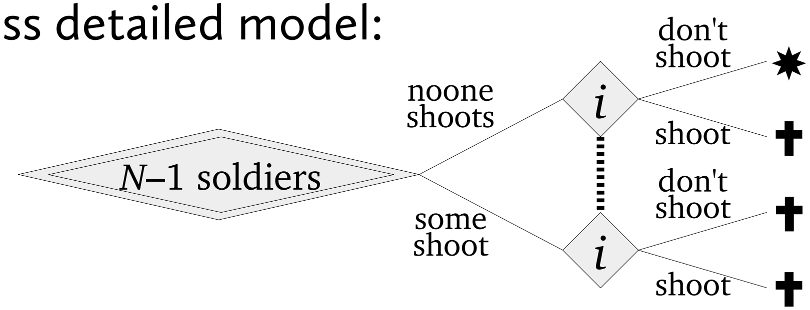
f



Example 1 (contd.): Whose Bullet was it?

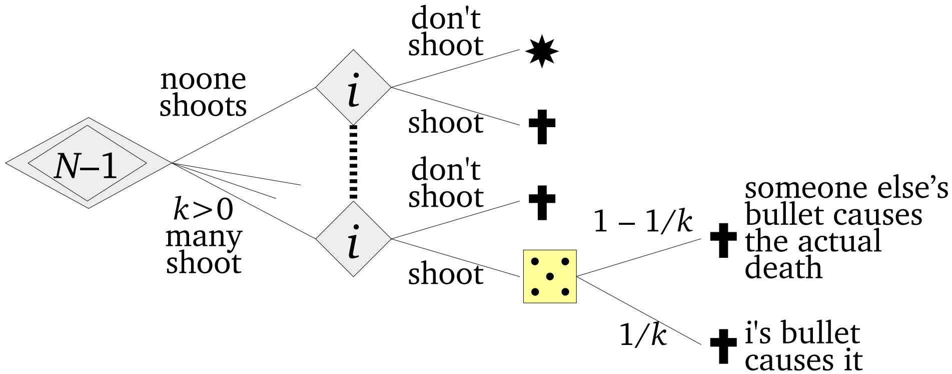
Less detailed model:

b'

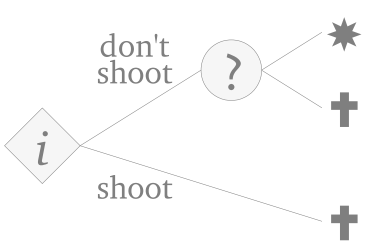


More detailed model:

b''



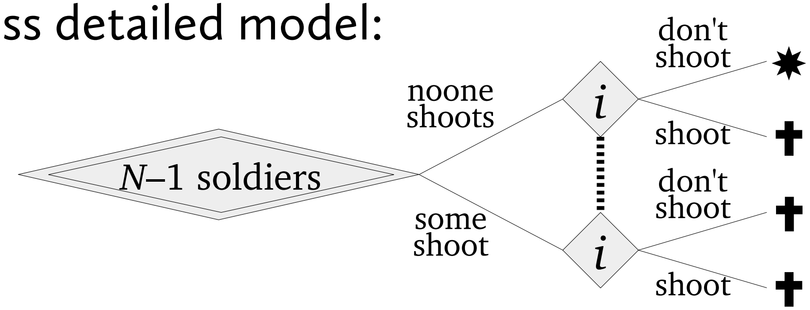
c



Example 1 (contd.): Whose Bullet was it?

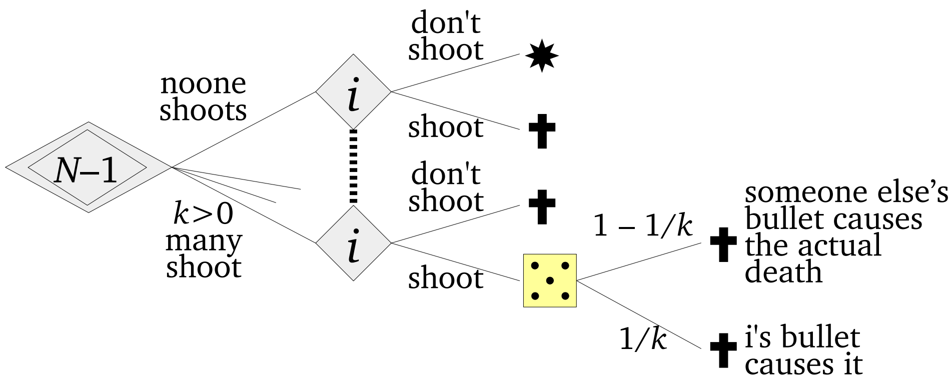
Less detailed model:

b'

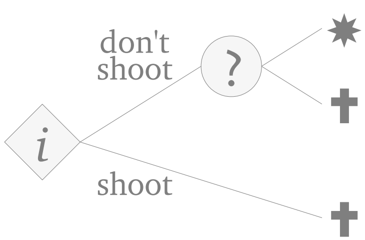


More detailed model:

b''



c



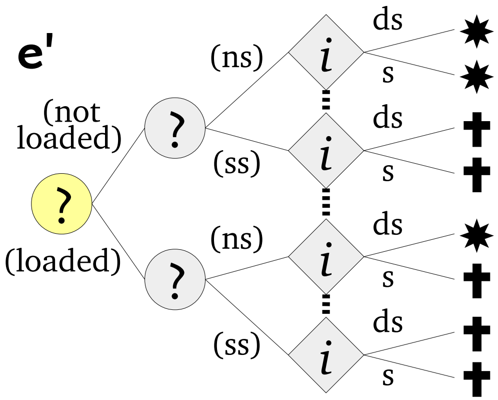
Thesis 9:
If each shot bullet *would* have killed, it doesn't matter which bullet *did* kill.
b'' is still *equivalent* to **b'** (and **b,c**) w.r.t. the assessment of i's ex-post responsibility.

(So if i shoots, she has full ex-post responsibility for the actual result.)

