Facing rare and catastrophic disasters : Four essays on the economics of nuclear safety regulation

Romain Bizet

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à MINES ParisTech

Facing Rare and Catastrophic Disasters: Four essays on the economics of nuclear safety regulation

École doctorale n°396

ÉCONOMIE, ORGANISATION, SOCIÉTÉ

Spécialité ÉCONOMIE ET FINANCE

COMPOSITION DU JURY :

Mme Estelle CANTILLON
Université Libre de Bruxelles
Présidente du jury

M. Pierre PICARD
École polytechnique
Rapporteur

M. Jean-Marc TALLON
Paris School of Economics
Rapporteur

M. Geoffrey ROTHWELL
OECD Nuclear Energy Agency
Examinateur

M. Ulrich WAGNER
Mannheim University
Examinateur

M. François LÉVÊQUE
Professeur - MINES ParisTech
Examinateur

Soutenue par **Romain BIZET**
le 15 Novembre 2017

Dirigée par **François LÉVÊQUE**

MINES ParisTech | PSL RESEARCH UNIVERSITY PARIS

"S'il est théoriquement possible, du point de vue de la théorie des échanges de chaleur, qu'en mettant un poulet dans un four chaud pendant une heure, celui-ci en ressorte congelé, la probabilité de cet évènement est suffisamment faible pour qu'aucun cuisinier ne tienne compte de cette possibilité."

- Émile Borel, *Traité du calcul des probabilités et de ses applications (1921-1939)*

Acknowledgements

First and foremost, I would like to express my gratitude to François Lévêque, for trusting me with this challenging research topic, and for providing me with a dedicated and pedagogical supervision. I have tried to learn a lot during these past three years from his ability to grasp accurate senses of complex economic mechanisms, and to turn them into clear and simple intuitions.

I would also like to thank Pierre Fleckinger, whose supervision was key to the realization of the fifth chapter of this thesis. His Google-scholar-like knowledge of the literature and his ability to design *cute* economic models were a constant example to look up to.

Finally, I want to express my gratitude to Petyo Bonev, whose careful and enthusiast supervision was instrumental to the realization and quality of the empirical chapters of this thesis.

Next, I would like to thank professors Jean-Marc Tallon, David Martimort and Jean-François Laslier for their courses on decision theory, mechanism design and social choice. I believe that learning from them shaped my journey in the field of economics, and somehow influenced the development of this thesis. I would also like to thank professor Pierre Picard for his time and valuable advice in the first step of this thesis.

I also thank the faculties from CERNA for the atmosphere they developed in our lab, and from which I benefited a lot during the past three years. In particular, I thank Matthieu Glachant for his always direct and high-quality comments on several parts of this thesis, and Petyo Bonev and Margaret Kyle for providing us with high-quality seminars and doctoral workshops. I would like to thank Sesaria Ferreira, CERNA's administrative officer, whose multiple abilities - ranging from bureaucratic problem-solving to the transformation of austere lunch breaks into a lively gathering of souls - were among the greatest external contributions to this thesis.

This thesis would not have been possible without the support I obtained from the Autorité de Sûreté Nucléaire (ASN) and Institut de Radioprotection et de Sûreté Nucléaire (IRSN). In particular, I would like to thank the president and former director of ASN, Pierre-Franck Chevet and Jean-Christophe Niel, as well as the former president of IRSN Jacques Repussard for giving us access to the data with which the empirical part of this thesis. I would also like to thank Hervé Bodineau, Vincent Crutel and Antoine Pellarin from IRSN; Julien Collet and Éric Gaucher from ASN; and Antoine Assice, Laurent Kurkowski and Dominique Minière from EDF, whose expertise, knowledge of the inner workings of the French nuclear regulation, and support in the analysis of this dataset were crucial in the design of my two empirical chapters. I would also like to thank the project managers from all French Commissions for Local Information (CLIs), as well as the members of ANCCLI, for their time and for providing me with the additional data that led to the fourth chapter of this thesis.

