

Filosofia Unisinos  
Unisinos Journal of Philosophy  
23(3): 1-16, 2022 | e23305

Unisinos – doi: 10.4013/fsu.2022.233.05

Article

## Not Too Risky. How to Take a Reasonable Stance on Human Enhancement

Nem tão arriscado. Como assumir uma posição razoável sobre o melhoramento humano

**Murilo Mariano Vilaça**

<https://orcid.org/0000-0001-9720-5552>

Fundação Oswaldo Cruz (ENSP/DIHS) / Programa de Pós-graduação em Bioética, Ética Aplicada e Saúde Coletiva (PPGBIOS) / Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro - FAPERJ (JCNE), Rio de Janeiro/RJ, Brazil. Email: [contatoacademico@hotmail.com](mailto:contatoacademico@hotmail.com).

**Andrea Lavazza**

<https://orcid.org/0000-0003-2608-2609>

Centro Universitario Internazionale, Arezzo; University of Pavia, Italy. Email: [lavazza67@gmail.com](mailto:lavazza67@gmail.com)

### ABSTRACT

Following a trend in bioethical/applied ethics approaches, one of the frustrating features of studies on technological human enhancement is their dichotomous tendency. Often, benefits and risks of technological human enhancement are stated in theoretically and empirically vague, polarized, unweighted ways. This has blocked the debate in the problematic 'pros vs. cons' stage, leading to the adoption of extremist positions. In this paper, we will address one side of the problem: the focus on risks and the imprecise approach to them. What motivates our approach is the identification of the weaknesses of the anti-enhancement criticism, which stem from its use of the concept of risk, as well as the heuristic of fear and the precautionary principle. Thus, 'taking a step back' to move forward in the debate, our purpose is to establish some theoretical foundations concerning the concept of risk, recognizing, at the same time, its complexity and importance for the debate. Besides the concept of risk, we emphasize the concept of existential risk, and we make some considerations about epistemic challenges. Finally, we highlight central features of more promising approaches to move the debate forward.

**Keywords:** Human enhancement technologies, risk, uncertainty, conceptual problem, epistemic challenges.



Artigo está licenciado sob forma de uma licença  
Creative Commons Atribuição 4.0 Internacional.

## RESUMO

Seguindo uma tendência das abordagens bioéticas/em ética aplicada, uma das características frustrantes dos estudos sobre o melhoramento humano tecnológico é a sua tendência dicotômica. Muitas vezes, os benefícios e riscos do melhoramento humano tecnológico são afirmados de modo teórico e empiricamente vago, polarizado e não ponderado. Isto 'emperra' o debate no problemático estágio 'prós versus contras', ensejando a adoção de posições extremistas. Nesse artigo, abordaremos um lado do problema: o foco sobre os riscos e a abordagem imprecisa destes. O que motiva nossa abordagem é a identificação das fragilidades decorrentes da utilização do conceito de risco na retórica anti-melhoramento, bem como a heurística do medo e o princípio de precaução em que esta corrente se baseia. Assim, 'dando um passo atrás' para avançar no debate, nosso objetivo é estabelecer alguns fundamentos teóricos relativos ao conceito de risco, reconhecendo, ao mesmo tempo, a sua complexidade e importância para o debate. Além do conceito de risco, enfatizamos o conceito de risco existencial, e fazemos algumas considerações sobre os desafios epistêmicos. Por fim, destacamos as características centrais de abordagens mais promissoras para fazer o debate avançar.

**Palavras-chave:** Tecnologias de melhoramento humano, risco, incerteza, problema conceitual, desafios epistêmicos.

## 1. Technological Human Enhancement and its Critics

### 1.1. What is Human Enhancement?

Human beings have always tried to improve their performance with all the means they have been able to devise. Exercise and constant practice have been used in this sense since ancient times. But natural stimulating substances have been in use for millennia as well. Today, we have at our disposal a number of new technologies that promise a quantum leap in human enhancement (HE) possibilities. Among these, off-label drugs used to increase athletic performance, attention, and concentration; memory specific molecules; targeted psychotropic drugs; brain stimulation to increase cognition and creativity; interfaces (direct or indirect) between brain and machine; bodily prostheses and brain implants; genetic engineering through Crispr-Cas9.

In order to have a conceptual clarity useful for the discussion about the criticism of human enhancement, it is helpful to introduce some definitions and distinctions. *Enhancement* means the improvement of one's natural resources (actual or yet to be expressed) by heterogeneous means, to have more skills and, therefore, more opportunities. However, *enhancement* also means *change*, and this may raise unprecedented problems, different forms of competition or new social inequalities, or even changes in the "human" condition. The enhancement of human performance is the increase in skills (abilities, attributes, or competences) by means of medical or technological methods, to improve a person's overall performance (see better, run faster, remember more accurately). Examples are intraocular lenses; doping in sport; but also a special training programme that resorts to technological aids (Lavazza and Colzato, 2018).

*Enhancement* as *empowerment* indicates a process of individual growth, based on increasing self-esteem, self-efficacy and self-determination to bring out latent resources and consciously appropriate the potential with which one is naturally endowed. *Empowerment* is perhaps a *soft version* of enhancement. Or, rather, it is something different, which recalls the idea of "guided exercise". It is a growth that can be sustained, but not directly caused, by devices.

Improvement can be understood as a qualitative concept (individual and intersubjective well-being, fulfilment of one's existence), as excellence and progress in absolute terms, for us and for others,

and not in comparison with others or against others. Enhancement can also be understood as a quantitative concept, as extension of scope and duration (physical strength, intellectual capacity), but in a competitive and positional sense.

Concerning cognition and human enhancement, which seems to be the area where objections may be greatest, a distinction can be made between cognitive enhancement and neurocognitive enhancement.

Cognitive enhancement is an amplification or extension of the central capacities of the mind through the (qualitative) improvement or (quantitative) increase of the systems - internal or external - of information processing. Education is an ancient and traditional method of cognitive enhancement. Several other strategies are well known and commonly used. There are new techniques available today, not all of which are well established, and which serve different purposes. Cognitive enhancement today refers to techniques - or (also electronic) tools used as training aids - that increase 'normal' abilities, increasing or improving the various components of cognition (memory, attention, executive functions, etc.). The distinctive feature of cognitive enhancement is the conscious participation of the subject in the process.

Neurocognitive enhancement refers to technologies and drugs that increase 'normal' abilities in the brain, increasing or improving one's mood as well as the components of cognition. The distinctive feature of neurocognitive enhancement is its (at least partial) automaticity and the subject's unawareness of the process (passivity and dependence).