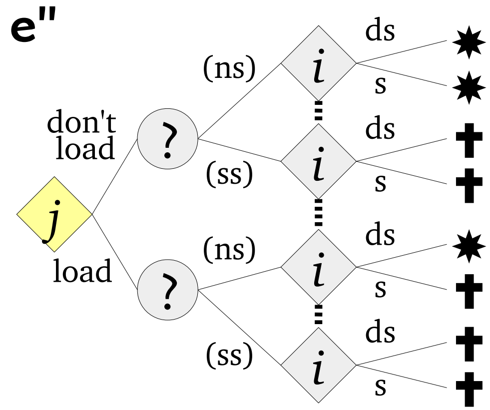
→ *Factual causes* matter less than *potential consequences and their probabilities*

“If you hadn't loaded the gun...” (Still Example 1)

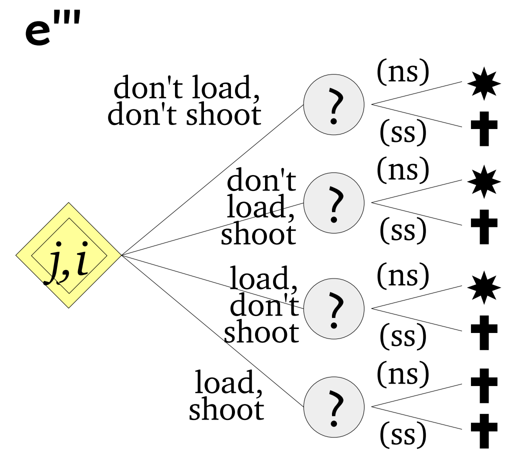
We had this model:



More specific model:



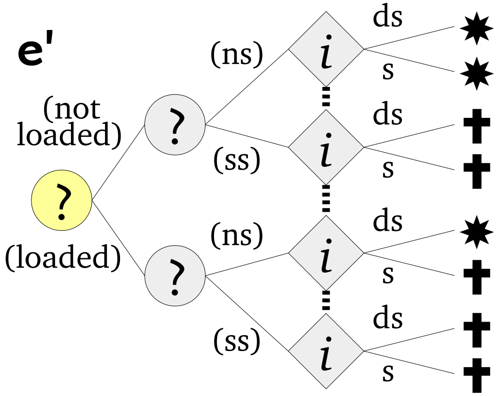
Less detailed model, treating $\{i,j\}$ as a group:



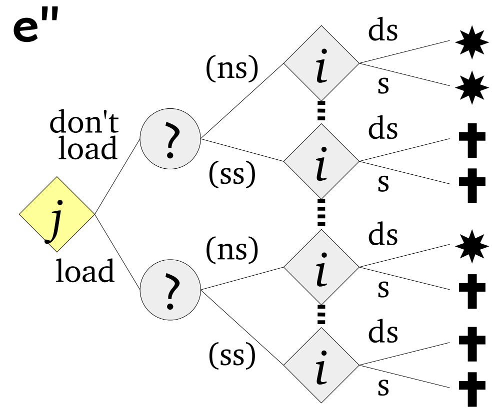
Is j less responsible than i?

"If you hadn't loaded the gun..." (Still Example 1)

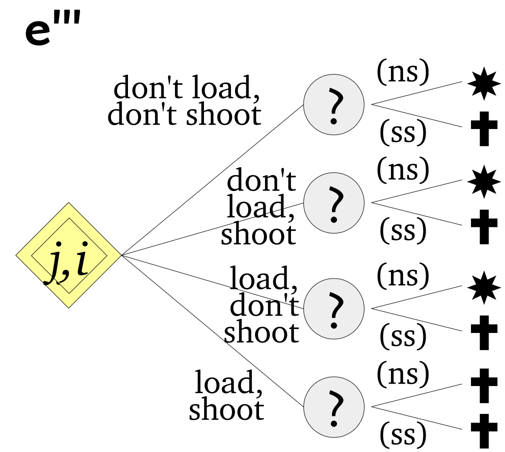
We had this model:



More specific model:



Less detailed model, treating {*i*,*j*} as a group:



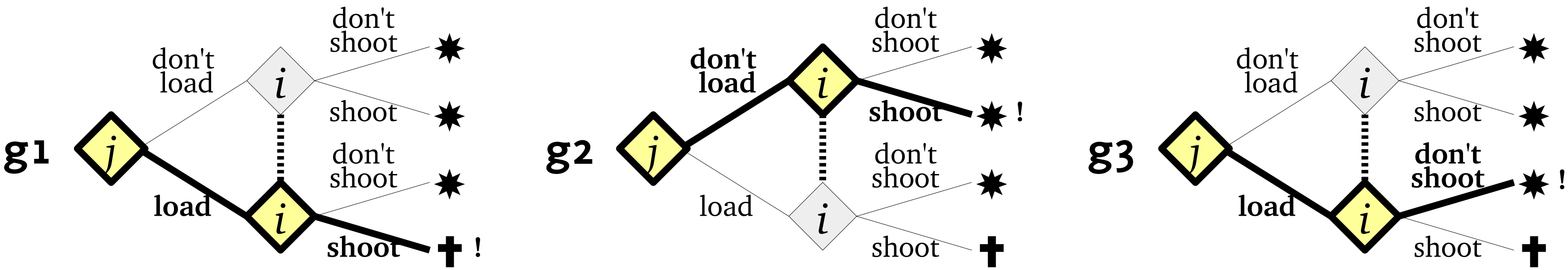
Is j less responsible than i?

Thesis 10: *i must consider the possibility* that *j* has loaded the gun,
but ***likewise*** *j must consider the possibility* that *i* will shoot,
hence ***both*** *have full ex-ante responsibility (i.e., "degree 1").*
Also the group {i,j} has resp. degree 1 < 1+1, hence responsibility is *not additive.*

(Note: some authors focussing on attribution consider a group only responsible if no smaller group is responsible)

Factual vs. *Counterfactual* ex-post Responsibility

Different *realizations* of the same, simplified model:

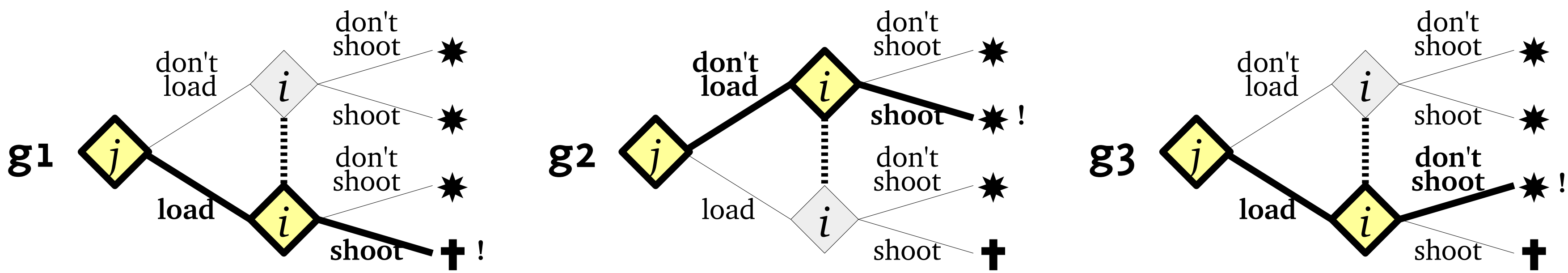


Both i and j are *fully ex-ante* responsible to avoid the target's death.
In g1, both i and j are *fully ex-post* responsible for the factual death.

Should either of i or j be blamed in g2 or g3?

Factual vs. *Counterfactual* ex-post Responsibility

Different *realizations* of the same, simplified model:



Both i and j are *fully ex-ante* responsible to avoid the target's death.
In g1, both i and j are *fully ex-post* responsible for the factual death.

