

Reasonable precaution or unjust discrimination? Applying a lexical utility model of the precautionary principle to moral choices.

Thomas Boyer-Kassem* and Sébastien Duchêne†

working paper — July 10, 2023

Abstract

In some applications to human beings, the precautionary principle seems to raise specific ethical concerns. For instance, it has been used to exclude men who have sexual relations with men from donating blood on the basis of a higher HIV prevalence in this group. Does not the precautionary principle here amount to justifying a form of discrimination? To help describe what is at stake from a decision-theoretic viewpoint and to consider the possibility of ethical catastrophes, we offer a lexical utility model of the precautionary principle, which generalizes a previous model by Bartha and DesRoches (2021) and Steel and Bartha (2022). Traditional expected utility theory also comes as a special case of the model. Our proposal opens new perspectives for considering ethical dilemmas within a precautionary approach.

*Université de Poitiers, MAPP, Philosophy Department, Poitiers, France. E-mail: thomas.boyer.kassem@univ-poitiers.fr. ORCID: 0000-0002-7621-0253

†Montpellier Business School, Montpellier, France. Email: s.duchene@montpellier-bs.com. ORCID: 0000-0002-3781-7677

1 Introduction

The precautionary principle (PP), as it offers to guide decisions when uncertainty prevails in the context of potential catastrophes, has attracted much discussion over the last decades, both in society and in academia. In its Wingspread version, PP states that “When an activity raises threats of harm to human health or the environment, precautionary measures should be taken even if some cause and effect relationships are not fully established scientifically.” (Ashford et al. 1998) That is, PP recommends taking action based on incomplete knowledge in a way that departs from traditional decision theory. As a consequence, it has been under attack since its proposal in the 1990s. Here, we address a new class of problems PP may raise, which are ethical in nature. According to the analysis we present, tension seems to arise between what PP recommends and what moral principles recommend, in cases where PP is applied to human beings, i.e., when the uncertainty (crucial to the triggering of PP) bears on a human being and, consequently, when the recommended action impacts these human beings. For instance, is it morally permissible to exclude men who have sexual relations with men from donating blood out of precaution that they could be HIV positive? Or to have people who are classified as dangerous by the intelligence service be remanded in custody out of precaution? PP seems here to conflict with moral principles about human beings, such as non-discrimination or presumption of innocence. As far as we know, this paper is the first one which studies the moral tension that stems from applying PP to human beings.

How should the conflict be solved? Can a general solution be recommended, or should it be on a case-by-case basis? These questions are particularly pressing, both fundamentally to better understand the legitimacy of PP, and from a social viewpoint, to address volatile issues — e.g. gay exclusion from blood giving is regularly in the headlines. We do not think that there is a general answer to the question of whether precautionary considerations should always prevail over discrimination ones, or *vice versa* — because the concrete cases can vary in many ways, making this or that moral feature take prominence, and also because different people may legitimately have different views on whether an event qualifies as a (moral) catastrophe. Anyway, we do not argue here in favor of a general solution for or against the PP being applied to human beings, not even for or against PP *tout court*. Instead, we aim at providing a theoretical framework to model how the moral tension we have identified can be resolved. Our model takes as input agents’ judgments about what they consider as (ethical) catastrophes or just ordinary events and then tells which decision should be made given these preferences. In that sense, the present work is not primarily ethical in nature, but decision-theoretical. It is normative *given agents’ preferences*, and does not defend any particular moral theory.

The core idea of our model is that agents may judge that some outcomes qualify as catastrophes of various kinds (in particular, ethical catastrophes), and put these

kinds of catastrophes in a hierarchy in various ways. People may either give absolute priority to the avoidance of one type of catastrophe or regard different catastrophes as comparable harms that can be traded off against each other. For instance, someone may consider that limiting the freedom of a suspect who has not been convicted is a catastrophe that is acceptable in no circumstance, while someone else may consider that it is, provided it may avoid a terrorist attack. Technically, the utility of an outcome is represented by a vector with several dimensions — belonging to an upper dimension means being a catastrophe of an incomparable nature, which should be avoided in priority. The model we propose is a lexical utility model of PP which generalizes a previous model by Bartha and DesRoches (2021) and Steel and Bartha (2022). Our model displays several interesting features. First, like in the above works, traditional expected utility theory comes as a special case of the model, which means that PP can have a peaceful relationship to expected utility theory, contrarily to a traditional view (Boyer-Kassem 2017a, 2017b, Christiansen 2019, Gollier et al. 2000, Peterson 2006, 2017). Second, our model invites to reconsider which questions are worth asking oneself in the morally tricky cases of PP: whether catastrophes can be identified, and if so, whether they can be ranked on some scale of incomparability. The question is not whether “PP should prevail” anymore. Our proposal thus opens new perspectives for considering ethical dilemmas within a precautionary approach.

The paper is organized as follows. Section 2 introduces PP and an existing lexical utility model of it. Section 3 presents cases of application of PP to human beings and investigates the common source of the moral tension. Section 4 presents our suggestion of a generalized lexical utility model with an additional ethical utility dimension.

2 The precautionary principle and its moral critics

2.1 Presentation of the precautionary principle(s)

It is generally considered that decision-making in policy should be driven by scientific evidence. For instance, if some product is scientifically shown to be noxious, then the use of that product should be restricted (*ceteris paribus*). But what if scientific studies are absent or cannot give a definite answer regarding the product’s dangerousness? A new decision rule was proposed in the 1980s as PP, with the general idea that, when facing a threat of particular harm, partial evidence only can be sufficient to take precautionary measures. For instance, if studies suggest that the use of neonicotinoids might be harmful to the bee population, but do not manage to establish it fully, then the use of these neonicotinoids should be restricted.

PP is considered here as a *decision rule*, i.e., a rule which gives as an outcome a recommended action (e.g., restrict the use).¹ PP involves three dimensions: (i)

¹Ahteensuu and Sandin (2012, p. 971-2) distinguish other understandings of PP, as an epistemic

the harm condition (e.g., several deaths), (ii) the knowledge condition (e.g., partial scientific evidence), and (iii) the recommended action (e.g., restriction of use). Depending on how these dimensions are tuned, this gives a bunch of *versions* of PP, not just *one* PP, ranging from “If it is almost certain from an action that a global catastrophe will occur, the action should be slightly diminished” to “If there is a slight doubt that an action can lead to some harm, the action should be prohibited”. The Wingspread version (cf. Section 1) falls in between. In this paper, we do not take a stand on which versions of PP are acceptable and discuss various ones depending on the cases.