## 1.2. The Criticism to Human Enhancement

Human enhancement can be contested because:

- (a) It is unnatural, since it affects an endowment which has been given to us and which in any case is the result of a crystallised balance.
- (b) Diminishes willpower and the ability to withstand difficulties or unforeseen circumstances.
- (c) Represents an attitude of hubris, of Promethean forcing of our humanity.
- (d) It may increase inequalities or create new inequalities, with a class of 'superhumans' able to access the most expensive enhancers.
- (e) It may create new forms of coercion, since even those who do not want to resort to enhancement will have to do so, so as not to be overtaken by those who use it.
- (f) It constitutes a form of doping that distorts competition in selective contexts, as is the case in sport, even when enhancement is not explicitly prohibited.

No participant in the EH debate who pretends to be reasonable can deny that there are risks involved in it. Thus, it is not around this question - whether or not there would be risks - that there seem to be relevant controversies, which deserve to be taken seriously. Regarding the conceptual and epistemic aspects, our attention should be on how to define, identify, differentiate, prioritize, estimate and, if possible/feasible, prevent/mitigate risks.

But, for some authors, these tasks seem to be irrelevant, since the human enhancement would itself be a threat that would impose only catastrophic risks on humanity. Such authors are often defined as bioconservatives or, in Buchanan's terms (2011), anti-enhancement. Aware that there are some nuances that differentiate them<sup>1</sup>, it can be said that the authors of the bioconservative front agree on some fundamental points: there is a human nature; some of the most relevant human values<sup>2</sup> are based on

<sup>1</sup> On the characteristics of bioconservative thinking, as well as on similarities and differences between some of the main bioconservatives - Leon Kass, Francis Fukuyama, Michael Sandel and Jürgen Habermas, see Roache and Clarke (2009) and Vilaça (2021).

<sup>2</sup> Dignity, autonomy, authenticity, equality, human rights, solidarity, gratitude, etc.

it; technological HE seriously threaten<sup>3</sup> human nature and therefore human values; thus, there are radical and unavoidable normative problems in human enhancement, what makes it *prima facie* immoral; advance towards the human enhancement will necessarily cause catastrophic and irreversible effects.

With a difference of emphasis, explicitly or implicitly, they share the bad prognosis, which is what characterizes the so-called Heuristics of Fear. Such heuristic is based on a simple aphorism: “[...] *in dubio pro malo* – when in doubt, listen to the worst prognosis rather than the best” (Jonas, 1997, p. 49). This is, for Jonas, the answer to the question of how to put precaution in the face of risks in technological civilization.

The vagueness that characterizes the uses of the concept of risk is a key point of the generalized/hyperbolic fear present in the anti-enhancement discursive line. Vagueness is directly associated with another feature of ‘anti-enhancement rhetoric’, namely, catastrophism. Authors of the anti-enhancement current claim that there are risks and that, in addition, they would be catastrophic and irreversible, but they often do not provide explicit arguments or data. Despite this, they can exert great influence on part of those interested in the human enhancement subject.

Faced with the announcement of the ‘end of humanity’, who, in their right mind, should not panic, even more so if the ‘news’ of such an imminent risk of extinction comes from internationally famous thinkers? Bioconservatives proclaim a kind of ‘end of the world’, their claims being based not on elements of empirical nature but more often on a principled view against the very idea of enhancement. They are following in the footsteps of Jonas, for which

[...] an imaginative “heuristics of fear”, replacing the former projections of hope, must tell us what is possibly at stake and what we must beware of. the magnitude of those stakes, taken together with the insufficiency of our predictive knowledge, leads to the pragmatic rule to give the prophecy of doom priority over the prophecy of bliss (Jonas, 1984, p. x).

Interestingly, *prophets of doom* are committed to a kind of ‘boomerang argument’. In other words, if we cannot trust our rational ability to predict, what would be the epistemic basis of a bad prognosis, if not a *pessimism linked to conservative traditionalism*?

As stated by John Kekes, a renowned conservative scholar, pessimism and traditionalism characterize conservative thinking (Kekes, 1998). According to Gordon, Burckhart and Segler (2014), the heuristic of fear is typically a *conservative precautionary principle*, according to which, with regards to the unpredictable effects of technological development, poor prognosis should be prioritized over good prognosis. According to Harris (2009), one of the most frequent objections to interventions on ‘human nature’, that is, the named claim of Playing God, is heavily tied to The Precautionary Principle (Harris, 2009).

In defending conservative pessimism, Jonas (1984, p. 81, p 204) extrapolates, from an argumentative point of view, some limits that should be strictly respected. We will highlight only two emblematic passages, in which the author tries to defend his pessimistic stance, but ends up exposing his lack of commitment to the epistemic content of catastrophic predictions.

He states that “the reproach of ‘pessimism’ leveled at such partisanship for the ‘prophecy of doom’ can be countered with the remark that the greater pessimism is on the side of those who consider the given to be so bad or worthless that every gamble for its possible improvement is defensible” (Jonas, 1984, p. 81). In addition, he claims that “the prophecy of doom is made to avert its coming, and it would be the height of injustice later to deride the ‘alarmists’ because ‘it did not turn out so bad after all.’ To have been wrong may be their merit” (Jonas, 1984, p. 204).

In other words, the bioconservative would have a kind of safe conduct: he can make catastrophic predictions and draw conclusions about the consequences of technoscientific interventions. But if none

<sup>3</sup> To instrumentalize human beings (Habermas), insulting the gifts received by God, nature or chance (Sandel), committing acts that cause disgust (Kass) and/or endorsing the most dangerous idea in the world (Fukuyama).

of what he predicted came to pass, he could not be faulted. Your mistake would be a merit, says the 'father' of the heuristic of fear. It seems like this understanding goes beyond the most basic rules of rational debate. As a result, one can point out that a considerable part of Jonas' assumptions fit into at least three of the five "frustrating" aspects of the biomedical enhancement debate indicated by Buchanan (2011), namely, murky rhetoric masquerading as argument, sweeping empirical claims, without evidence, and fundamental non-clarity.

## 2. What is at Stake with Human Enhancement

The debate around the technological human enhancement is complex and involves multiple key concepts. Two of them, as said, form a kind of inseparable binomial, namely, risk and benefit. Such concepts have been present in debates on human interventions via science and technology for a long time. It can even be said that a significant part of these debates has always been guided by them.

