UNIVERSITY OF GOTHENBURG

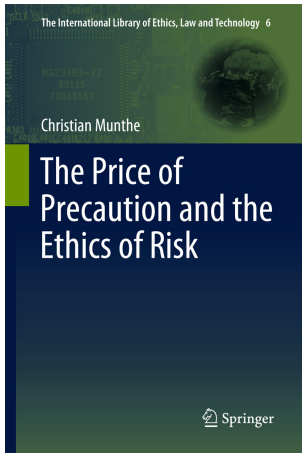
THE PRICE OF PRECAUTION

EVALUATING ACTIONS ACTUALISED BY EXTREME AND EXTREMELY UNCLEAR RISKS

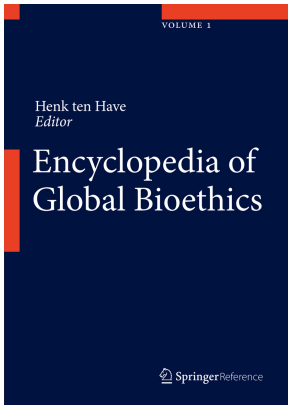
CHRISTIAN MUNTHE, PROFESSOR OF PRACTICAL PHILOSOPHY

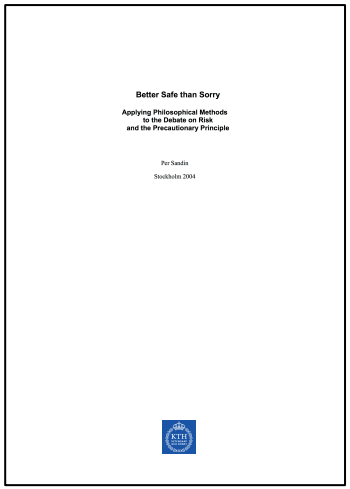
UNIVERSITY OF GOTHENBURG

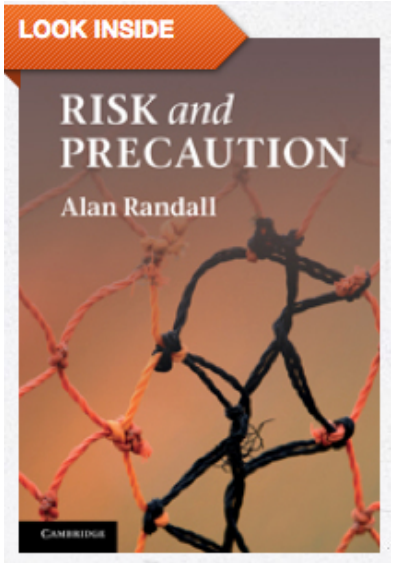
DEPARTMENT OF PHILOSOPHY, LINGUISTICS AND THEORY OF SCIENCE)

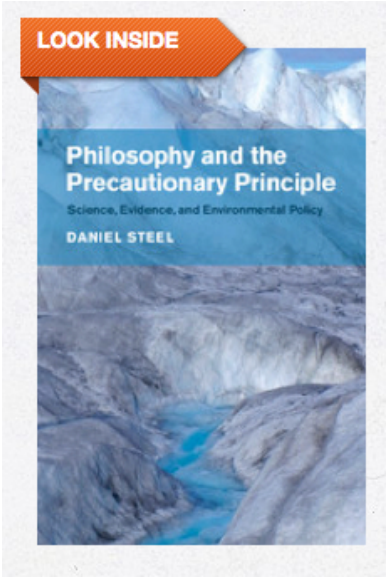


Extended article accepted, in press, 2015









Basic Challenge of Technology Development: Uncertainty, Ignorance and ”Black Swans”

- New technologies (hopefully) bring benefits (if not, we have no reason to develop or use them)
- We know that they *may* also effect substantial downsides, but not whether or not they *will*, or exactly *which*
- Some of these are (claimed to be) *extreme*
- Widespread harm / vast eradication of value / eradication of humanity

There is good reason to *care* about these things – the problem is *how much* and *what we should do as a consequence!*

- Calculated risk-taking: any small/unclear probability may combine with massive negative outcome values into huge risks
- Ethical theories support the general idea that we have more reason to avoid larger losses of value / harms
- Less clear, however, what this implies in practical terms: traditional ethical theory typically void of guidance
- **The idea of precaution**: we have reason to halt development, clarify dangers, apply prevention, possibly abstain altogether

The reason for precaution cuts both ways: opportunity costs and additional costs and risks

- Many technologies may be necessary to mitigate or prevent various extreme risks
- Geoengineering, AI-tech, med-tech, synbio, space tech, etc.
- Thus: halting their development, crippling their potential with preventive inhibitions, etc. may itself impose extreme risk
- In addition 1: the resources could have been spent on mitigating or preventing more clarified and more easily affected, albeit less extreme risks
- In addition 2: spending resources to clarify unclear risks, apply preventions, etc. will add costs and risks of their own

The price of precaution

- Not to care / do anything about potential dangers has an obvious price (in ethical terms): recklessness, negligence, irresponsibility
- But caring and acting on this reason *also* has a price: what is lost in terms of possible benefits, lost opportunities and added costs/risks
- This price must not be implausibly high – but what determines plausibility here? – as mentioned, ethical theory provides little clue
- The problem is not new: well-known issue, e.g., in medicine and hinted at since ancient times, e.g., in virtue ethical ideals
- Extreme risks scenarios would seem to pose a particular challenge: in view of the enormity of what's at stake, *whenever should we relax our precautions?*

UNIVERSITY OF GOTHENBURG

DEPARTMENT OF PHILOSOPHY,
LINGUISTICS AND THEORY OF SCIENCE)

What determines the price of precaution?