The relation between PP and standard decision theory (i.e., expected utility theory or cost-benefit analysis) is a debated question. Opponents of PP argue that PP is unscientific, irrational, inconsistent with standard decision theory, or that it aims at replacing it (e.g., Harris and Holm 2002, Peterson 2006, 2017, Sunstein 2005), while proponents argue that these charges are undeserved (e.g., Ahtensuu 2007, Boyer-Kassem 2017a, 2017b, Sandin et al. 2002, Sandin 2006, Steel 2015) and that PP is only supplementing or extending, not contradicting, standard decision theory in areas where it is silent (e.g., Bartha and DesRoches 2021, Christiansen 2019, Gollier, Jullien and Treich 2000). It is often considered that PP does not apply if probabilities can be meaningfully attributed to possible outcomes, i.e., applying PP requires some “scientific uncertainty” about the occurrence of the harm. In this picture, expected utility theory prevails when probabilities can be attributed, and PP can step in only if they cannot. Interestingly, Bartha and DesRoches (2021) and Steel and Bartha (2022) extend PP to cases where probabilities can be defined (more on that below).

PP has been initially proposed for harms that belong to human health and the environment, e.g., to avoid illnesses or oil spills. More recent and tentative applications have considered preemptive war (Sunstein 2005, Steel 2013, 2015), terrorism (Sarkozy in Europe 1, 2016), or personal safety (Steel 2015). In Section 3, we consider examples from both human health and security, but this does not mean that we are taking a stand on whether PP should indeed be extended to security. Our proposal in Section 4 applies equally well should PP be limited to human health. So, a reader skeptical of applying PP to security issues will still find the remainder of this paper profitable.

Can PP be formalized? Peterson (2006; 2017) and then Steel (2015, Appendix, p. 218-224) answer affirmatively by offering qualitative decision-theoretic formulations of PP, not without critics.² We present here in detail Bartha and DesRoches’ (2021) proposal (summarized and applied in Steel and Bartha 2022), as we shall reuse it in Section 4.³ Their model abandons a very classical assumption of classical

rule or a procedural requirement. For reviews on PP more generally, cf. Stirling (2009), Ahtensuu and Sandin (2012), Munthe (2020).

²Boyer-Kassem (2017a; 2017b) argues that Peterson’s proposal is inadequate. Bartha and DesRoches (2021) identify a few problems in Steel’s account and regret that it is not quantitative.

³For justifications of the foundations of lexical utility models in general, cf. these two articles,

decision theory, namely the continuity axiom for preferences. This axiom states that if someone prefers outcome C to outcome B to outcome A , then there exists some $\lambda \in [0, 1]$ such that the person is indifferent between B and the gamble which gives C with probability λ and A with probability $1 - \lambda$. It expresses the idea that all outcomes are commensurable and that their utilities can be represented on a continuous line. Now, PP brings the idea that catastrophic outcomes are special, and Bartha and DesRoches suggest they cannot be represented on such a continuous line. They represent utilities in a discontinuous way, typically on a two-level lexical utility model: the first level represents the catastrophic part of outcomes and the second level their non-catastrophic part. Mathematically speaking, the utility of an outcome O is represented as a two-dimension vector, noted $(u_1(O), u_2(O))$; $u_1(O)$ is worth -1 if the outcome is catastrophic, 0 otherwise;⁴ $u_2(O)$ is the outcome's traditional utility. The decision rule of this model is: to maximize expected *lexical* utility. That is, consider the first component of the outcomes utility and compute the expected utility of each act. If one act maximizes it, then this act is recommended. In case of draws, turn to the second component of utilities, compute the expected utility of each act, and select the act which maximizes it. This presentation suggests that outcomes must receive meaningful probabilities, contrary to the traditional view that PP only applies when probabilities cannot be given. But Bartha and DesRoches argue that the probabilities need not be precisely known and can just be theoretically postulated. They offer examples where their model recommends an act on the simple condition that the probability of an outcome is not null. They also show that the model can be expanded to imprecise probabilities. So, the requirement of probabilities to compute expected utilities should not be viewed as a limitation of the model.

It is worth noting that if all outcomes have null u_1 , then BDR's model boils down to a traditional utility model, which respects the continuity axiom, and the decision rule is just the maximization of expected utility. It is an empirical question, Bartha and DesRoches point out, whether people actually have lexical utilities or not, but this should not be ruled out beforehand. Their model can thus be considered as a generalization of standard utility theory. They show in several examples that their model captures the spirit of PP, leading to the expected precautionary recommendations.

Finally, note that “the lexical model does not need an independently postulated PP. Precautionary reasoning is an automatic consequence of the usual principle to maximize expected utility, together with representation of the qualitative distinction between catastrophic and non-catastrophic outcomes.” (p. 15). That is, the oft-debated question of the rationality of PP receives here a neat answer: precautionary

as well as references therein.

⁴It is up to agents to judge whether an outcome is considered catastrophic. Another view could be to define a catastrophe threshold theoretically, but this is not required here and shall not be assumed.

reasoning can be embedded in a generalization of the standard model and apply even when probabilities make sense.

2.2 Moral critics of the precautionary principle

As this paper investigates moral problems PP can raise, let us picture the landscape of existing moral critics. The upshot of this short review is that existing works have not focused on the moral tension we are after.

Although PP has been the focus of a large body of literature, the works which consider *moral* arguments⁵ are quite meager (cf. Ahtensuu and Sandin 2012, p. 974 or Munthe 2020, p. 7). The literature seems convergent on the fact that traditional ethical theories cannot, as such, criticize PP (Munthe 2020 p. 4; Resnik 2021, chap. 3) and that a specific ethical theory has to be devised to address the novel ethical issues when facing risk situations in which PP applies (Hansson 2013; Munthe 2011, 2020; Steel 2014).