The current debate on technological HE is no different. Speculations about the possible benefits and risks of improving interventions sometimes compete, creating a radical polarization of the debate (pros vs. cons). Although it does not represent the totality of the approaches present in the debate (specifically, it is a feature of consequentialist approaches), this risk-benefit opposition is not rare, while some seem to focus on the possible benefits, others seem to invest all their attention and argumentation on the risks.

Although they are well known, the concepts of risk and benefit are still used in very vague ways. Benefits and risks are stated, without their fundamental aspects, such as the conceptual and epistemic aspects, as well as their normative implications, being seriously addressed. In this vein, it is helpful to address the generic use of the concept of risk in the debate on human enhancement. Our objectives can be considered modest, but we would like to define them as basic, in the sense of fundamental and necessary: (1) emphasize the necessary clarification of the conceptual problem and highlight some of the main epistemic challenges related to the risk associated with technoscience; and (2) contribute to the debate around the best critical approach to HE, in order to shed light on another important issue, that is, the normative one, since the concept of risk underlies claims for banning HE.

Risk is related to a relevant characteristic of human beings, namely, valuing, caring, worrying and, therefore, investing in things that they consider significant in themselves or for them. Although there is a wide variety of things that are valued/valuable, and that some things are valuable in themselves (have intrinsic value), everything that is valued is something significant and important, something that deserves our attention and concern, which does not mean that only what we care about has value, for a very simple reason: it would be impossible to care about everything that has value (Kahane, 2014).

According to Kahane (2014, p. 750), "what attention and concern is merited by something is a function not only of its own value, but also of what else of value is in view". Thus, if something is valuable, intrinsically or for us, it means that it is the object of our attention and concern, which is expressed by initiatives of preservation, promotion or, at the limit, improvement. In other words, valuing implies making decisions and acting in the world, sometimes making interventions, which contribute to maintaining, promoting or improving what is important for us or in itself.

Although meaning should not be confused with value, "for something to be significant, it needs to possess (or at least bring about) some value" (Kahane, 2014, p. 749), understanding *significant* as *important*, as that which *makes a real difference*, being, therefore, worthy of our attention and concern. Conversely, when something is regarded as insignificant, it does not *necessarily and totally lack* some value, but rather would not be *important enough* to someone (who regards it as such) that he would not care about it – at least not in the same way as you care about what matters most to you.

Much of what is important/significant/valued by humans is vulnerable to threats, which generate risks. In a way, then, the equation is formed by what *matters to us* (people or material or immaterial goods, tangible or intangible, what some call *Asset*) + *Threat* + *Vulnerability* = *Risk* or  $A + T + V = R$  (Vellani, 2022).



According to this understanding, if there is something/someone that matters to us, threats, but there are no vulnerabilities involved, there is no risk. If there are threats, vulnerabilities, but there is not something/someone important (valued) involved, then neither. In short, there is only risk if there is a threat to something/someone that we value and that is vulnerable.

Such conditions are present in the HE debate: humans matter, they are vulnerable, they can be threatened by certain interventions, which can expose them to risks. Like any intervention on human life, HE can, at the same time, express attention/concern for humans or represent threats (of different types or at different levels) to them. In this sense, it becomes understandable that we overestimate the risk of jeopardizing the value we hold so dear. This is why it is important to clarify conceptually what risk means in this context and to circumscribe the tendency to overestimate risk in relation to human enhancement via technological means.

### 3. The Conceptual Problem of Risk and Some Epistemic Challenges

In a democracy, anyone can freely express their opinion about anything, within certain constitutional limits. But the exercise of freedom of speech does not imply the epistemic condition of knowing. It's one thing to have an opinion, another to know. The same goes for the discourse about risk. People tend to express their opinions about risk without epistemic concerns, based simply on their perceptions or personal convictions. In the scientific discourse, however, generic and imprecise claims about risk can be extremely problematic, as such opinions encumber the debate with something indeterminate, but which, despite its epistemic fragility, have the potential to mobilize, influence or bias the decision-making process.

Generic assumptions without theoretical-empirical support can give rise to hyperbolic and incorrect perceptions about the risks involved in a specific activity. This can have undue restrictive effects. For example, scientific research and new technological applications can be hindered or even prohibited, due to supposed risks considered inevitable, unavoidable, and catastrophic. Accordingly, it is useful to make a distinction between what can be verified/falsified/criticized, because it claims objectivity/truth, and what is immune to this process of scrutiny, since it is a mere expression of a subjectivity. Thus, wrong decisions can be made, given what Andersson et al. (2020) call "noisy behavior or decision noise", which lead to bias in the show elicitation tasks.

Obviously, a distinction should be made between the criticism of human enhancement that is brought by the experts and the irrational fears of the public, who do not have an in-depth knowledge of the topic. However, what can be noted is that even in the scientific literature concerning human enhancement, there is almost always a lack of detailed analysis of the concept of risk and its epistemic aspects. For this reason, a pedagogy of risk, which highlights the basic aspects and contemporary elaboration on the subject, is particularly useful both for the specialist debate and for all those who tend to be misled by a naive or emotional consideration of the risk involved in human enhancement.

For Vaz (2004), the enormous relevance that the concept of risk has acquired requires, from the outset, a conceptual clarification. In this sense, Vaz highlights what constitutes the core of the idea of risk: the "[...] attempt to bring an undesirable future event to the present, calculate it and define the ways to face it" (Vaz, 2004, pp. 112-113). Accordingly, "the concept of risk is in opposition to the philosophical concept of *necessity* in its epistemological and existential dimensions. It applies in situations where the future is neither necessary, absolutely foreseen, nor totally unknown" (Vaz, 2004, p. 113).