I would now like to thank my fellow Ph.D. students and friends from CERNA. I thank Lina, Marine, Damien, Charlène, Jiekai, Carlotta, Btissam, Laurie, Dandan, Maddalena, Sabrine, Guillaume, Victor, Connie and Ekaterina, for making these last three years a very challenging and gratifying intellectual adventure. I would also like to thank the cheerful Ph.D students from the French Association of Energy Economists: Antoine, Clara, Cyril, Déborah, Martin, Nicolas, Simon, Victor, and Xavier. A final special thanks goes to Philippe and Paul-Hervé, for the time suffered together in improvised econometrics lessons, for the long hours spent looking for incongruous interpretations of uncanny equations, and for the moments enjoyed discussing together every other newsworthy event under the light of our favourite economics textbooks.

Last but not least, I would like to thank my parents, brothers, family and close friends, whose unwavering support has always been my deepest source of determination.

Foreword

This thesis is embedded in a research program called *The new economics of nuclear energy* directed by François Lévêque and financially supported by EDF. This thesis is composed of four main chapters that tackle the uncertainties that characterize the risks associated with the use of nuclear power. The results presented in this document do not necessarily reflect the views of EDF.

The first part of this thesis deals with the measurement of nuclear safety when observations of accidents are extremely rare. The second part of this thesis tackles the implementation of safety regulation and disaster mitigation strategies. In both parts, I use both empirical and theoretical methods taken from the economics literature to develop new tools for the analysis of nuclear safety, to derive new results regarding its evolutions, and to propose policy guidelines for further improvements of nuclear safety regulation.

I first propose an introductory chapter in which I review the economics literature dedicated to safety regulation, show that many existing contributions can be used to analyse the particular case of nuclear safety, and yet, that these contributions fail to account for two specific aspects of nuclear safety: the rare and catastrophic nature of nuclear power accidents. Based on this observation, I review recent theoretical results which could be used to account for these characteristics, and explain how this thesis aims to contribute to this literature.

Two chapters of this thesis provide theoretical analyses of two questions related to nuclear safety: the estimation of the cost of nuclear accidents, and the design of credible communication strategies in the wake of major nuclear disasters. Though, the questions tackled in these chapters are broad, and arise more generally when facing catastrophic risks or risks whose objective probabilities of occurrence are not well-defined. I believe these chapters should not be regarded as purely specific to nuclear power.

Chapter 2 tackles model uncertainty and the social cost of nuclear power accidents. This chapter is joint work with F. Lévêque and is based on a report on the cost of nuclear accidents made for the European Investment Bank, and on a communication prepared for the 2015 International Workshop on Nuclear Safety held in U.C. Berkeley in 2015 (see e.g. Bizet and Lévêque (2017)). This chapter was presented at the 2016 IAEE international conference in Bergen, at the 2017 SBCA conference in George Washington University, and in dedicated seminars in 2016 in Mines ParisTech, the universities of Waseda, Tokyo and Kyoto, the Beijing Institute of Technology, the North China Electric and Power University, and in Tianjin and Beihang Universities.

Chapter 5 provides a game-theoretic analysis of crisis communication strategies. This chapter is joint work with P. Fleckinger. In this chapter, we propose a cheap-talk model to study the communication of disastrous events, focusing on the cases in which the actions individuals would like to take to protect themselves differ from those the government would like them to take. Avoiding panic or destroying contaminated food stocks to prevent exposure to radiations are examples of such cases. We characterize features of credible communication strategies, and study their coordination with preparedness and mitigation actions. This rather early work has not been presented in conferences so far.

Chapters 3 and 4 are more specific to nuclear power. They both propose empirical analyses of nuclear safety in France, based on a new dataset containing all significant safety events declared in the French fleet. In both chapters, I try to go beyond the safety and reliability analyses that these events have to offer, and to identify interesting economic questions related to the occurrences, detection and reporting of these events.

Chapter 3 estimates the variations of safety in the French fleet over time and across reactors of different ages and technologies, while accounting for the fact that safety events can fail to be detected or reported. This chapter was supervised by F. Lévêque and P. Bonev, and was presented at the 2016 YEEE seminar in the University of Edimburg, and in a dedicated seminar in Kyoto University in 2016. It is currently under revision for publication in *Nature Energy*.