The main moral objection against PP is that it “place[s] extra weight on human-caused risks to health and the environment” (Hourdequin 2007, p. 343), or even amounts to “a lexical ordering according to which harms to the environment or human health are prioritized over everything else” (Steel 2015, p. 82). However, this objection does not hold in front of current explications of PP: in Steel’s (2015) formulation or in Bartha and DesRoches’ (2021), no lexical ordering is presupposed between, say, the environment and the economy. Steel considers a *proportionality* requirement, which makes it possible that economic harms trump environmental harms in applying PP (Steel 2015, p. 82). Bartha and DesRoches (2021) do consider a lexical ordering of harms or utilities, but only between catastrophic and non-catastrophic outcomes, which is not supposed to be identical to the partition between environmental and economic harms. Overall, objecting to PP that preserving the environment is not the highest moral value attacks a straw man.

A few works apply PP by relating it to discrimination issues. Stone (2003) studies the relationship between PP and human rights in an adoption case. The “Communication on the Precautionary Principle” (European Commission, 2000) is discussed by Jensen (2002) and Cheyne (2007). Overall, moral arguments against PP are scarce in the literature and often under-developed. We would like to convince the reader that promising moral tensions to study exist in this domain.

3 Setting the stage: some examples

This Section introduces the moral tension we want to study through four examples, which are diverse in terms of the topic (health or security) and are real or invented. Some readers may think that moral solutions are obvious, for instance,

⁵We do not consider here decision-theoretic arguments, which have been mentioned earlier.

one should just always stick to PP. But we would like these readers to acknowledge that the cases either have raised a moral tension in society (for the real cases) or that it would (for the invented case). Our aim here is not to endorse or criticize the moral objections to PP but just to point at the moral tension.

3.1 Blood donation

Blood giving is subject to particular scrutiny from the public health authorities sector for fear that donors could unknowingly transmit illnesses. Some can be directly detected in the blood sample, but others display a “silent window” after someone has been contaminated, for instance, ten days for HIV. During that window, detecting the illness in the blood is technically impossible, although it will contaminate the recipient. To reduce the risk during the silent window, sanitary agencies rely on questionnaires to exclude donors who display risky profiles. In Europe, “the precautionary principle appears to be the predominant reason behind the set of eligibility criteria” (de Kort et al 2016, p. 101). Men having sex with men (MSM) have long been excluded from donating blood because of a higher HIV prevalence and because the number of new HIV infections each year is much higher compared to the rest of the population. PP is then applied to exclude MSM from donating blood, given the higher possibility that they could transmit HIV to donors.

Some have fought what they consider to be discrimination against homosexuals. The complaint is that the selection criteria is not individual risky behaviors, like a recent change of sexual partner (as it is asked to heterosexuals), but sexual orientation *simpliciter* (cf. Véran 2013, p. 35 for the French case). The public debate is typically framed as a conflict between PP and anti-discrimination views.

3.2 Covid-19 and quarantine

Since the beginning of the COVID-19 crisis, many restrictions on population mobility have been implemented to contain the pandemic. PP has been a widely used approach in the European Union in this respect. For instance, the *Joint European Roadmap towards lifting COVID-19 containment measures* (2020) states that “the restrictive measures introduced by Member States [...] have been based on available information in relation to the characteristics of the epidemiology of the disease and followed a precautionary approach” (cited by Goldner Lang 2021, p. 11). Mobility restrictions have consisted, in particular, in setting up restrictions on entering populations from specific countries only. For example, in May 2020, the Baltic states formed a “ ‘travel bubble’, whereby a group of states allow each other’s citizens and residents to enter freely” (Goldner Lang 2021, p. 20), while residents from other European countries entering these countries had to be isolated for 14 days (Euractiv 2020). Goldner Lang (2021, p. 20) notes that travel bubbles “are discriminatory in character since they treat certain Member States and their nationals more favorably than others.” In addition, these measures can be questioned

as they reduce civil liberties (impossibility to travel, isolation measures), stigmatize specific populations compared to others, and somehow assume a “presumption of sickness”. The European Commission responded by the need to apply PP by being non-discriminatory, in particular towards “geographical origin” (Joint European Roadmap (2000), §6.3.2; see also Communication from the Commission (2020)).

3.3 Sarkozy and the ‘S’-labeled people

In 2016, former French president Nicolas Sarkozy declared: “It seems to me that the precautionary principle applies just as much to the security of the French people as to environmental protection.” He suggested a concrete application: “I ask that any person who is labeled ‘S’ [i.e., who is filed by the intelligence services because he is suspected of threatening the security of the State] be controlled and that the few hundreds of people who display dangerousness criteria [...] be remanded in administrative custody, in virtue of the precautionary principle.” (Europe 1, 2016).

These proposals were not without raising a debate. The French minister of Justice, Jean-Jacques Urvoas, set them aside as it would be a source of arbitrary decisions and would call into question the presumption of innocence, against highly established legal precedents (Le Monde 2016). PP was thus opposed by the argument of civil liberties. This conflict has a clear origin: the presumption of innocence requires that one is considered innocent until proven guilty, with the uncertainty threshold being set very low, generally “beyond reasonable doubt”; conversely, PP will trigger liberty penalizing measures with a much higher threshold of uncertainty — clues possessed by the intelligent services could suffice. In other words, PP expresses a preference for few false negatives (few dangerous people outside, more innocents in custody), whereas the presumption of innocence expresses a preference for few false positives (few innocents in custody, more culprits outside).

3.4 Hiring

Is it permissible to hire in discriminatory ways between ethnic groups for one’s safety? A French Jewish boss claimed so and published a job advertisement which specified that Maghrebi people would be refused “out of precaution”. Sued in court, he developed that from his Jewish viewpoint, every Maghrebi person represents a risk, as she may have relatives who “don’t like” Jewish people and he invoked the Merah case (in 2012, a French-Algerian jihadist named Merah killed seven people among which three children in a Jewish school in the south of France). The boss claimed to apply PP to Maghrebi people as others apply it to GMOs (20 Minutes, 2013, Le Point 2013). The court condemned him to a fine of € 5000 for violating an anti-discrimination law which forbids to select job applicants according to their ethnicity or religion. Even if it contravenes a French law, the Jewish boss’s reasoning seems to formally abide by PP, although with a high uncertainty threshold, as it generalizes to all Maghrebi people the possibility that they may attack Jewish people

because some did it once. When used to trigger specific actions towards people according to their ethnicity, PP frontally contradicts a non-discrimination principle and is particularly stereotypical.⁶

Table 1 overviews the examples discussed in this Section.