In Souza's terms,

*The simplest understanding of the concept of risk is as "the probability of a dangerous event ( $p(E)$ ) multiplied by the amount of the expected damage ( $D$ ) connected to this event:  $R(E) = p(E) \times D$ ".*

*In common speech and practice, however, that clear concept quickly becomes murky as talk of risk appears to reflect a confusing multiplicity of meanings (Souza, 2010, p. 17).*

According to the glossary of *The Society for Risk Analysis* – SRA,

*We consider a future activity [interpreted in a wide sense to also cover, for example, natural phenomena], for example the operation of a system, and define risk in relation to the consequences (effects, implications) of this activity with respect to something that humans value. The consequences are often seen in relation to some reference values (planned values, objectives, etc.), and the focus is often on negative, undesirable consequences. There is always at least one outcome that is considered as negative or undesirable (Aven et al., 2018, p. 4).*

A fundamental distinction was already foreseen in one of the classics on the concept of risk. In *Risk, Uncertainty and Profit* (1921), Knight states that risk is a measurable form of uncertainty, but ‘true’ uncertainty is that which cannot be measurable or quantifiable. According to Vaz,

*In a simple definition of the difference between risk and uncertainty, if we don’t know for sure what will happen, but we know the chances of events, we have risk; if we do not know its probability, if no estimate is possible, we are faced with pure and simple uncertainty (Vaz, 2004, p. 213).*

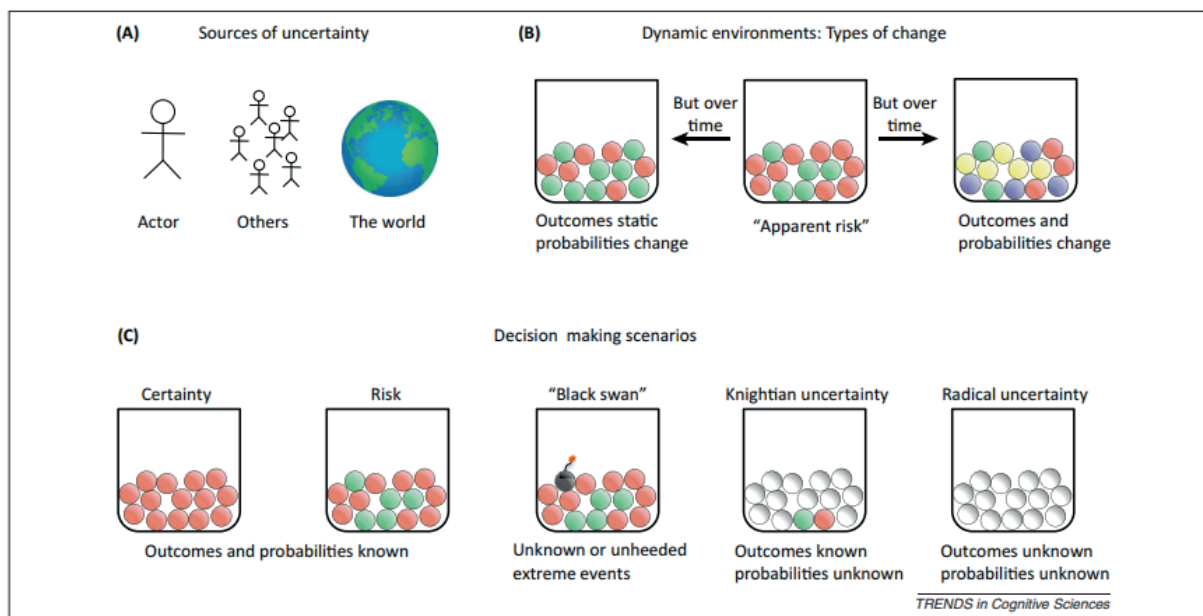
To summarize schematically, we can resort to the table devised by Campbell and Clarke (2018), in which they differentiate between certainty, risk and uncertainty both in descriptive terms and in terms of the appropriate approach (we made a small stylistic change and included a category, namely, measures of control):

**Table 1:** Certainty vs risk vs uncertainty

Type of risk	Certainty (no risk)	Risk (known risk)	Uncertainty (unknown risk)
Description	All possible information is known; outcomes are certain	Potential outcomes and probability are quantifiable/measurable using statistics and modelling (e.g. the lottery, games of chance)	Either the outcome is known but the probability is not known (‘Knightian uncertainty’ [...]), or both the possible outcomes and their probability are not known (‘radical uncertainty’ – ibid.) (e.g. earthquakes, shocks, health, romance)
Approach required		Ability to understand what predictions mean	Ability to recognize when something cannot be known
		Logic	Intuition
		Statistical analysis	Heuristics
Control measures		There is the possibility of developing control measures, which, if applied in a timely manner, can prevent or mitigate the anticipated damage, without eliminating it	Beyond control

Source: Campbell and Clarke (2018, p. 3).

Also based on Knightian concept of uncertainty, Meder, Lec and Osman (2016) propose a very enlightening scheme<sup>4</sup>, in a text in which they rethink the question of how to make decisions under uncertainty:



Source: Meder, Lec and Osman (2016, p. 259).

The above approach indicates some important conceptual and epistemic differences between a risk and an uncertainty.

Another important point is the differentiation between *objective* and *subjective risks*. "A so-called 'objective' risk means nothing more, nothing less than a risk constructed by socially authorized experts, that is, an *objective risk* is one that has been scientifically constructed using the best available data and knowledge" (Vaz, 2004, p. 117). In turn, "the perceived risk would be based on subjective impressions" (Vaz, 2004, p. 117).

From an objective point of view, Vaz (2004, p. 113) states that "risk designates an epistemological relationship of partial knowledge of the future", and "[...] there would be no point in talking about risk if the concept had not included the effort [and, we would add, the real possibility] of avoiding the undesirable. If we had not – since the 17th century, albeit in part, unfortunately – overcome the *fatalistic acceptance of future events*, there would be no risk, concludes Vaz. The future, which is relatively unknown, has to be, even if partially, transformable, he says.

Here, an observation is in order. When we point to the limits of subjective perception, we are not postulating that it is possible to assess risks independently of it. Subjective factors, relating to individuals or groups, are unavoidable. Research has been carried out in several fields, indicating, in different

<sup>4</sup> **Figure 1.** Uncertainty in its various guises. Illustrating sources of uncertainty and situations of decision making under uncertainty, using an urn model. (A) Uncertainty can reside in the mind of the boundedly rational agent. Uncertainty can also result from the decisions of and influences from other agents and from genuine randomness in the external environment (i.e., the data-generating process). (B) Examples of dynamic environments that involve changes in the decision-making situation over time. Left: The proportion of balls changes in unpredictable (or unknown) ways over time; therefore, probability estimates obtained at t1 are of little use at t2. Right: The outcomes themselves change over time, requiring a reformulation of the decision situation. (C) Examples of decision-making scenarios. From left to right: In situations of certainty and risk, the outcomes and their probabilities are known. In a 'black swan' situation, the urn contains a rare but highly consequential event (a 'bomb' or, in the case of a positive event, a 'diamond') that is either unknown to the decision maker or ignored in the representation of the decision situation. In a situation of Knightian uncertainty, the outcomes are known but not their probabilities. The right-most example is a situation of radical uncertainty, in which both the outcomes and their probabilities are unknown (Meder, Lec and Osman, 2016, p. 259).



and sometimes conflicting ways, that social, educational, cognitive, genetic, neurobiological, gender, age factors, etc. are correlated not only with perception, but also with the preference for low or high risk activities in the most diverse domains (Chilton *et al.*, 2002; Bonem, Ellsworth and Gonzalez, 2015; Nicolaou and Shane, 2019; Bouchouicha *et al.*, 2019; Andersson *et al.*, 2020; Gross *et al.*, 2021; Globig, Blain and Sharot, 2022). That takes Nicolaou and Shane (2019, p. 261) stating that “the description of people as risk-taking or risk-avoiding types is not a rhetoric flourish”.