Chapter 4 estimates the causal effect of a French local monitoring policy on the behaviour of nuclear plant managers. To do so, we gathered data describing the intensity of the monitoring performed by local public agencies and measured how this monitoring affected reporting behaviours. This paper is joint work with P. Bonev, and was presented at the 2017 ZEW Energy Conference, the 2017 TSE-IDEI Energy Economics Conference, and the 2017 EAERE, ELAEE and IAEE annual conferences.

To conclude this foreword on a more personal note, I hope that through these chapters, one may find tools to better apprehend the various issues that characterize nuclear safety: from the problems associated with its definition and measurement, to the implementation of effective safety regulation and disaster mitigation policies.

Contents

List of Figures

List of Tables

Résumé du chapitre 1

Dans ce chapitre introductif, je propose une revue de la littérature économique dédiée à la réglementation de la sûreté, en l'illustrant par des exemples de pratiques réglementaires employées dans l'encadrement de l'industrie nucléaire, comme par exemple l'utilisation combinée de standards de sûreté et de règles de responsabilité stricte pour les opérateurs. Ce chapitre décrit ensuite les incertitudes qui existent quant aux probabilités et aux dommages associés à ces accidents, ainsi que les biais qui caractérisent les perceptions de ce risque chez les individus, experts ou non. Ces spécificités du risque nucléaire remettent en question la pertinence des résultats de la littérature économique, basés sur l'hypothèse que les régulateurs et les firmes régulées ont connaissance, au moins de façon subjective, des risques auxquels ils sont exposés. Enfin, cette introduction présente des éléments récents de théorie économique permettant d'éclairer la question de la réglementation des risques dans un environnement incertain, et présente les questions auxquels les quatre chapitres principaux de la thèse répondent.

CHAPTER 1

Introduction

Recent studies dedicated to the economics of nuclear power have paid particular attention to the overnight cost of construction of nuclear power stations (see e.g. Du and Parsons (2009); MIT (2009); Grubler (2010); Boccard (2014); Berthélemy and Rangel (2015); Rangel and Lévêque (2015); Rothwell (2016) or Lovering et al. (2016))[1]. This recent rise of interest for this rather marginal source of production of electricity (11% of the world consumption of electricity in 2016)[2] has been motivated by a combination of factors: e.g. the fall of gas prices in the United-States and the Fukushima-Daiichi nuclear accidents on March 11, 2011 (Davis, 2012). The rise in concerns over nuclear safety and the enhanced competitiveness of other baseload technologies led economists to question the evolution of the competitiveness of nuclear power, largely driven by the large overnight construction costs of nuclear power reactors.

Nevertheless, as of the end of the year 2016,[3] the world counted 451 operating nuclear power stations. Today, the average age of these reactors is 29.6 years,[4] and 61 nuclear plants are still under construction. So even if the rise of the overnight cost of construction of nuclear reactors was to drive them out of the market for new baseload capacities in a near future, ensuring appropriate safety levels within existing nuclear stations will remain a concern until these plants are decommissioned. Under the hypothesis that new builds will be operated during 60 to 80 years, this may not occur before the end of the twenty-first century.

Moreover, the stakes associated with the safety of nuclear power plants are substantial,

[1] Multiple institutional reports have also been dedicated to this leading issue: see e.g. D'Haeseleer (2013) or Ecofys (2014) for the European Commission, a report from the French court of auditors (Cour des Comptes, 2014), a technological roadmap from the International Energy Agency (IEA/NEA, 2015), or a recent report from the U.S. Energy Information Administration (EIA, 2016).

[2] Source: The World Nuclear Association Information Library and the International Atomic Energy Agency PRIS database.

[3] Source: The International Atomic Energy Agency PRIS database.

[4] The median and mode of the age distribution of existing reactors are respectively 32 and 33 years.

and uncorrelated with the amount of electricity nuclear power stations produce annually, as some nuclear plants are located in the vicinity of densely populated areas. According to Butler (2011), 152 nuclear power stations have more than 1 million people living within a 75-kilometre radius. To illustrate the potential damage at stake, von Hippel and Schoeppner (2016) depict how a realistic Fukushima-like nuclear accident could unfold in the United-States, and lead to the evacuation of New-York City. An additional example - meeting lower academic standards - is the estimation by former Japanese prime-minister Naoto Kan that the Fukushima-Daiichi catastrophe could have led to the evacuation of up to 50 million people, had the wind blown in the direction of Tokyo.[5]

One of the aims of this introductory chapter is to show that despite the extent of the damage at stake, few studies in the economics literature have paid attention to the interplay between the specificities of nuclear accidents and the incentive schemes used to induce safety in nuclear power stations. In particular, I argue that the catastrophic nature and low frequencies of nuclear accidents are not accounted for in the economics literature dedicated to safety regulation. Hence, how these two characteristics affect the prescriptions of the economic theory of safety regulation, and how the specific mechanisms used by nuclear regulators empirically affect nuclear safety are two questions that remain, so far, largely unanswered.