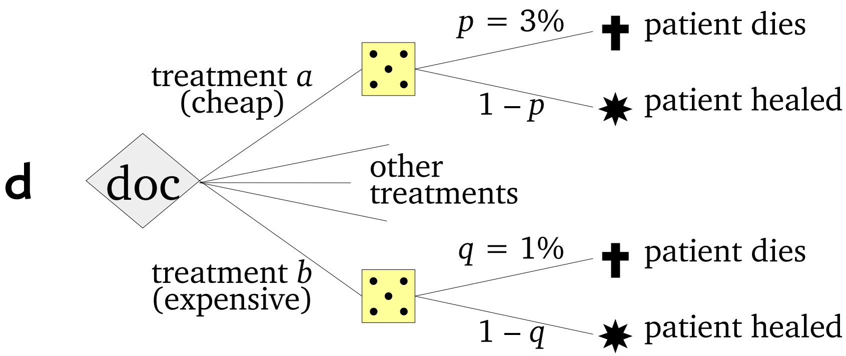
Should either of i or j be blamed in g2 or g3?

Thesis 11:
In **g2**, since the factual outcome is good, *i is not factually responsible* for a bad one, still *i's action did make a bad outcome **possible***, so *i* should be blamed for this:
*i is fully **counterfactually** responsible regarding the target's possible death.*

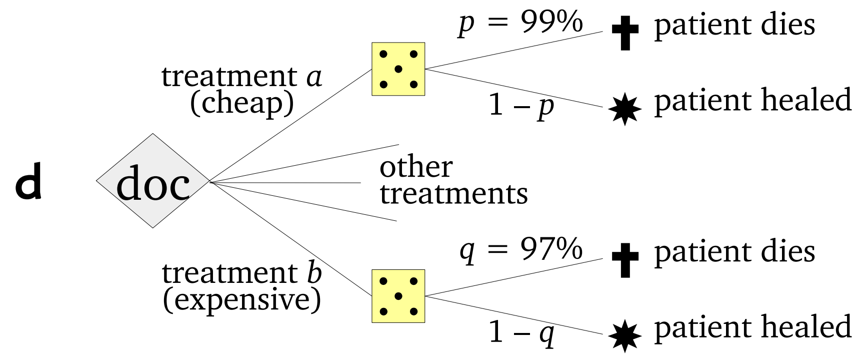
(and similarly *j* in g3)

Example 2: Risk of Death in Medical Treatment

Low overall risk of death:

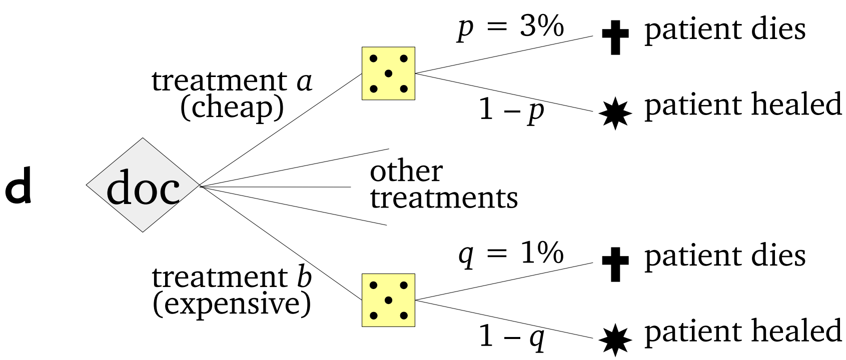


High overall risk of death:

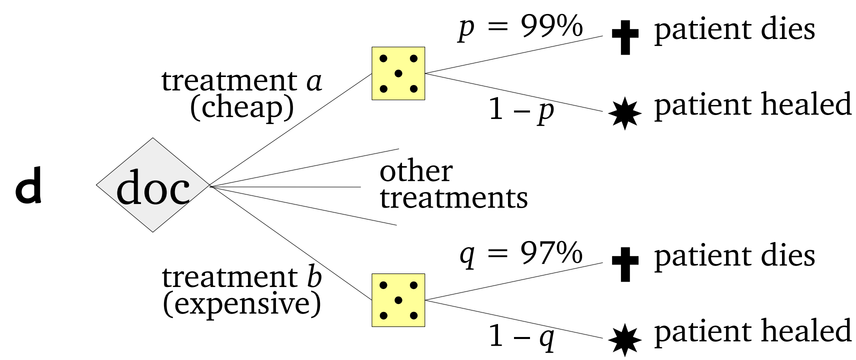


Example 2: Risk of Death in Medical Treatment

Low overall risk of death:



High overall risk of death:



Thesis 12:
Doctor's degree of ex-ante responsibility regarding the patient's possible death is a certain *function f(p,q) of the two probabilities* p *and* q, which increases with larger p and decreases with larger q (assuming $p > q$).

Maybe simply f(p,q) = p – q = 3% – 1% = 99% – 97% = 2% ?
Or should f (3%,1%), f (51%,49%), and f (99%,97%) all differ, and if so, how?

Suggested general principles & some candidate formulae

"[...] it seems rational to [...] concentrate [...] on the actual decision in light of the probabilities."
(Nagel 1979: Moral luck)

Principles for Measuring Responsibility (I)

derived from the preceding theses (nos. in brackets)

Unquantified uncertainty:

- ***Others' unknown choices*** (if at free will) should be treated as *unquantifiable uncertainty*, and not as stochasticity with some assumed probabilities. (1.+4.)
- Facing unquantified uncertainty, one must consider that ***the "best" branch may have probability one.*** (2.)
- Facing unquantified uncertainty, one must also consider that ***the "worst" branch may have probability one.*** (7.)

Stochastic uncertainty:

- Actual stochasticity may reduce the ***degree*** of responsibility in dependence of the resulting *probabilities*. (5.,8.,12.)

Principles for Measuring Responsibility (II)

Three forms of responsibility:

"Forward-looking": ***options*** *(potential actions) may lead to*

- ***ex-ante*** responsibility. (2.)

"Backward-looking": ***choices*** *(factual actions) may lead to*

- ***factual ex-post*** responsibility or
- ***counterfactual ex-post*** responsibility. (3.+11.)

Principles for Measuring Responsibility (III)

Relevant information:

- Consider ***all potential consequences of all options***, not only the factual consequences or their factual causes. (9.,10.,11.)
- The factual outcome only determines the type of ex-post responsibility (factual or counterfactual), not its degree. (11.)
- Consider what agents ***can be expected to know/can rightfully claim to know*** at the time they act, not what they choose to believe. (4.)
- Timing issues that do not affect this knowledge are irrelevant. (6.)