For Hacking (2003), each of us has a kind of *risk portfolio*, that is, a standard set of precautions to which we are exposed or presented from a very early age (as an example, he cites looking at both sides before crossing the street), generating a *special care* (related to a special fear) with some things. “The ‘portfolio’ of risks changes markedly over time. The choice of risks to worry about is rarely determined by what experts assure us are the ‘objective’, ‘real’ probabilities and disutilities of the dangers” (Hacking, 2003, p. 22). Resorting to some important historical facts that could not have been foreseen by any general theory of risk, Hacking highlights the importance of identifying and understanding collective risks, stating that “what we might hope for is an understanding of necessary conditions under which a collective risk should become high priority, part of the communal risk portfolio of causes to work for (or against), or at least to worry about” (Hacking, 2003, p. 22).

Regarding the elucidation of the conceptual problem of risk, we can take advantage of the contribution by Nick Bostrom, which includes the new concept of existential risk:

*Existential risk – One where an adverse outcome would either annihilate Earth-originating intelligent life or permanently and drastically curtail its potential. An existential risk is one where humankind as a whole is imperiled. Existential disasters have major adverse consequences for the course of human civilization for all time to come (Bostrom, 2002, p. 2).*

According to Bostrom (2002, p. 1), “it’s dangerous to be alive and risks are everywhere. Luckily, not all risks are equally serious”. The magnitude of a risk should be measured by three dimensions: scope, intensity and probability. “By ‘scope’ I mean the size of the group of people that are at risk. By ‘intensity’ I mean how badly each individual in the group would be affected. And by ‘probability’ I mean the best current subjective estimate of the probability of the adverse outcome” (Bostrom, 2002, p. 1).

Bostrom (2002) proposes a *typology of risk*, dividing it into six types, taking into account the first two dimensions (scope and intensity). In terms of scope (or reach), risks can be personal, local or global. In terms of intensity, they can be bearable<sup>5</sup> or terminal. According to him,

*“Personal”, “local”, or “global” refer to the size of the population that is directly affected; a global risk is one that affects the whole of humankind (and our successors). “Endurable” vs. “terminal” indicates how intensely the target population would be affected. An endurable risk may cause great destruction, but one can either recover from the damage or find ways of coping with the fallout. In contrast, a terminal risk is one where the targets are either annihilated or irreversibly crippled in ways that radically reduce their potential to live the sort of life they aspire to. In the case of personal risks, for instance, a terminal outcome could for example be death, permanent severe brain injury, or a lifetime prison sentence. An example of a local terminal risk would be genocide leading to the annihilation of a people (this happened to several Indian nations). Permanent enslavement is another example (Bostrom, 2002, pp. 1-2).*

<sup>5</sup> “To say that a particular global risk is endurable is evidently not to say that it is acceptable or not very serious. A world war fought with conventional weapons or a Nazi-style Reich lasting for a decade would be extremely horrible events even though they would fall under the rubric of endurable global risks since humanity could eventually recover” (Bostrom, 2002, p. 2).

Although the conceptual problem is a fundamental aspect, the questions around risk are far from being closed there. Risk analysis or assessment is a complex and highly demanding field of study in many aspects. It is beyond the scope of this paper to detail them. Thus, we close this section by returning to epistemic aspects, making brief considerations on this point, completing what we highlighted above about the epistemic differences between risk and uncertainty.

According to Rose (2007), *risk assessment*, which has a long history, involves the search for the identification of factors that delimit *groups, behaviors or risk profiles* on which preventive or prophylactic interventions will be made. This requires the establishment of risk measurement units, from which future occurrences can be calculated and predicted. In this sense, "risk [...] denotes a family of ways of thinking and acting that evolve calculations about probable futures in the present followed by interventions into the present in order to control that potential future" (Rose, 2007, p. 70). When mentioning the *statistical inference techniques that calculate risk*, Vaz (2004, p. 114) states that "[...] no individual has zero risk in relation to something; there are only groups with different levels of risk".

Estimates are not certainties. So they can fail. And, according to some experts, estimates, especially when referring to risks that would imply major catastrophes (such as existential risks), depend on subjective judgments and can be very imprecise, so that "the most reasonable estimate might be substantially higher or lower" (Bostrom, 2013, p. 15). Thus, "[...] perhaps the strongest reason for judging the total existential risk within the next few centuries to be significant is the extreme magnitude of the values at stake. Even a small probability of existential catastrophe could be highly practically significant" (Bostrom, 2013, p. 15).

In a paper suggestively titled *probing the improbable*, Ord, Hillerbrand and Sandberg (2010) highlights the methodological challenges related to catastrophic risks, with low probability, but which receive *high stakes*. The authors focus on estimating the probabilities of global catastrophic risks, postulating that the approach developed by them is more useful than the dichotomy present in risk assessment between *model and parameter uncertainties*. From a complex analysis of the relationship between theories, models and their calculation methods, they conclude that

*When estimating threat probabilities, it is not enough to make conservative estimates (using the most extreme values or model assumptions compatible with known data). Rather, we need robust estimates that can handle theory, model and calculation errors. The need for this becomes considerably more pronounced for low-probability high-stake events, though we do not say that low probabilities cannot be treated systematically (Ord, Hillerbrand and Sandberg, 2010, p. 202).*

When one recognizes the scope for flaws in science-based risk estimates, it seems unreasonable to succumb to the view that prophecies and hunches are, at bottom, epistemic equivalents to estimable risk. If there are limits in scientific models of risk assessment/analysis/estimation, the way out is to advance in the formulation of more adequate and accurate models, and not give up scientific rationality in the name of fatalistic subjectivism.

The concept of risk has emerged together with the process of rationalization of the world and the scientific and technological advances that operated some of the most important inflections in the history of mankind. The paradox is that we need to resort to reason and science to deal with the risks they themselves create.