This thesis contributes to filling in this gap by considering first how the rare and catastrophic nature of nuclear accidents affect the measurement of safety. We argue that the rare frequency of these accident precludes the assessment of objective measures of the risks of incurring harm, and show that alternative measures of safety can be developed by considering minor but significant safety incidents. Second, this thesis also aims to question how the implementation of safety regulation and accident management strategies is affected by the rare and catastrophic nature of nuclear accidents. We do so by studying empirically a specific French policy dedicated to the monitoring of nuclear plant managers by local commissions, and by providing a game theoretic analysis of disaster communication strategies.

The following of this introductory chapter is organized as follows. First, I present some key institutional, technical, and economic features of nuclear safety and of its regulation. In particular, I argue that the regulatory practices observed in the nuclear industry are well grounded in the literature on the economics and governance of safety regulation. I review this literature and illustrate its multiple results regarding the use of liability rules and safety standards, or its results regarding safety governance, through specific examples taken from nuclear safety

[5] Source: The Japan Times (2013).

regulation.

Second, I focus on the specificities of nuclear accidents. I review the literature dedicated to the assessment of the risk associated with nuclear power, i.e. the probabilities and damage associated with major nuclear accidents. I argue that the uncertainties that characterize these accidents prevent the empirical measurement of the effect of safety regulation on the level of nuclear safety, that these uncertainties lead to the coexistence multiple risk perceptions among society, and that these uncertainties challenge some hypotheses that underlie the traditional results of the economics of safety regulation. I review some recent theoretical contributions to the economics literature that aim to address similar questions.

Finally, I state in more details the research questions that motivate this thesis, and present the specific contributions made in the different chapters of this dissertation.

1.1 The economics and governance of nuclear safety regulation

1.1.1 An introduction to nuclear safety regulation

National and international regulatory set-ups

As of July 2017, 31 countries operate 446 nuclear power stations to sustain their electrical consumption, for a total installed capacity of 390 electrical gigawatt (GWe), and an annual electrical production of 2,500 terawatt.hours (TWh).[6] Nuclear safety regulation is a national responsibility, which is organized differently in every state, but usually entails two main features: strict liability rules for nuclear plant operators, and mandatory safety standards enforced by a monitoring agency.

First, plant operators in all countries relying on nuclear energy are held strictly liable for the damage caused by nuclear accidents, up to a certain amount and up to a certain point in time. These limits are set by several international treaties,[7] which are implemented in several ways depending on local organizations of the nuclear industry.

For instance, in Western Europe, the 2004 revised Paris protocol holds nuclear plant operators liable for harm for up to €700 million per installation, the State in which the disaster occurred for up to €500 million, and other States member of the protocol are held collectively liable for up to €300 million.[8] In the United-States, the Price-Anderson act holds nuclear utili-

[6]Source: The International Atomic Energy Agency PRIS database.

[7]Source: The World Nuclear Association website. Only four countries participate to no international convention on nuclear liability: China, Iran, Korea, Pakistan and Taiwan.

[8]The revised Paris protocol is not yet implemented, as some parties still have to ratify the treaty. Liability

ties collectively liable to up to $13 billion.[9] In Japan, each nuclear plant operator can be held liable for up to JPY 120 billion, i.e. $1.12 billion. Indian nuclear operators are held liable for up to $250 millions. International conventions also limit the period of time during which an operator can be held responsible for the consequences of an accident.