Principles for Measuring Responsibility (IV)

Sharing/division of responsibility:

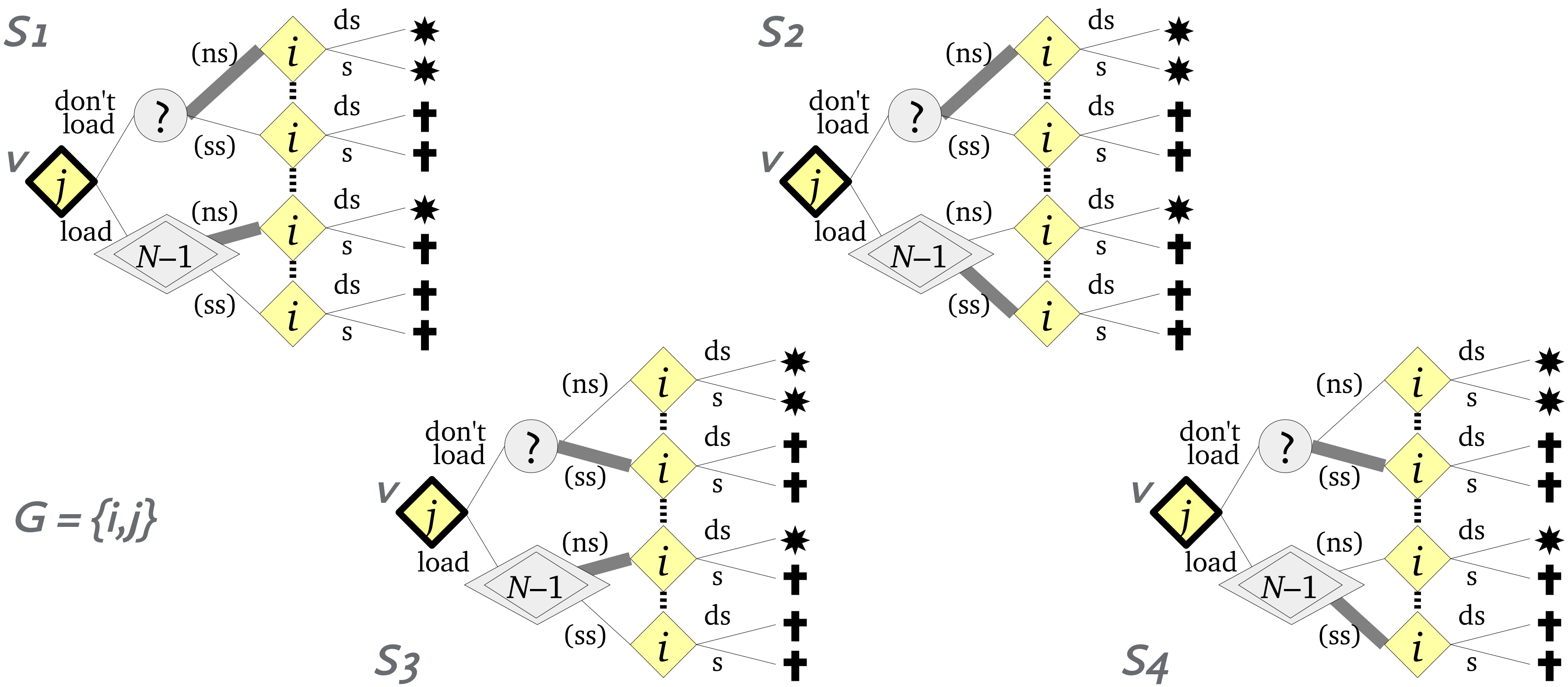
- Several agents may ***all*** individually be ***fully ex-ante*** responsible ***and fully ex-post*** responsible for the same outcome. (10.)
- A ***group's*** collective degree of responsibility (from 0 to 1) can be *smaller* than the sum of its members' individual degrees of responsibility. (10.)

  (It is unclear at this point whether it can also be *larger*)

The *Max-diff*-Formula for *Ex-Ante* Responsibility

Verbal definitions:

A ***scenario S for agent group G at node v*** is a choice of branch for each unquant. uncertainty node and all other agents' information states in the branch rooted at *v*.



The *Max-diff* -Formula for *Ex-Ante* Responsibility

Verbal definitions:

A ***scenario S for agent group G at node v*** is a choice of branch for each unquant. uncertainty node and all other agents' information states in the branch rooted at *v*.

A ***strategy s for agent group G at node v*** is a choice of action for each of *G* 's information states in the branch rooted at *v*.

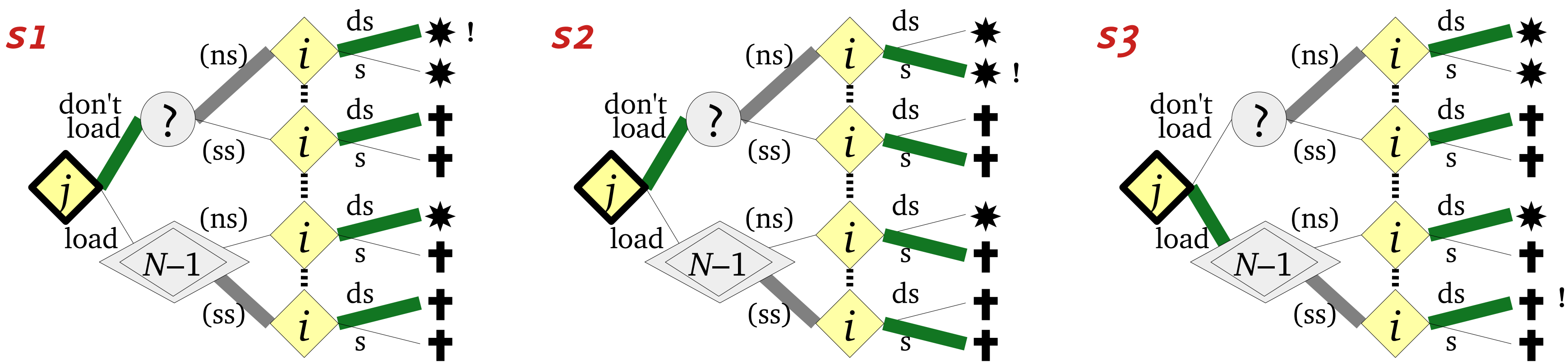
The *Max-diff* -Formula for *Ex-Ante* Responsibility

Verbal definitions:

A ***scenario S for agent group G at node v*** is a choice of branch for each unquant. uncertainty node and all other agents' information states in the branch rooted at *v*.

A ***strategy s for agent group G at node v*** is a choice of action for each of *G* 's information states in the branch rooted at *v*.

G = {i,j}, scenario S2:



s4 = (load, shoot)

The *Max-diff* -Formula for *Ex-Ante* Responsibility

Verbal definitions:

A ***scenario S for agent group G at node v*** is a choice of branch for each unquant. uncertainty node and all other agents' information states in the branch rooted at *v*.

A ***strategy s for agent group G at node v*** is a choice of action for each of *G*'s information states in the branch rooted at *v*.

The ***conditional value of G's strategy s at v given S*** is a certain strictly increasing function *f* (e.g., *f*(P) = P or *f*(P) = logit P) of the probability P, evaluated at *v*, of a good outcome, conditional on *S* and *s*.