Case	Domain	Moral conflict with PP?
Blood donation	Health & Env.	Sexual orientation discrimination
Covid-19 & quarantine	Health & Env.	Discrimination & civil liberties
Sarkozy & ‘S’-labeled	Security	Presumption of innocence
Hiring	Security	Discrimination (racial, religious, ...)

Table 1: Overview of the examples discussed.

3.5 A common source of the moral tension?

The moral objections to applying PP in the above cases are somewhat diverse: non-discrimination, civil liberties, or presumption of innocence. What unites them is that the moral tension always stems from applying PP *to human beings*.⁷ By this, we mean two things: first, the knowledge condition of PP (cf. Section 2.1) is about whether some human being can be the *source* of the harm;⁸ second, the recommended precautionary action negatively impacts these human beings.

One might object that the uncertainty is not always about a person, but about an object: it is whether an HIV molecule or a Covid molecule is present somewhere. But the uncertainty relates the molecule *and* the person in an inseparable way: the question is not whether some molecule is dangerous wherever it is, but whether some people because they could bear the molecule, could (indirectly) be dangerous.

The tension does not simply stems from an undue application of PP to security issues since it still arises in health cases like in 3.1 or 3.2. And it does not either

⁶There is a large literature, both theoretical and empirical, on racial discrimination in the labor market and the emergence of taste-based and statistical discrimination concepts (for reviews, see Lang and Lehmann, 2012, or Lang and Spitzer, 2020) and on ways to address it such as affirmative action (for a review, see Arcidiacono and Lovenheim, 2016). Sociologists have also provided a broader perspective through the concepts of institutional discrimination, historical discrimination, or perceived discrimination (for a review, see Small and Pager, 2020). For a philosophical viewpoint, see e.g., Altman (2020). None of these works explicitly discuss PP.

⁷PP can indirectly raise a moral tension for humans when applied to a chemical, say, if authorizing the chemical may alter some people’s health and denying its authorization will make some workers lose their jobs. But this is not a case where PP is directly applied *to human beings*, only to a chemical.

⁸Note that the uncertainty may be located in two places: (1) whether this human being indeed displays the characteristics generating the harm (and it is sure that, if she does, then she will causally lead to some harm), like in the donating blood case; (2) whether someone displaying the characteristics will causally lead to some harm (and it is sure that the person displays the characteristics), like in the ‘S’-labeled people case.

come from the suffering some people will experience from precautionary measures since this is already the case in traditional applications of PP (e.g. if PP bans a chemical product, some workers may lose their jobs).

Why can applying PP to human beings create specific moral problems? To begin with, note that PP is typically not applied to all human beings but only to a subset of humans who share a given characteristic — men having sex with men, people from a specific country, ‘S’-labeled people, Maghrebi people, for instance. In statistics, one would say that PP is applied to human beings who share some modality of a variable, e.g., the modality “yes” for the variable “S-labeled”. The typical justification for PP is that this modality is (strongly suspected to be) correlated with some harm to be avoided, and precautionary measures should apply to people with this modality. The moral objections stem from four reasons, all related to the use of (implicit, weak) correlations about human beings.

First, it may not be morally permissible to single out a particular category of humans. This seems to be the moral objection in the Hiring case: picking out “Maghrebi people” for the measure “don’t hire” is not morally permissible, as it is equated with racial discrimination or racism. This kind of moral objection would not occur with categories of inanimate objects. Second, a correlation, even a strong one, does not always mean causality. That is, the characteristics in question may not be the *cause* of the harm. For instance, in the case of the Covid-19, nationality as such cannot be the cause of being ill or positive for Covid-19, but various social and medical factors probably are. Precautionary measures will target the wrong set of people and will unfairly affect them. Third, the correlation may be a causality, but the people may not be *responsible* for the modality of the variable. A cause of contracting HIV can be to have sex with a category of people for which the prevalence of HIV is high (e.g., gays). But arguably, one is not *responsible* for being sexually attracted by other men and for belonging to the MSM category. So, the moral objection goes, it is morally wrong to target MSM and negatively impact them because they are not responsible for belonging to this category.⁹ Fourth, the correlation may be only weak, which would morally prevent from using it to impose some measure on human beings. This is the kind of objection in the case of ‘S’-labeled people: the presumption of innocence requires that people can only be considered as guilty if proven so, with a very high proof threshold. A mere partial correlation typically does not meet this threshold. Here again, the moral objection triggers because PP is applied to human beings, who come with moral rights.

⁹Lack of responsibility may not in itself be a sufficient criterion for a moral objection. Cf. for instance Boyer-Kassem and Duchêne (2020), who also discuss the relationship between correlation, causality and responsibility.

4 Modeling the decision with an additional ethical utility dimension

The moral tension arising when PP is applied to human beings may be solved in one way or another by sticking to PP or rejecting it in favor of anti-discrimination or civil liberties considerations. Our aim is not to defend some normative position in this respect, and we shall not advance any moral theory in favor or against PP.¹⁰ Our aim, in this last Section, is to *model* and *conceptualize* the decision with which people solve the moral tension. Our thesis is that the lexical utility model of PP (presented in Section 2.1) can be adequately extended to that aim.

4.1 Limits of existing models

Can existing models successfully account for the cases of Section 3? Let us first be explicit on the desiderata.

- (i) The model does not just always prescribes or forbids PP; it can describe both agents who judge that a version of PP applies to, say, the donating blood case, and to those who judge that no version does.
- (ii) The model enables fine-grained judgments on all cases, like changing one's opinion on whether MSM should be authorized to donate blood, depending on the data (e.g., rate of infection among MSM, blood reserve shortage).
- (iii) The model is a good model of PP (when it applies to decisions taken with PP, cf. (i)), i.e., it correctly represents decisions taken in the name of PP.
- (iv) The model applies to human beings (like in Section 3) and also to traditional cases of PP in human health and the environment.
- (v) The model is meaningfully connected to decision theory or moral theory, e.g. it encompasses expected utility theory or some particular moral theory.