## 4. The Society of Risk

According to Moynihan, the possibility of intervening in the world through rational action takes us back to the Enlightenment, a period in which there was the consolidation "[...] of the various scientific vocabularies requisite for the first explicit prognoses on existential catastrophe" (Moynihan, 2020, p. 1). Also,

such concern was configured in a period from which *being rational* and *being responsible* became like 'faces of the same coin' and in which it was possible to start to separate values and facts (Moynihan, 2020).

In a scenario of secularization, humans were given the task of thinking about moral values and their prescriptive norm character as something distinct from natural facts or a higher order. Associated with this, it was given the power to transform the world, as well as the burden of being responsible, as a species, for its effects. Ben-Haim's claim that "the freedom to decide is an opportunity to err; but every opportunity is also a potential for success" (Ben-Haim, 2006, p. 1) seems, despite its cliché tone, to sum up this new anthropological framework well. Ulrich Beck and Anthony Giddens, in their classic approaches to the *risk society* and the relationship between *risk and responsibility*, reinforce this understanding.

According to Beck, the fundamental aspects of the risk society can be presented in the following terms:

*Risk society begins where nature ends. [...] this is where we switch the focus of our anxieties from what nature can do to us to what we have done to nature. A central paradox of risk society is that these internal risks are generated by the processes of modernization which try to control them.*

[...]

*Risk society begins where tradition ends, when, in all spheres of life, we can no longer take traditional certainties for granted. The less we can rely on traditional securities, the more risks we have to make* (Beck, 1998, p. 10).

In dialogue with Beck, Giddens states that

*A risk society is a society where we increasingly live on a high technological frontier which absolutely no one completely understands and which generates a diversity of possible futures. The origins of risk society can be traced to two fundamental transformations which are affecting our lives today. Each is connected to the increasing influence of science and technology, although not wholly determined by them.* (Giddens, 1999, p. 3).

So, according to Giddens,

*To analyse what risk society is, one must make a series of distinctions. First of all, we must separate risk from hazard or danger. Risk is not, as such, the same as hazard or danger. A risk society is not intrinsically more dangerous or hazardous than pre-existing forms of social order. [...] The idea of risk is bound up with the aspiration to control and particularly with the idea of controlling the future* (Giddens, 1999, p. 3).

This process that Beck and Giddens mention can be called *detraditionalization* (or *secularization*), which, according to Heelas,

*[...] involves a shift of authority: from 'without' to 'within'. It entails the decline of the belief in pre-given or natural orders of things. Individual subjects are themselves called upon to exercise authority in the face of the disorder and contingency which is thereby generated. 'Voice' is displaced from established sources, coming to rest with the self.* (Heelas, 1996, p. 2).

In this vein, Furedi (2007) claims that, *in practically all areas of society*, we see an *explosion of risks*, a social phenomenon that is at the basis of what the author calls *Culture of Fear*. According to the author, "one of the principal features of our culture of fear is the belief that humanity is confronted by powerful destructive forces that threaten our existence. With so much at stake how can responsible people fail to raise the alarm?" (Furedi, 2007, p. x).

For Beck, in modern, detraditionalized or risky societies, there is both a multiplicity of definitions and an *increasing number of risks*:

*There occurs, so to speak, an over-production of risks, which sometimes relativize, sometimes supplement and sometimes outdo one another. [...] This pluralism is evident in the scope of risks; the urgency and existence of risks fluctuate with the variety of values and interests.*

*[...]*

*The causal nexus produced in risks between actual or potential damaging effects and the system of industrial production opens an almost infinite number of individual explanations (Beck, 1992, p. 31).*

A kind of omnipresence of science and risk in people's daily lives creates an increasingly complex challenge, namely, dealing in an enlightened way with the relationship between them. In this sense, we turn to Vaz again, who highlights some implications of the *pre-eminence of risk in contemporary culture*:

*[...] more and more individuals use scientific knowledge when organizing their lives. A central feature of society is the new relationship between lay actors and experts [...]. Science is more present in our minds than ever, which poses a problem of credibility for scientists, insofar as they can differ on the risks that exist and how much we should be concerned about them, and insofar as their opinion about risks becomes the basis for public policy. [...] The concept of risk] points to the weight of science in politics and everyday life, as well as the complex play of forces between experts, social movements, the State and the media [...] (Vaz, 2004, p. 117).*

This cultural background provides the breeding ground for the heuristic of fear and the anti-enhancement rhetoric. But at the same time, it can be considered a basic motivation for human enhancement. In fact, faced with the increasing risks that the individual has to face in a complex society for which he or she is not prepared, does not have sufficient physical and cognitive endowments, instruments that favour forms of enhancement can diminish risks and not increase them, if one correctly understands the type of risk that is at stake. Take for example the anti-enhancement argument that natural evolution has found a good balance for the human being and it is therefore risky to change it with direct interventions of which we do not know the possible side effects or undesirable effects.

This is certainly an argument that takes into account robust scientific aspects, such as fitness in one's environment. However, it is also an argument that does not adequately consider all aspects involved. Indeed, very rapid technical progress has already radically changed the physical and social environment in which we move. This means that the endowments we have developed in the course of evolution through adaptation mechanisms are no longer well matched to our new environment. Natural evolution cannot keep pace with scientific and technological progress, so the risks we run by not trying to enhance ourselves may be greater than the risks we would run if we did not try to improve our physical and cognitive endowments to make them better suited to our environment.

In this sense, a better understanding of the concept of risk and of the epistemic context in which risk assessment is carried out can be of help in avoiding cognitive traps by which we end up exaggerating the risk associated with human enhancement via technological tools compared to other, more real types of risks that accompany us every day in our environment.

## 5. Considering Reasonable Risks

Since risk is a key point in moral reflection and in the debate on technological human enhancement, we draw attention to the existence of a very eclectic and interdisciplinary field of studies that has produced a significant amount of approaches and knowledge on the topic, which points to its complexity

and also to the impropriety of rhetorical simplifications, that is, mere assertions that there is a risk in a specific activity and that this would serve to proscribe it.

Given the relationships between science, technology, risk, harm, politics, morals, decision-making and public policies today, it is imperative to overcome the simplistic way that lacks minimal epistemic consistency that characterizes a far from irrelevant part of the uses of the term risk in debate over HE.