The *Max-diff* -Formula for *Ex-Ante* Responsibility

Verbal definitions:

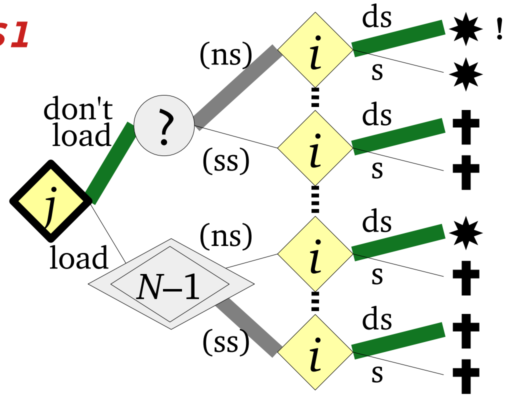
A ***scenario S for agent group G at node v*** is a choice of branch for each unquant. uncertainty node and all other agents' information states in the branch rooted at *v*.

A ***strategy s for agent group G at node v*** is a choice of action for each of *G*'s information states in the branch rooted at *v*.

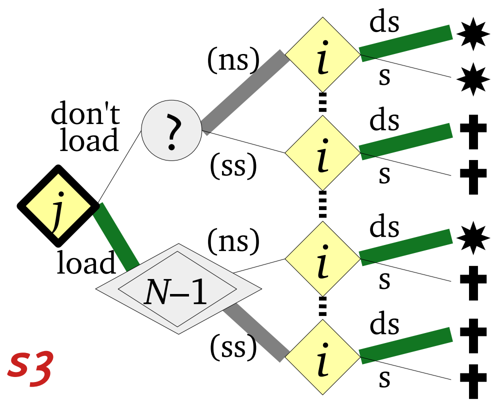
The ***conditional value of G's strategy s at v given S*** is a certain strictly increasing function *f* (e.g., *f*(P) = P or *f*(P) = logit P) of the probability P, evaluated at *v*, of a good outcome, conditional on *S* and *s*.

G = {i,j}, scenario S2, using *f*(P) = P:

s1



$f($ P(good$|v,S2,s1)$ $) = 1$



s3

$f($ P(good$|v,S2,s3)$ $) = 0$

The *Max-diff* -Formula for *Ex-Ante* Responsibility

Verbal definitions:

A ***scenario S for agent group G at node v*** is a choice of branch for each unquant. uncertainty node and all other agents' information states in the branch rooted at *v*.

A ***strategy s for agent group G at node v*** is a choice of action for each of *G* 's information states in the branch rooted at *v*.

The ***conditional value of G's strategy s at v given S*** is a certain strictly increasing function *f* (e.g., *f* (P) = P or *f* (P) = logit P) of the probability P, evaluated at *v*, of a good outcome, conditional on *S* and *s*.

G's conditional influence at v given S is the difference between the largest & smallest conditional values of all of *G* 's strategies in *S* at *v*.

$$\mathrm{ci}(G,v|S) = \max\{\, f(\,P(\mathrm{good}|v,S,s)\,) : \text{strat. } s \text{ for } G \text{ in } S \text{ at } v\,\}$$
$$-\min\{\, f(\,P(\mathrm{good}|v,S,s)\,) : \text{strat. } s \text{ for } G \text{ in } S \text{ at } v\,\}$$

The *Max-diff* -Formula for *Ex-Ante* Responsibility

Verbal definitions:

A ***scenario S for agent group G at node v*** is a choice of branch for each unquant. uncertainty node and all other agents' information states in the branch rooted at *v*.

A ***strategy s for agent group G at node v*** is a choice of action for each of *G*'s information states in the branch rooted at *v*.

The ***conditional value of G's strategy s at v given S*** is a certain strictly increasing function *f* (e.g., *f*(P) = P or *f*(P) = logit P) of the probability P, evaluated at *v*, of a good outcome, conditional on *S* and *s*.

G's conditional influence at v given S is the difference between the largest & smallest conditional values of all of *G*'s strategies in *S* at *v*.

$$\mathrm{ci}(G,v|S) = \max\{ f(\,P(\mathrm{good}|v,S,s)\,) : \text{strat. } s \text{ for } G \text{ in } S \text{ at } v \} - \min\{ f(\,P(\mathrm{good}|v,S,s)\,) : \text{strat. } s \text{ for } G \text{ in } S \text{ at } v \}$$

G = {i,j}, scenario S2, using *f*(P) = P:

s1: $f(\,P(\mathrm{good}|v,S2,s1)\,) = 1$

s3: $f(\,P(\mathrm{good}|v,S2,s3)\,) = 0$

Difference: 1

The *Max-diff* -Formula for *Ex-Ante* Responsibility

Verbal definitions:

A ***scenario S for agent group G at node v*** is a choice of branch for each unquant. uncertainty node and all other agents' information states in the branch rooted at *v*.

A ***strategy s for agent group G at node v*** is a choice of action for each of *G* 's information states in the branch rooted at *v*.

The ***conditional value of G's strategy s at v given S*** is a certain strictly increasing function *f* (e.g., *f* (P) = P or *f* (P) = logit P) of the probability P, evaluated at *v*, of a good outcome, conditional on *S* and *s*.

G's conditional influence at v given S is the difference between the largest & smallest conditional values of all of *G* 's strategies in *S* at *v*.

G's degree of ex-ante responsibility at v
is its maximum conditional influence at *v* over all possible scenarios at *v*.

The *Max-diff* -Formula for *Ex-Ante* Responsibility

Verbal definitions:

A ***scenario S for agent group G at node v*** is a choice of branch for each unquant. uncertainty node and all other agents' information states in the branch rooted at *v*.

A ***strategy s for agent group G at node v*** is a choice of action for each of *G* 's information states in the branch rooted at *v*.

The ***conditional value of G's strategy s at v given S*** is a certain strictly increasing function *f* (e.g., $f(\mathrm{P}) = \mathrm{P}$ or $f(\mathrm{P}) = \mathrm{logit}\ \mathrm{P}$) of the probability P, evaluated at *v*, of a good outcome, conditional on *S* and *s*.

G's conditional influence at v given S is the difference between the largest & smallest conditional values of all of *G* 's strategies in *S* at *v*.

G's degree of ex-ante responsibility at v is its maximum conditional influence at *v* over all possible scenarios at *v*.

Formally:

$$\mathrm{ear}(G,v) = \max \{ \max \{ f(\,\mathrm{P}(\mathrm{good}|v,S,s)\,) : \text{strat. } s \text{ for } G \text{ at } v \} - \min \{ f(\,\mathrm{P}(\mathrm{good}|v,S,s)\,) : \text{strat. } s \text{ for } G \text{ at } v \} : \text{scenario } S \text{ for } G \text{ at } v \}$$

Ex-post Responsibility Version 1: The *Sum-max-diff-max* -Formula

Verbal definitions:

G's ***reachable target in S at v***
is the maximum conditional value of all of G's strategies in *S* at *v*.

G's ***shortfall in S at decision node v due to action a***
is the difference between G's reachable targets in *S* at nodes *v* and *v:a*.

G's ***incremental ex-post-degree of responsibility due to a***
is its maximum shortfall at *node(a)* due to *a* over all possible scenarios at *node(a)*.