Desiderata (i) and (ii) are because we want the model to be descriptive; (ii) is more demanding than (i). (iii) is a minimal theoretical demand because of the centrality of PP. (iv) and (v) are desiderata of generality of the model to avoid *ad hoc* objections.

The desideratum (iii) already limits much the scope of the investigation. Models that focus on the economic implications of these moral tensions, for instance neoclassical models and an expected utility theory framework (Lang and Lehmann, 2012; Lang and Spitzer, 2020), are not tuned to model PP. Among existing models of

¹⁰Our own views are that the moral tension may not be solved in a general, uniform way, but rather on a case-by-case basis, sometimes precautionarily, sometimes not. In addition, the two authors do not have the same moral or even definite intuitions in all cases. But all this does not matter in any way for the thesis we defend in this Section and in the paper.

PP, Peterson’s (2006) model suffers from some inadequacies; Steel’s (2015) model is a very good one, but some limitations are pointed by Bartha and DesRoches (2021), who have proposed a new model that improves on it (cf. Section 2.1). So Bartha and DesRoches’ model, presented in Section 2.1, looks like the right starting point.¹¹ In addition, their model already fulfills (v), since expected utility theory comes as a special case, when one judges that no outcome qualifies as a catastrophe. It partly fulfills (i), as agents are free to consider some outcomes or none as catastrophes, which makes them follow or not follow PP. But what if an agent considers that discriminating MSM is a catastrophe and that contaminating receivers is also a greater or lesser catastrophe? Then, Bartha and DesRoches’ model is of little help because its representation of catastrophes is binary¹². It also has some difficulties with (ii), as it cannot represent a continuous tradeoff between these two kinds of catastrophes. The key lies in extending Bartha and DesRoches’ model by enabling the representation of two kinds of catastrophes (e.g. one of health, and one of ethical nature). This is detailed in the next Section.

4.2 Our model on the donating blood case

Let us consider how Bartha & DesRoches’ lexical utility model can be used in the blood donating case. There are two states of the world: some potential MSM donors are infected with HIV without being detected (I), or this is not the case ($\neg I$). Two actions are considered: authorize MSM donating blood (A), or do not ($\neg A$). Recall that Bartha & DesRoches’ lexical utility model represents the utility of an outcome O as a two-dimensional vector $(u_1(O), u_2(O))$, the first component indicating whether the outcome is catastrophic or not (resp. -1 or 0), and the second component representing the traditional utility. Here, the catastrophe is an HIV infection of a blood receiver, which occurs only if action A is made in state I . $u_2(O)$ is typically the health benefit reaped from a blood donation.¹³

Assume first that no discrimination issues are considered. This gives Table 2. As soon as the probability of the state I is not null, maximizing the expected lexical utility requires considering the first dimension only, and $\neg A$ is the recommended

¹¹Because of (v), one may instead want to start from some moral theory or traditional decision theory. Still, then one should incorporate Bartha and DesRoches’ model in one way or another (because of (iii) and (iv)). So it is equivalent to starting from their model.

¹²Steel and Bartha (2022), who extend Bartha and DesRoches’ model to tradeoff issues, note that “Catastrophes that are categorically worse than other catastrophes could also be modeled within a lexical utility framework by utility vectors with three or more components. However, we do not introduce such models here because formulations of PP typically focus on a single level of catastrophe.” (footnote 5 p. 4). Our present contribution can be seen as developing this modelization by focusing on overlooked cases of PP which precisely require distinguishing between several levels of catastrophe.

¹³Some refinements might be considered, like the number of months of abstinence or the percentage of HIV infections, but the structure would remain similar.

action:¹⁴ PP recommends excluding MSM from donating blood.

action	state	
	I	$\neg I$
A	$(-1, u)$	$(0, v)$
$\neg A$	$(0, w)$	$(0, x)$

Table 2: Decision table for the blood donating case in Bartha and DesRoches' lexical utility model, without discrimination considerations.

Now, add discrimination considerations. We propose to extend Bartha and DesRoches' model by adding an ethical utility dimension u_e , which represents the catastrophe of discriminating against people because of their sexual orientation: $u_e(\neg A) = -1$ and $u_e(A) = 0$. This catastrophe is distinct, and of a different nature from the catastrophe of transmitting HIV, so this *a priori* justifies defining a distinct utility dimension. One may have different views on the relative importance or comparability of the two catastrophes. Let us consider three people, Anna, Bob and Carmen. Anna thinks that HIV transmission is the worse catastrophe, Bob believes that the discrimination of MSM is the worse catastrophe, and Carmen thinks that the two are actually comparable. In the model, Anna's utility will be represented by inserting the ethical dimension u_e between u_1 and u_2 ; Bob's utility will have u_e before u_1 and u_2 ; Carmen's utility will merge u_1 and u_e . Table 3 represents their respective decision tables. Recall that, in this lexical utility model, placing a utility dimension before another means that it will have full preeminence in the decision-making: to maximize expected lexical utility, one first maximizes the expected utility's first component and, only in case of draws between two actions, one turns to the expected utility's second component, and so on.

Assuming that the probability of neither state is null, the rational decision is reached by considering the first dimension only: $\neg A$ for Anna, A for Bob, and A for Carmen. That is, Anna does not authorize MSM to donate blood, while Bob and Carmen do. Anna's and Bob's decisions within the model are what could be expected from their description — Anna considers HIV infections to be a greater kind of catastrophe than discrimination, and Bob thinks the reverse. For Carmen, things are a bit more complicated, and the model needs further improvements. Carmen thinks that the two catastrophes (infection and discrimination) are comparable, so they are represented in the same dimension. But comparability does not mean equality: she might think that one is twice worse than the other. Besides, more information is needed on the *number* of additional infections to compare it with discrimination (without this information, a comparison in the same dimension is meaningless).

¹⁴For the first dimension, noting $p(I)$ the probability of an infection, the expected utility of action A is $-1 \times p(I) + 0 \times (1 - p(I)) = -p(I)$. The expected utility of action $\neg A$ is $0 \times p(I) + 0 \times (1 - p(I)) = 0$. So, whatever $p(I) \neq 0$, $\neg A$ has the higher expected utility.