So, according to More, “The Proactionary Principle emerged out of a critical discussion of the well-known precautionary principle that developed in Europe and has been used in the United States and elsewhere as a type of model for dealing with change” (More, 2013, p. 259). Based on the formulation of Soren Holm and John Harris, The Precautionary Principle states that:

*When an activity raises threats of serious or irreversible harm to human health or the environment, precautionary measures that prevent the possibility of harm shall be taken even if the causal link between the activity and the possible harm has not been proven or the causal link is weak and the harm is unlikely to occur (More, 2013, p. 259).*

The advocates of the Proactionary Principle identify at least six pain points in The Precautionary Principle, related to assuming worst-case scenarios; distracting attention from established threats to health, especially natural risks; assuming that the effects of regulation and restriction are all positive or neutral, never negative; ignoring potential benefits of technology and inherently favoring nature over humanity; illegitimately shifting the burden of proof and unfavorably positioning the proponent of the activity; conflicting with more balanced, common-law approaches to risk and harm.

Also, according to Harris (2009), although it is widely accepted, the Precautionary Principle is inconsistent. Harris lists a series of questionable assumptions present at the beginning, concluding that it does not offer a rational basis that contemplates, at the same time, the defense of the interruption of intervention on something (*leaving things alone*) and prioritization of what is already given (*status quo*). For Harris,

*[...] it is unclear why a precautionary approach should apply only to proposed changes rather than to the status quo. In the absence of reliable predictive knowledge as to how dangerous leaving things alone may prove, we have no rational basis for a precautionary approach which prioritizes the status quo (Harris, 2009, p. 133).*

However, one must consider that fear heuristics is a strategy that comes from our evolutionary history and proves to be adaptive in so many ways, as has been shown recently:

*Already as infants humans are more fearful than our closest living primate relatives, the chimpanzees. Yet heightened fearfulness is mostly considered maladaptive, as it is thought to increase the risk of developing anxiety and depression. How can this human fear paradox be explained? The fearful ape hypothesis [...] stipulates that, in the context of cooperative caregiving and provisioning unique to human great ape group life, heightened fearfulness was adaptive. This is because from early in ontogeny fearfulness expressed and perceived enhanced care-based responding and provisioning from, while concurrently increasing cooperation with, mothers and others (Grossmann, 2022).*

So, if it seems important that the Heuristic of Fear is replaced by other approaches, which are more consistent at epistemic level (although not infallible), in order to properly deal with the risks of HE, one cannot forget the reality of the Heuristic of Fear and its role in society.

A relevant proposal has been put forward by Bostrom and Sandberg (2009). The so-called *Evolutionary Heuristic*:



*By understanding both the sense in which there is validity in the idea that nature is wise and the limits beyond which the idea ceases to be valid, we are in a better position to identify promising human enhancements and to evaluate the risk-benefit ratio of extant enhancements. If we are right in supposing that intuitions about the wisdom of nature exert an inarticulate influence on opinion in contemporary bioethics of human enhancement, then the evolution heuristic – while primarily a method for addressing empirical questions – may also help to inform our assessments of more normatively loaded items of dispute (Bostrom and Sandberg, 2009, pp. 408-409).*

So, the prospects of human enhancement must be assessed in the light of a realistic understanding of human psychology, which in turn derives from the evolution of species in a very different environment. In that environment, the concept of risk was only implicit and manifested in the instinctual reactions that individuals had learned over generations as useful for the survival and propagation of offspring.

Today, those psychological mechanisms, such as the heuristics of fear, are still part of us but have to deal with a radically changed environment compared to the savannah where homo sapiens evolved. That is why, when the possibility of empowering human beings in the face of new environmental challenges becomes available, the inherited psychological mechanisms may prove dysfunctional and must be contextualised and sterilised within a scientific framework of a correct understanding of risk and its dimensions in contemporary society.

This does not mean that every form of human empowerment is in itself beneficial, without real risks to be pursued. Risks are always present and must be correctly presented, without them becoming a form of preventive rejection of any kind of intervention.

There is no doubt that the empirical-normative issues surrounding the risks related to HE are many and complex. Dealing with such issues demands, in part, anticipatory or preparatory research (Ferrari, Coenen and Grunwald, 2012; Ankeny, Munsie and Leache, 2021). This imposes limits and challenges, such as those referring to the link between risks and values. There is a need to foresee not only the risks, from the point of view of possible threats to what we currently value. Furthermore, we have to consider the need to systematically investigate future axiological changes, anticipating future trajectories (*space of possible axiological trajectories*), and how this would interfere with risk analysis, a question raised by Danaher (2021) recently, naming this effort *axiological futurism*.

Finally, it should not be forgotten that some criticisms of human enhancement refer to questions of principle that are reducible to an estimation of the risk involved in the practice under consideration. A form of opposition to human enhancement is therefore not included in the discourse on risk as such. However, we believe that an informed and rigorous approach to the issue of risk and its measurement is a fundamental contribution to the debate on human enhancement through technological means and should find further development beyond the contribution offered here.

## References

- ANDERSSON, O. et al. 2020. Robust Inference in Risk Elicitation Tasks. *Journal of Risk and Uncertainty*, **61**: 195-209.
- ANKENY, R. A.; MUNSIE, M. J.; LEACH, J. 2021. Developing a Reflexive, Anticipatory, and Deliberative Approach to Unanticipated Discoveries: Ethical Lessons from iBlastoids. *The American Journal of Bioethics*, **22**(1): 36-45.
- AVEN, T. et al. 2018. *Society for Risk Analysis Glossary*. Available at: <https://www.sra.org/wp-content/uploads/2020/04/SRA-Glossary-FINAL.pdf>. Accessed on: april 27, 2022.
- BECK, U. 1992. *Risk Society: Towards a New Modernity*. London, SAGE Publications, 260 p.
- BECK, U. Politics of Risk Society. 1998. In: J. FRANKLIN (ed.), *The Politics of Risk Society*. Cambridge,