G's ***total ex-post-degree of responsibility (version 1)***
is the sum of its increm. ex-post-degrees over all actions actually taken by G.

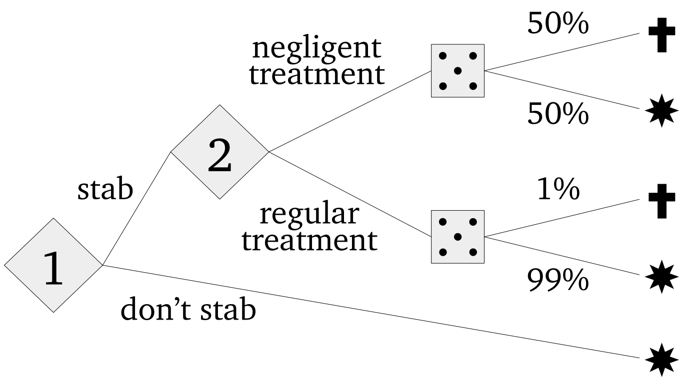
In short:

$\mathrm{rt}(G,S,v) = \max \{ f(\, \mathrm{P}(\mathrm{good}|v,S,s)\,) : \text{strategy } s \text{ for } G \text{ in } S \text{ at } v \}$

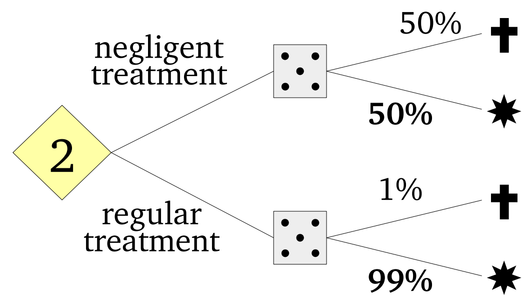
$\mathrm{epr}_1(G) = \mathrm{sum} \{ \max \{ \mathrm{rt}(G,S,v) - \mathrm{rt}(G,S,v{:}a) : S \text{ at } node(a) \} : \text{action } a \text{ taken by G} \}$

Further Examples

Example à la Canavotto & Giordani (yesterday):

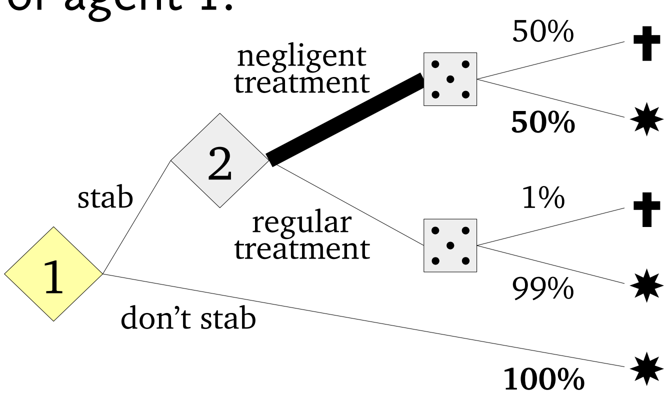


For agent 2, there is only one scenario:



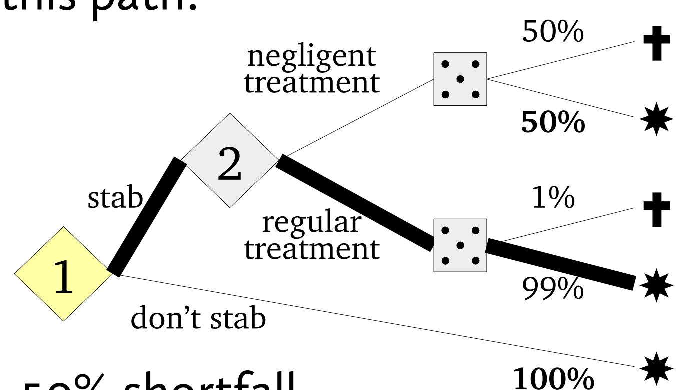
→ 49% ex-ante

Influence-maximizing scenario for agent 1:



→ 50% ex-ante

Ex-post resp. of agent 1 after this path:

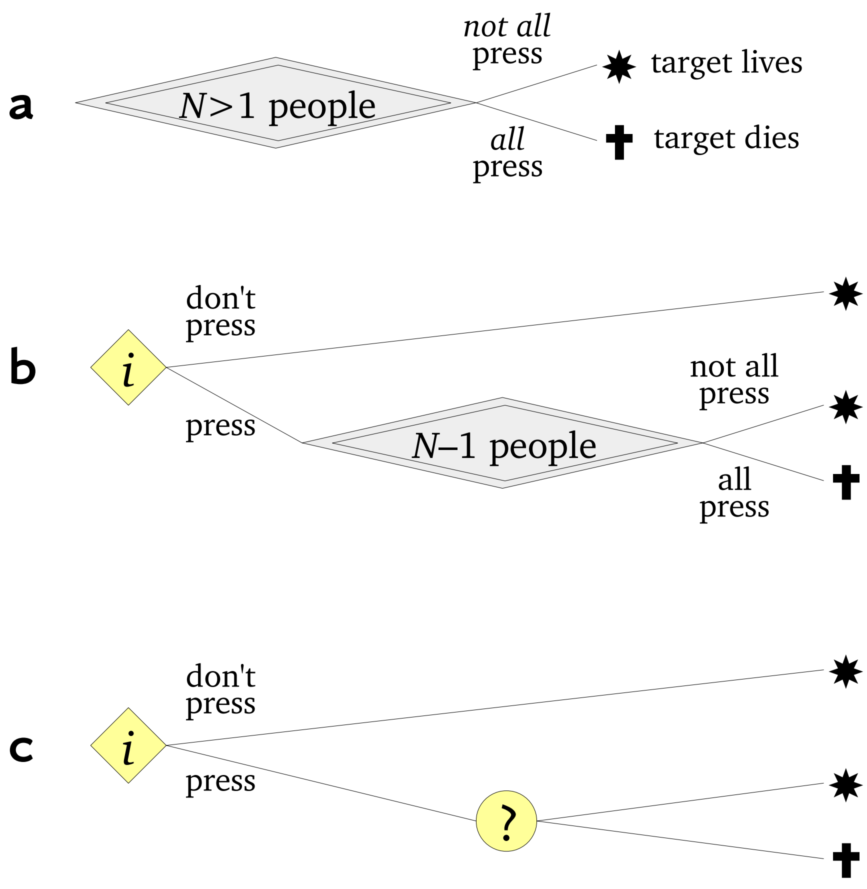


50% shortfall

→ 50% counterfactual ex-post$_1$

An Electric Chair

Target dies iff *all N* persons press a button.



Our formula implies:
a,b,c are all *equivalent*
w.r.t. the assessment
of *i*'s responsibilities;
the number *N* is irrelevant.

i has *full ex-ante responsibility*
(like any other subgroup of the *N*).