Anna			Bob		
action	state		action	state	
	I	$\neg I$		I	$\neg I$
A	$(-1, \mathbf{0}, u)$	$(0, \mathbf{0}, v)$	A	$(\mathbf{0}, -1, u)$	$(\mathbf{0}, 0, v)$
$\neg A$	$(0, -\mathbf{1}, w)$	$(0, -\mathbf{1}, x)$	$\neg A$	$(-\mathbf{1}, 0, w)$	$(-\mathbf{1}, 0, x)$

Carmen		
action	state	
	I	$\neg I$
A	$(-1+\mathbf{0}, u)$	$(0+\mathbf{0}, v)$
$\neg A$	$(0-\mathbf{1}, w)$	$(0-\mathbf{1}, x)$

Table 3: Decision table for the blood donating case, now with discrimination considerations (in bold). Anna, Bob, and Carmen have different views on the catastrophes hierarchy and place the additional ethical utility dimension accordingly.

So, instead of using -1 to represent an ethical catastrophe, one should use a real number, as suggested by Steel and Bartha (2022). Suppose Carmen actually judges a discrimination catastrophe as of utility $-\alpha$ instead of -1 (in the two cells on the line of action $\neg A$). Then the expected utility of action $\neg A$ will be $-\alpha$. The expected utility of action A will still be $-p(I)$. Depending on the relative values of α and $p(I)$, Carmen will select A or $\neg A$. In other words, someone who considers the two catastrophes comparable will choose to avoid one catastrophe or the other, depending on their specific characteristics, so the modeling makes full sense. The French government decisions in Section 3.1, which started with an interdiction that MSM donate blood followed by a progressive authorization, may be successively modeled by Anna, then Carmen, and perhaps Bob.

4.3 General model

Beyond this example, our general proposal is to model the ethical tension in the applying PP to human beings through a lexical utility model. Like in Bartha and DesRoches' (2021) model, standard utility is adjoined a dimension for catastrophes (e.g., in health, environment, or security).¹⁵ Then, beyond Bartha and DesRoches' (2021) model, we allow for the introduction of a third utility dimension,¹⁶

¹⁵As mentioned in Section 2.1, whether an event qualifies as a catastrophe is here not theoretically postulated but is left to the agent's judgment and preference.

¹⁶The idea of differentiating between catastrophes is expressed by Steel and Bartha (2022, p. 4): "PP is compatible with the notion that some catastrophes are worse than others," which leads to "utility vectors with three or more components". This was previously briefly noted by Bartha and DesRoches (2021, p. 8720): "The lexical approach can be broadened to handle multiple harm thresholds, but we are unable to consider cases of this type in this article."

nb of dimensions	3	3	2	2	2	1
utility vector	(u_1, u_e, u_2)	(u_e, u_1, u_2)	$(u_e + u_1, u_2)$	$(u_1, u_e + u_2)$	$(u_e, u_1 + u_2)$	$(u_e + u_1 + u_2)$
example	Anna	Bob	Carmen	Djawad	Élisa	Fuyumi

Table 4: Range of the possible models in our proposal. u_2 represents traditional utility, u_1 represents the catastrophic part of the utility as in Bartha and DesRoches (2021), u_e represents the ethical catastrophic part that we introduce (discrimination and presumption of innocence considerations).

which represents the possibility of an *ethical* catastrophe, for instance, because of discrimination or violation of civil liberties. This additional utility dimension echoes the common origin identified in Section 3.5, namely the moral rights of the human beings to which PP is applied. Further, we allow catastrophes (either standard or ethical) to be represented with real numbers, not just 0 and -1 , as in Bartha and DesRoches’ (2021) model. In standard expected utility theory, utilities are provided by agents; likewise, here, the relative position (or existence) of the ethical utility dimension with respect to other dimensions is provided by agents. We do not defend here any normative claim that, for instance, the ethical dimension should always be the first or always be the second dimension. Then, given the description of agents’ utilities, the rational decision is to maximize lexical expected utility, which means computing the expected utility using the first utility dimension and, in case of draws between actions, considering the next utility dimension, and so on. Overall, the general picture we propose is a multi-dimensional utility model, with the number of dimensions, between 1 and 3, given by the agent.¹⁷

Table 4 defines three other fictive agents, Djawad, Élisa, and Fuyumi. Djawad considers that the ethical considerations do not reach the “catastrophe” threshold and are comparable with traditional events. Élisa thinks only ethical considerations deserve the label “catastrophe”. Fuyumi considers that all consequences are comparable, which amounts to traditional expected utility theory (with ethical considerations taken into account). According to the traditional view in which PP and expected utility theory do not overlap, she cannot be said to follow PP. Depending on the specific number of lives saved or lost, she will decide in favor or against authorizing MSM to donate blood.¹⁸ As in Bartha and DesRoches’ model, the present model is an extension of the standard expected utility theory and encompasses it.

Besides, recall that there is not just *one* PP, but several versions of it depending on the levels of harm or uncertainty which trigger PP or on the recommended measures (cf. Section 2). Fortunately, our model is compatible with these nuances

¹⁷In case there are several ethical catastrophes or several standard catastrophes one could even consider more than 3 dimensions, for instance, with alternating kinds of catastrophes.

¹⁸To come back to the desiderata introduced in Section 4.1, this illustrates that (ii) is fulfilled.

and can account for the fact that one follows some versions of PP and not others. For instance, suppose someone subscribes to PP only in case the probability of additional HIV infections p is higher than some q . Then, if $p < q$, this person will adopt Fuyumi’s model and not follow PP, while if $p > q$, she will adopt Carmen’s model and follow PP.

Note that the lexical utility model may differ depending on the aggregate level at which the problem is considered. Take the hiring case, but assume now that the boss does not make his discrimination explicit in the advertisement — he just does not hire Maghrebi people. The harm made once to one Maghrebi person by not being hired can be somewhat moderate and is probable only — she may assume that she was not hired for another reason than her ethnicity. But suppose the same decision is repeated by every hirer against all Maghrebi people, at every place in the country. In that case, no doubt is permitted that Maghrebi people are intentionally not hired, and major social tensions may result. Back to the model, this means that the “assault” catastrophic dimension may be placed in front of the “discrimination” catastrophic dimension for a single case, but it may not be so at the level of a country.¹⁹ Individual and governmental decisions are typically evaluated with different moral standards.