- UK, Polity Press, p. 9-22.
- BEN-HAIM, Y. 2006. *Info-Gap Decision Theory: Decisions Under Severe Uncertainty*. London, Academic Press, 360 p.
- BONEM, E. M.; ELLSWORTH, P. C.; GONZALEZ, R. 2015. Age Differences in Risk: Perceptions, Intentions and Domains. *Journal of Behavioral Decision Making*, **28**(4): 317-330.
- BOSTROM, N. 2002. Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards. *Journal of Evolution and Technology*, **9**(1): 1-29.
- BOSTROM, N. 2013. Existential Risk Prevention as Global Priority. *Global Policy*, **4**(1): 15-31.
- BOSTROM, N.; SANDBERG, A. 2009. The Wisdom of Nature: An Evolutionary Heuristic for Human Enhancement. In: J. SAVULESCU; N. BOSTROM (eds.), *Human Enhancement*. Oxford, Oxford University Press, p. 375-416.
- BOUCHOUICHA, R. et al. 2019. Gender Effects for Loss Aversion: Yes, No, Maybe? *Journal of Risk and Uncertainty*, **59**(2): 179-184.
- BUCHANAN, A. 2011. *Beyond Humanity? The Ethics of Biomedical Enhancement*. Oxford, Oxford University Press, 286 p.
- CAMPBELL, L.; CLARKE, P.K. 2018. Making Operational Decisions in Humanitarian Response: A Literature Review. ALNAPPAPER. Available at: <https://www.alnap.org/help-library/making-operational-decisions-in-humanitarian-response-a-literature-review>. Accessed on: april 27, 2022.
- CHILTON, S. et al. 2002. Public Perceptions of Risk and Preference-Based Values of Safety. *Journal of Risk and Uncertainty*, **25**(3): 211-232.
- DANAHER, J. 2021. Axiological Futurism: The Systematic Study of the Future of Values. *Futures*, **132**: 1-14.
- FERRARI, A.; COENEN, C.; GRUNWALD, A. 2012. Visions and Ethics in Current Discourse on Human Enhancement. *Nanoethics*, **6**: 215-229.
- FUREDI, F. 2007. *Culture of Fear Revisited: Risk-taking and Morality of Low Expectation*. London, Continuum, 212 p.
- GIDDENS, A. 1999. Risk and Responsibility. *The Modern Law Review*, **26**(1): 1-10.
- GLOBIG, L. K.; BLAIN, B.; SHAROT, T. 2022. Perceptions of Personal and Public Risk: Dissociable Effects on Behavior and Well-Being. *Journal of Risk and Uncertainty*, <https://doi.org/10.1007/s11166-022-09373-0>.
- GORDON, J-S.; BURCKHART, H.; SEGLER, P. 2014. Introduction. In: J-S. GORDON; H. BURCKHART (eds.). *Global Ethics and Moral Responsibility: Hans Jonas and His Critics*. New York, Routledge, p. 1-8.
- GROSS, J. et al. 2021. When Helping is Risky: The Behavioral and Neurobiological Trade-off of Social and Risk Preferences. *Association for Psychological Science*, **32**(11): 1-14.
- GROSSMANN, T. 2022. The Human Fear Paradox: Affective Origins of Cooperative Care. *Behavioral and Brain Sciences*, in press, <https://doi.org/10.1017/S0140525X2200067X>.
- HACKING, I. 2003. Risk and Dirt. In: R. V. ERICSON; A. DOYLE (eds.), *Risk and Morality*. Toronto, University of Toronto Press, p. 22-47.
- HARRIS, J. 2009. Enhancement Are a Moral Obligation. In: J. SAVULESCU; N. BOSTROM (eds.), *Human Enhancement*. Oxford, Oxford University Press, p. 131-154.
- HEELAS, P. 1996. Introduction: Detraditionalization and Its Rivals. In: P. HEELAS; S. LASH; P. MORRIS (eds.), *Detraditionalization: Critical Reflections on Authority and Identity at a Time of Uncertainty*. Cambridge, MA, Blackwell Publishers, p. 1-20.
- JONAS, H. 1984. *The Imperative of Responsibility: In: Search of an Ethics for the Technological Age*.

- Chicago, The University of Chicago Press, 255 p.
- JONAS, H. 1997. *Técnica, medicina y ética: sobre la práctica del principio de responsabilidad*. Barcelona, Paidós Ibérica Edicioones, 206 p.
- KAHANE, G. 2014. Our Cosmic Insignificance. *Noûs*, **48**(4): 745–772.
- KEKES, J. 1998. *A Case for Conservatism*. London, Cornell University Press, 239 p.
- KNIGHT, F. H. 1921. *Risk, Uncertainty and Profit*. Boston and New York, Houghton Mifflin Company, 381 p.
- LAVAZZA, A.; COLZATO, L. S. 2018. Editorial Special Topic: Neuroethical Issues in Cognitive Enhancement. *Journal of Cognitive Enhancement*, **2**(4), 319-322.
- MEDER, B.; LEC, F.L.; OSMAN, M. 2013. Decision Making in Uncertain Times: What can Cognitive and Decision Sciences Say about or Learn from Economic Crises? *Trends in Cognitive Sciences*, **17**(6): 257-260.
- MORE, M. 2013. The Proactionary Principle: Optimizing Technological Outcomes. In: M. MORE; N. VITA-MORE (eds.). *The Transhumanist Reader: Classical and Contemporary Essays on the Science, Technology, and Philosophy of the Human Future*. Chichester, UK, Wiley-Blackwell, p. 258-267.
- MOYNIHAN, T. 2020. Existential Risk and Human Extinction: An Intellectual History. *Futures*, **116**: 1-13.
- NICOLAOU, N.; SHANE, S. 2019. Common Genetic Effects on Risk Taking Preferences and Choices. *Journal of Risk and Uncertainty*, **59**(3): 261-279.
- ORD, T.; HILLERBRAND, R.; SANDBERG, A. 2010. Probing the Improbable: Methodological Challenges for Risk with Low Probabilities and High Stakes. *Journal of Risk Research*, **13**(2): 191-205.
- ROACHE, R.; CLARKE, S. 2009. Bioconservatism, Bioliberalism, and The Wisdom of Reflecting on Regugnance. *Monash Bioethics Review*, **28**(1): 04.1-04.21.
- ROSE, N. 2007. *The Politics of Life Itself: Biomedicine, Power, and Subjectivity in the Twenty-First Century*. Princeton, New Jersey, Princeton University Press, 350 p.
- SOUZA, R. 2010. Here's How I Feel: Dont't Trust Your Feelings! In: S. ROESER (ed.), *Emotions and Risky Technologies*. Dordrecht, Springer, p. 17-36.
- VAZ, P. 2004. Risco e justiça. In: T. C. B. CALOMENI (org.). *Michel Foucault: entre o murmúrio e a palavra*. Campos, RJ, Editora Faculdade de Direito de Campos, p. 101-132.
- VELLANI, K. H. 2022. *Unraveled: An Evidence-Based Approach to Understanding and Preventing Crime*. Texas, Threat Analysis Group, 293 p.
- VILAÇA, M. M. 2021. Against Perfection, Human Enhancement or for the Gift? An Analysis of Michael Sandel's Argument on Genetic Engineering. *Síntese*, **48**(152): 779-805.

Submitted on December 28, 2021.

Accepted on May 13, 2022.