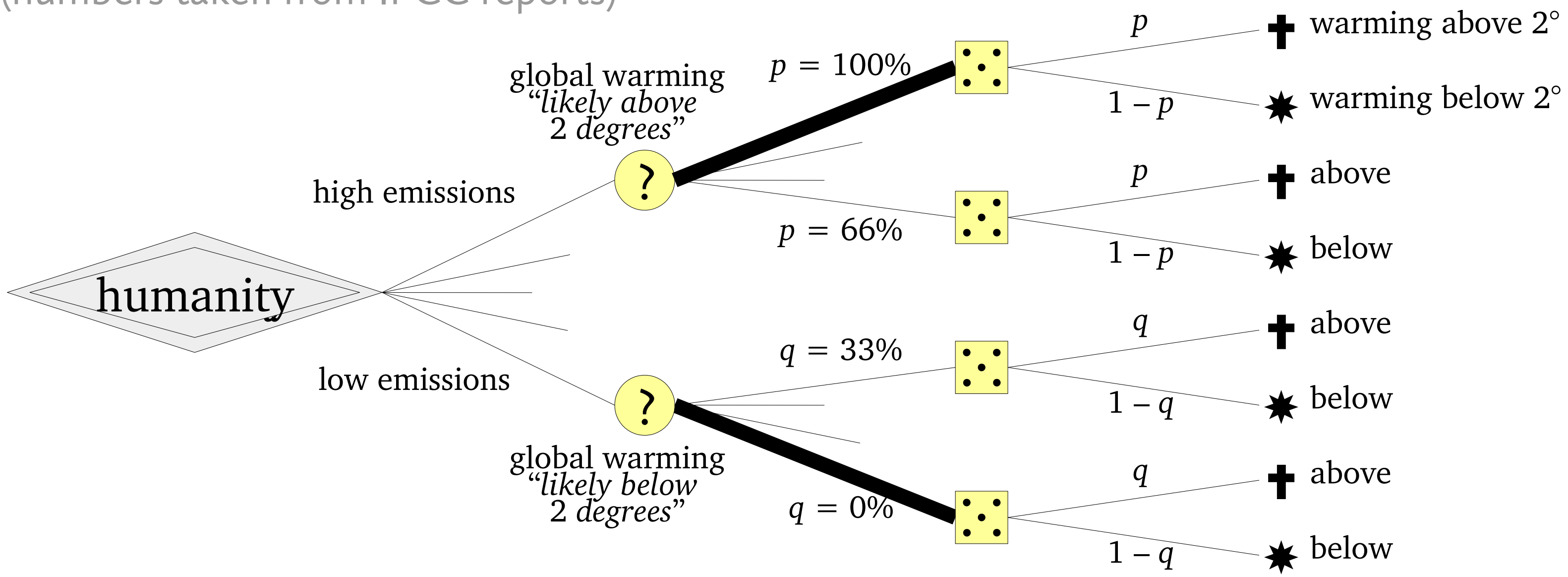
*i [*or some group *G]* has
either full or no ex-post resp.,
depending on whether
*i [*or at least one member of *G]*
has pressed her button.

These implications seem OK...

Intermediate Levels of Uncertainty

Exact probabilities unknown, but known to be within a certain range
→ Represent by combination of unquantified uncertainty and stochasticity nodes

E.g.: staying below two degrees of global warming
(numbers taken from IPCC reports)



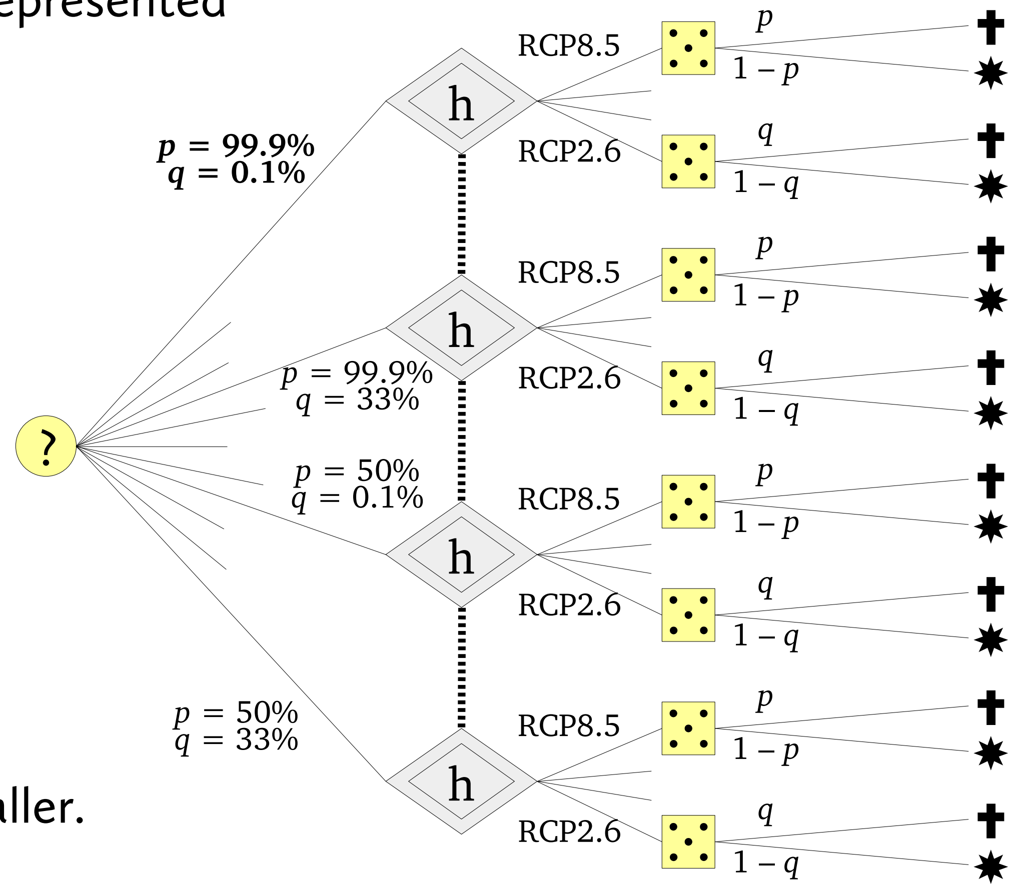
→ ex-ante resp. is 100% – 0% = 100% (and not only 66% – 33% = 33%)

Independent vs. dependent Uncertainty

The situation may be more clearly represented by this (formally equivalent) model:

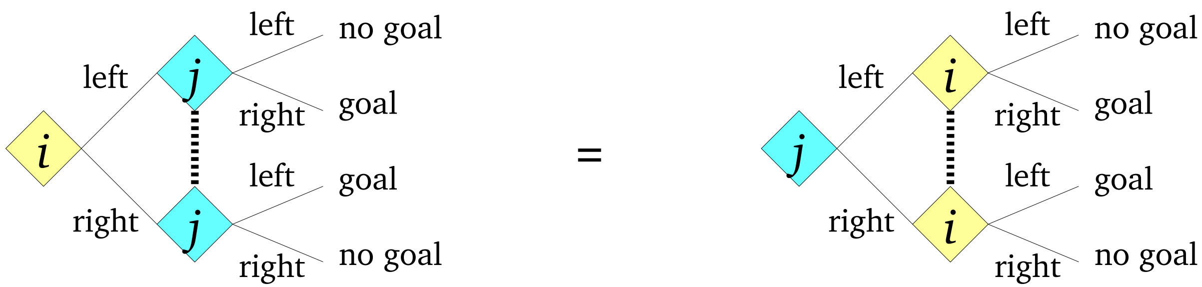
This representation shows that the model is adequate if the *uncertainty* about the effect of RCP8.5 and the uncertainty about the effect of RCP2.6 are *independent*.