4.4 Objection and reply

Consider the following objection: your model is not very specific and has many free variables (utility values, number of utility dimensions); so it is no real surprise that it can describe many diverse cases. What is your model’s added value? As it is not normative, concretely, what do we learn with it?

Our answer is multiple. First, the model is partly normative, in the same way as traditional expected utility theory is. Once utility values (plus here, utility dimensions) are given, the model says which decision is right. Second, the value of our model is to focus on the right moral questions. Instead of “Can PP be applied to terrorists? or to gays?”, the model asks: “Are these catastrophes commensurable? Should one be avoided as a priority?”. To reach a decision, the key is to organize catastrophes (or outcomes) into a possible hierarchy and, when they are at the same level, to give their magnitudes. Third, our model sheds new theoretical light on the role of PP in decisions in two ways. (i) Note that there is no equivalence between not authorizing MSM to donate blood and applying PP to MSM. Anna, and potentially Carmen and Fuyumi (for suitable parameter values), do not authorize MSM, but Anna and Carmen always apply PP, while Fuyumi never does.²⁰ Bob, Carmen, and Fuyumi (the last two for suitable parameter values) do authorize MSM, but

¹⁹For a discussion of utility aggregation problems, cf. Bartha and DesRoches (2022), p. 8732-8734.

²⁰It is important here to distinguish between the decision rule one uses and the decision one takes. Two people may make the same decision by following different decision rules.

Carmen applies PP to MSM, while Bob and Fuyumi never do. (ii) There is an unseen formal unity of decisions. As Anna does not authorize MSM to donate blood because of the risk they could transmit HIV, it is intuitive to say that she applies PP to MSM (as we have just done above). Conversely, it is intuitive to say that Bob, who advances anti-discrimination reasons, does not apply PP to MSM. But formally speaking, the decisions of both Anna and Bob are modeled with a three-dimensional lexical utility model. Anti-discrimination considerations, and (precautionary) considerations regarding additional HIV infections, can be modeled within the same framework. The case of the ‘S’-labeled people displays another parallel: the presumption of innocence, which is the objection Urvoas raised to Sarkozy’s proposal to apply PP to ‘S’-labeled people, can be considered as a reverse PP: it requires not to declare someone guilty (that is, to take a precautionary measure) even if there is a low probability of not being guilty (the knowledge threshold), because it could otherwise be unfair to an innocent (the harm to be avoided). So, the intuitive speaking of “PP *vs* anti-discrimination” or “PP *vs* presumption of innocence” can be misleading in suggesting that the conflict happens between radically different principles, whereas there are important formal similarities between them, and that they can be subsumed within a common framework. Instead, the moral tension should be seen as a competition between two kinds of catastrophes to be avoided, which can be placed in different hierarchies.

5 Conclusion

We have presented several examples of applications of PP that seem to raise new moral tensions or dilemmas (Section 3). The distinctive feature of these cases is that PP is applied to human beings (and not, say, to molecules), who come with moral rights (Section 3.5). We proposed to model the cases by extending an existing lexical utility model (Section 4). Our model distinguishes between ordinary events and catastrophes of different kinds, in particular ethical catastrophes. It associates them with distinct utility dimensions. Then, it enables to say that people have diverging moral views about the application of PP because they rank the various utility dimensions (about events and catastrophes) in different orders, or do not consider the same number of utility dimensions, or within a specific dimension do not evaluate catastrophes with the same magnitudes. In other words, they have diverging views on which harms should first be avoided. This model of PP is applicable in situations where outcomes receive precise probabilities, i.e., to the risk domain, in addition to the traditional case where outcomes cannot receive probabilities. Our proposal should help people who face ethical dilemmas when applying PP to clarify the problem and focus on the right questions: Is PP being applied to human beings? Can I identify possible catastrophes, i.e., incomparably bad events? If I identify several kinds of catastrophes (e.g., deaths, discriminations), how do I rank them? The question is not “Should PP prevail in this context?” anymore, which suggests

that many public debates in Section 3 are ill-posed. The point is not whether PP should apply to gays or to Maghrebi people.

Regarding the scope of application of PP, which is traditionally human health and the environment, we have neither defended an extension to security issues nor excluded it. It is true that, because the domain of security often involves applying PP to human beings, the kind of dilemmas discussed here are bound to arise. As suggested, their resolution will require moral inputs from agents about relevant catastrophes, probably on a case-by-case basis. So the question of the scope of PP, and for instance, its application to security issues, should not be expected to be solved by the same arguments that worked for its application to human health or the environment, like for instance Steel's (2015, chap. 4) historical argument in favor of precaution.

Acknowledgments

We thank Paul Bartha, Tyler DesRoches, Daniel Steel and participants at the workshop "Why be Cautious? On the Justification of Precautionary Principles" (Bern, 2017) and at the Paris Seminar for Social and Formal Epistemology (2023) for valuable comments.

References

- 20 Minutes. 2013. Discrimination à l'embauche: Un patron juif condamné à 5.000 euros d'amende. Retrieved March 27, 2023. <https://www.20minutes.fr/societe/1177335-20130620-20130620-discrimination-a-embauche-patron-juif-condamne-a-5000-euros-amende>
- Ahtensuu M. and P. Sandin. 2012. The Precautionary Principle. In Roeser et al (ed), *Handbook of Risk Theory*, Springer.
- Altman A. 2020. Discrimination. *The Stanford Encyclopedia of Philosophy* (Winter 2020 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/win2020/entries/discrimination/>>.
- Arcidiacono P. and M. Lovenheim. 2016. Affirmative action and the quality-fit trade-off. *Journal of Economic Literature*, 54(1), 3-51.
- Ashford N. et al. 1998. The Wingspread Statement on the Precautionary Principle. Wingspread, Racine, 1998. Retrieved March 27, 2023. URL = <http://www.gdrc.org/u-gov/precaution-3.html>
- Bartha, P. and C. T. DesRoches .2021. Modeling the precautionary principle with lexical utilities. *Synthese* 199, p. 8701–8740.
- Boyer-Kassem, T. 2017a. Is the Precautionary Principle really incoherent?. *Risk Analysis* 37(11): 2026–2034.