The requirement of precaution

"in the face of an activity that may produce great harm, we (or society) have reason to ensure that the activity is not undertaken, unless it has been shown not to impose too serious risks" (Munthe 2015)

- Eligibility criteria: "May" and "great harm"?
- Evaluation criteria: "too serious"?
- Proof standards: "show"

My basic thesis:

- Decision-making must not be systemtically paralysed ➔ No ban on the risking of certain outcomes
- We have a (non-absolute) moral reason to improve the evidence basis on which we decide
- Ingredients must jointly express a *normatively plausible* price of precaution ➔ the ethics of risk

Three paradigm examples: LHC, A.I. & Space tech

- LHC
  - possible advance of fundamental understanding of the workings of the natural world, the universe, etc. – who knows what that may lead to?
  - possible massive disasters of unfathomable magnitudes (or?)

- A.I.
  - Possible major advances of collective decision-making, wealth for all, etc.
  - Possible undermining of civilizations, machine takeover, etc.

- Space tech
  - The ultimate rescue option: humans likely to wreck the planet, and soner or later a major meteorite may come our way
  - An economic black hole, potentially able to claim all resources available, possibly to no good at all

Can *de minimis* risk resolve the issue?

- Some risks are not eligible, for some reason (?)
- Some ideas emphasise the probability side (counting against minding about extreme risks) or the outcome value side (counting for) of a risk. Most seem more or less arbitrary.ww
- Basic idea of decision-costs ➔ the issue of the proper price of precaution: what costs *are* too high and *why*?
- Whether or not a refinement of the evidence base would change the decision recommendation ➔ the issue of the proper price of precaution: assumes a criterion of good decisions and this applies also to the issue of whther or not to seek more evidence
- Extreme risk scenarios thus not *easily* dismissed from a precautionary agenda, and we have some reason to improve our evidence on the matter, but unclear how much

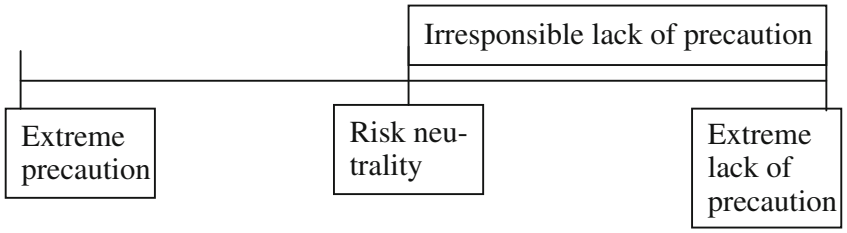
Two basic models: 1. risk neutrality

- Risks and chances of equal magnitude balance each other out in responsibility terms when options are compared
- Irresponsible lack of precaution is to decide something where the risks are not thus balanced by chances, more precaution than this is to pay an implausibly high price of precaution
- Fits well with standard approaches to risk analysis within the maximizing expected utility paradigm
- Leads over to the evaluation of the respective outcome scenarios, as precaution cuts both ways and applies also to the issue of possibly refining the evidence basis

Two basic models: 2. increased weight of evil

- We have reason to pay a higher price of precaution than what risk neutrality requires in order to certify avoidance of (certain) risks due to their outcome aspect.

- Many possible models that reflect different ethics/value stances, but basically



- My 'indirectly sufficientarianist' idea: If an option secures an "acceptable risk-chance mix" relative to what affected parties stand to loose or gain, risks of other options become more difficult to justify; it is worth paying a higher price of precuation in terms of lost benefits to avoid them.

In both cases:

- Since precaution cuts both ways, our reasons to halt or cripple technological advance may not be that demanding beyond our reason to avoid unnecessary hazards – extreme risks are all around …

- To discriminate further, some justifiable eligibility criterion giving us reason to ignore some of them needs to be presented

- But the presence of less extreme, though more clarified and easy to meet challenges that require resources may be a reason to abstain further technological advance until these have been met

- The increased weight of evil approach may provide more reason of that kind

Extreme risk *outcomes*: four ways in which the eradication of humanity need not be (such) a bad thing

- Humanity is in fact not that valuable, this idea is mainly a product of either unjustifiable dogma or bilogically programmed wishful thinking (David Benatar)

- Humanity (naturally and/or aided by technology) is transformed into another type of biological being that still possess as valuable qualities (or even more)

- Humanity is (peacefully and painlessly) replaced by (originally human-made) ”superintelligent” machines, which possess as valuable qualities (or even more)

- On the whole and in the long run: the downside of a non-violent disappearance of humanity is well balanced by benefits to other types of creatures