If the science says they are somehow correlated, some of the many branches may not exist and hence the responsibility may be smaller.



Unavoidable Ex-Post Responsibility

Penalty kick: penalty-taker (i) kicks into one corner, goalie (j) jumps into another?



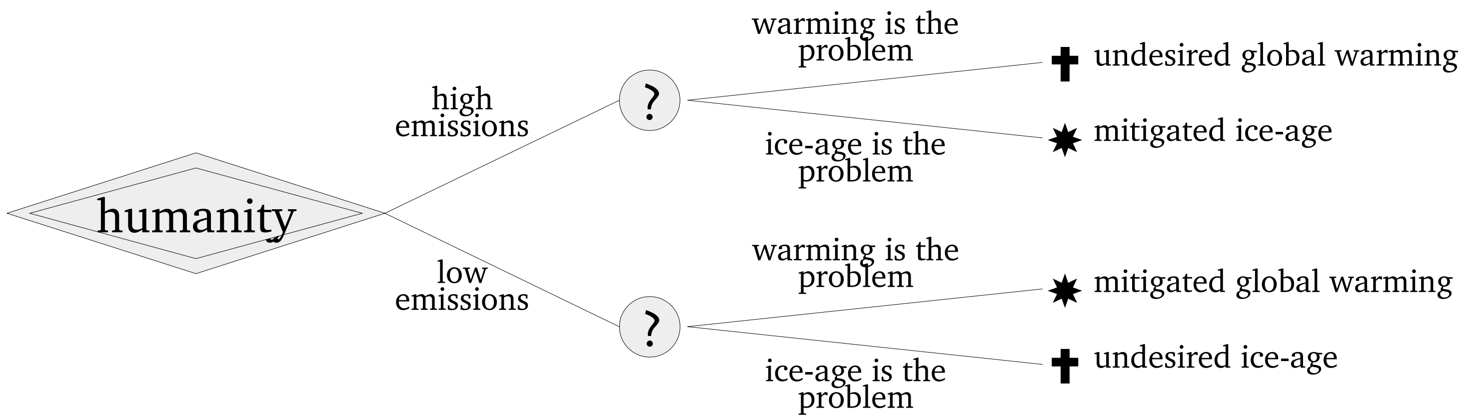
The complete symmetry of the model implies:

- both have the same degree of ex-ante responsibility
- both have the same degree of ex-post resp., no matter what they do
- the degree itself also does not depend on what they do!

In our theory:
*No matter what they do, both have **full** ex-post responsibility*
(and thus both might be held accountable for the outcome).

Sometimes There is no "Right Thing" to Do?

1980ies: Unclear whether GHG emissions support undesired global warming or help preventing an undesired imminent ice-age (private comm. with B. Hoskins)



If this model represents their knowledge at the time, our theory implies:
They had to expect to be ex-post (either factually or counterfactually) responsible anyway, no matter what they would do.

Is this reasonable?

Ex-post Responsibility Version 2: The *Sum-diff-maximin* -Formula

Verbal definitions:

G's ***guaranteed value of strategy s at node v***
is the minimum conditional value of *s* over all scenarios *S* of *G* at *v*.

G's ***precautionary target at v***
is the maximum guaranteed value of all of *G*'s strategies at *v*.

G's ***shortfall at v due to action a***
is the difference between *G*'s precautionary targets at nodes *v* and *v:a*.

G's ***total ex-post-degree of responsibility (version 2)***
is the sum of its shortfalls over all node-action pairs actually traversed by *G*.

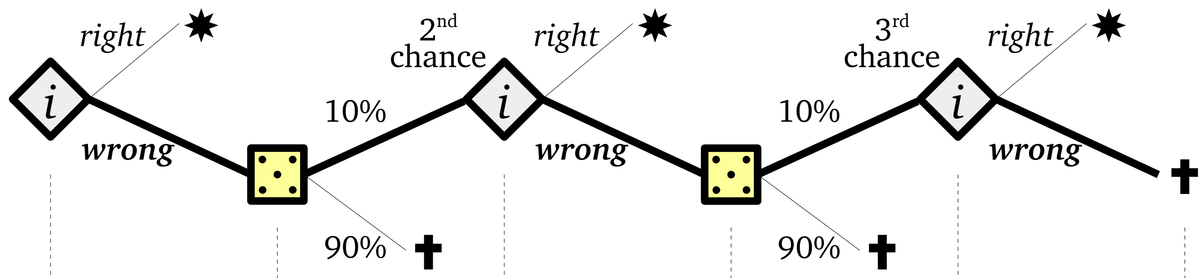
In short:

$\mathrm{pt}(G,v) = \max \{ \min \{ f(\,\mathrm{P}(\mathrm{good}|v,S,s)\,) : \text{scen. } S \text{ of } G \text{ at } v \} : \text{strat. } s \text{ for } G \text{ in } S \text{ at } v \}$

$\mathrm{epr}_2(G) = \mathrm{sum} \{ \mathrm{pt}(G,v) - \mathrm{pt}(G,v{:}a) : \text{node-action pair } (v,a) \text{ traversed by } G \}$

Repeated Failure can Increase Ex-Post Responsibility Beyond 100%!

Unexpected 2^{nd} and 3^{rd} chances, all failed:



Quantities:

smallest prob. of ✷	0	0	0	0	0	0
reachable target = largest prob. of ✷	1	0.1	1	0.1	1	0
influence = diff.	1	0.1	1	0.1	1	0
ex-ante resp.	100%		100%		100%	
shortfall = diff.		0.9		0.9		1
ex-post resp.		90%		180%		280%

Some Further Properties

- responsibility **increases** with
  - more options or repeated chances to act
  - higher uncertainty, earlier chances to act
- responsibility **does not decrease** with
  - more other players (no division or “diffusion of resp.”)
  - luck after a wrong decision

Some Next Steps

- Read all the stuff you guys have already done on this (Sorry for not citing anything – we were ignorant of much of the existing formal literature)
- Figure out which formula for ex-post makes more sense
- Relate to logics (use game trees in semantics, include "degree-of-responsibility" quantifiers, etc.?)
- Analyse natural-language use of "responsibility"
- Apply to somewhat realistic model of climate change
- *… Suggestions?*

Thank you for your attention!

heitzig@pik-potsdam.de