- Boyer-Kassem, T. 2017b. The Precautionary Principle has not been shown to be incoherent: A reply to Peterson. *Risk Analysis* 37(11): 2039–2040.
- Boyer-Kassem T. and S. Duchêne. 2020. On discrimination in health insurance. *Social Choice and Welfare* 55: 5–23.
- Cheyne I. 2007. “Gateways to the Precautionary Principle in WTO Law”, *Journal of Environmental Law* 19(2):155-172.
- Christiansen, A. 2019. “Rationality, Expected Utility Theory and the Precautionary Principle”, *Ethics, Policy & Environment* 22(1), p. 3–20.
- Euractive. 2020. Baltics open Europe’s first pandemic ‘travel bubble’ as curbs ease. Retrieved March 27, 2023. <https://www.euractiv.com/section/justice-home-affairs/news/baltics-open-europes-first-pandemic-travel-bubble-as-curbs-ease/>
- Europe 1. 2016. Fichés S : Sarkozy veut appliquer le ‘principe de précaution’ au terrorisme. Retrieved March 27, 2023. <https://www.europe1.fr/politique/internement-de-fiches-s-sarkozy-veut-appliquer-le-principe-de-precaution-au-terrorisme-2847931>
- European Commission. 2000. Communication from the Commission on the Precautionary Principle (The Commission of the European Communities, Brussels, 2 February 2000).
- European Commission. 2020. Towards a phased and coordinated approach for restoring freedom of movement and lifting internal border controls: COVID-19. 2020/C 169/03, C/2020/3250, OJ C 169, 15 May 2020, pp 30-37.
- de Kort et al. 2016. Blood donor selection in European Union directives: room for improvement. *Blood Transfus* 14: 101-8.
- Gollier, C., B. Jullien and N. Treich. 2000. Scientific progress and irreversibility: an economic interpretation of the ‘Precautionary Principle’. *Journal of Public Economics* 75, 229–253.
- Goldner Lang, I. 2021. ‘Laws of Fear’ in the EU: The Precautionary Principle and Public Health Restrictions to Free Movement of Persons in the Time of COVID-19. *European Journal of Risk Regulation*, 1-24.
- Gollier, C., B. Jullien, and N. Treich . 2000. Scientific progress and irreversibility: an economic interpretation of the ‘precautionary principle’. *Journal of Public Economics* 75 (2), 229–253.
- Harris, J. and S. Holm. 2002. Extending Human Lifespan and the Precautionary Paradox. *Journal of Medicine and Philosophy* 27(3):355–368.
- Hansson, S. O. 2013. *The Ethics of Risk*. Palgrave Macmillan.
- Jensen, K. K. (2002), “The Moral Foundation of the Precautionary Principle”, *Journal of Agricultural and Environmental Ethics* 15: 39-55.
- Joint European Roadmap towards lifting COVID-19 containment measures. 2020/C 126/01, C/2020/2419, OJ C 126, 17 April 2020, pp 1–11. [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020XC0417\(06\)](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020XC0417(06))
- Lang, K. and J. Y. K. Lehmann. 2012. Racial discrimination in the labor market:

- Theory and empirics. *Journal of Economic Literature*, 50(4), 959-1006.
- Lang, K. and A. K. L. Spitzer. 2020. Race discrimination: An economic perspective. *Journal of Economic Perspectives*, 34(2), 68-89.
- Le Monde. 2016. Jean-Jacques Urvoas : ‘Ces ‘arguties juridiques’ qui sont notre force’. Retrieved March 27, 2023. https://www.lemonde.fr/idees/article/2016/07/28/jean-jacques-urvoas-ces-arguties-juridiques-qui-son-t-notre-force_4975616_3232.html
- Le Point. 2013. Le douteux ‘principe de précaution’ d’un petit patron juif. Retrieved March 27, 2023. https://www.lepoint.fr/societe/le-douteux-principe-de-precaution-d-un-petit-patron-juif-20-06-2013-1684046_23.php
- Munthe, C. 2011. *The Price of Precaution and the Ethics of Risk*. Springer.
- Munthe, C. 2020. Precautionary Principle. In H. LaFollette (ed.), *International Encyclopedia of Ethics*, Chichester: Wiley.
- Peterson, M. 2006. The precautionary principle is incoherent. *Risk Analysis* 26(3):595–601.
- Peterson, M. 2017. Yes, The Precautionary Principle Is Incoherent. *Risk Analysis* 37: 2035–2038.
- Resnik, D. B. (2021), *Precautionary Reasoning in Environmental and Public Health Policy*, Cham (Switzerland): Springer.
- Sandin, P. (2006), “A Paradox Out of Context: Harris and Holm on the Precautionary Principle”, *Cambridge Quarterly of Healthcare Ethics*, 15: 175–183.
- Sandin, P., S. O. Hansson, C. Rudén and A. Juthe (2002), “Five charges against the precautionary principle”, *Journal of Risk Research* 5 (4), 287–299.
- Small, M. L., and D. Pager, (2020), “Sociological perspectives on racial discrimination”, *Journal of Economic Perspectives*, 34(2), 49-67.
- Steel D. (2015), *Philosophy and the Precautionary Principle*, Cambridge: Cambridge University Press.
- Steel D. and P. Bartha (2022), “Trade-offs and the precautionary principle: A lexicographic utility approach”, *Risk Analysis*, <https://doi.org/10.1111/risa.13892>
- Stirling A. (2009), “The Precautionary Principle”, in Olsen et al (eds.), *Blackwell Companion to the Philosophy of Technology*, Oxford: Blackwell.
- Stone T. W. (2003), “Margin of Appreciation Gone Awry: The European Court of Human Rights’ Implicit Use of the Precautionary Principle in *Fretté v. France* to Backtrack on Protection from Discrimination on the Basis of Sexual Orientation”, *Connecticut Public Interest Law Journal* 3(1):218-236.
- Sunstein C. R. (2005), *Laws of Fear. Beyond the Precautionary Principle.*, Cambridge: Cambridge University Press.
- Véran O. (2013), “La filière du sang en France”, report for the Prime Minister, France. https://solidarites-sante.gouv.fr/IMG/pdf/Rapport_Veran_filiere-sang.pdf