Second, each country organizes the monitoring and enforcement of adequate safety standards in a specific manner. Yet, safety regulations usually entail two main features: a licensing process which aims to validate the design basis of new nuclear reactors, and a monitoring process, which consists in periodic and random audits of operated plants. In the U.S., the licensing process grants operators the right to operate a power station during 40 years, provided they meet a set of requirements. The monitoring process then entails the permanent presence of two inspectors in the power station. In France, the licensing of a nuclear plant does not come with a definitive period of operation. A nuclear operator can operate the station as long as it complies with the standards and requirements made by the French nuclear safety authority during the inspections of the plants. One major inspection is made every ten years, during which major refurbishments can be required by the safety authority to pursue operation. More minor inspections occur on a yearly basis.[10]

As nuclear accidents affect the nuclear industry worldwide, all countries that use commercial nuclear power stations are members of the International Atomic Energy Agency (IAEA), an international agency that promotes the peaceful use of nuclear power, ensures its non-military use, provides institutional and technical guidance to nuclear safety regulators from member-states, and fosters cooperation between these member-states. The nuclear safety guidance provided by the IAEA can be considered as a minimum common standard, as most member-states at least adopt the standards recommended by the agency (IAEA, 2013).[11]

thresholds defined by the former Paris Convention are thus enforced in many coutries. This original Paris protocol holds its parties liable for a total coverage of up to €360 millions. The contracted parties to the 2004 revised Paris protocol were Belgium, Finland, France, Germany, the Netherlands, Slovenia, Spain, Sweden, Switzerland and the United-Kingdon.

[9] Within this collective liability, the Price-Anderson act distinguishes the nuclear operator who owns the damaged plant from other operators. The former is held liable for up to U.S.$ 450 million. Then, all other reactor operators are held liable for up to U.S.$ 121 million.

[10] A more in-depth discussion of national regulatory characteristics can be found in Lévêque (2015a).

[11] In the United-States, the Institute of Nuclear Power Operators (INPO) gathers representatives of all U.S. utilities operating nuclear plants and conducts collective mandatory safety inspections leading to a form of naming and shaming procedure. Likewise, the World Association of Nuclear Operators (WANO) also conducts visits of nuclear stations, and gathers data regarding safety events in order to foster knowledge spillovers among its members. Contrarily to the case of INPO, membership to WANO is purely voluntary.

Definitions and perimeter of the thesis

If nuclear safety may seem like an obvious concern, national regulators, international institutions, and academic economists do not necessarily agree on its definition. Therefore, in order to clarify the scope and perimeter of this thesis, I consider here multiple definitions of nuclear safety, and discuss which definition will be considered throughout this dissertation.

The International Atomic Energy Agency defines three important and complementary concepts that describe the risks of nuclear accidents: nuclear safety, nuclear protection and nuclear security.[12]

Nuclear safety is defined as "*the achievement of proper operating conditions, prevention of accidents and mitigation of accident consequences, resulting in protection of workers, the public and the environment from undue radiation hazards*". On the other hand, *nuclear protection* is defined as the dispositions that aim to prevent and mitigate the consequences of exposure on health or the environment.

The IAEA distinguishes nuclear safety and nuclear protection from *nuclear security*, defined as "*the prevention and detection of, and response to, theft, sabotage, unauthorized access, illegal transfer or other malicious acts involving nuclear material, other radioactive substances or their associated facilities*". These intentional and man-made sources of risks will not be addressed specifically in this thesis.

The IAEA defines these three concepts as features of any source of radioactivity, regardless of the purpose of its use: electricity generation, medical diagnosis through radiography or body scans, or other industrial or military applications. Although this broad definition of nuclear safety may refer to a very large number of uses of radioactivity, this thesis will exclusively focus on the safety of nuclear power stations. This particular focus is motivated by the observation that the other risks related to radioactivity - apart perhaps from the risks associated with the proliferation of nuclear weapons - do not entail the same type of rare and catastrophic accidents.

Nuclear safety definitions often vary among countries, who are not bound to adopt the definitions suggested by the IAEA. For instance, the U.S. Nuclear Regulatory Commission does not directly define nuclear safety, but refers to safety through the *safety culture* concept, defined as "*the core values and behaviours resulting from a collective commitment by leaders and individuals to emphasize safety over competing goals to ensure protection of people and the environment*".[13]

[12] All definitions of nuclear-safety-related concepts are available on the IAEA's website.

[13] Even though nuclear safety is not directly related to the probability of undergoing a major nuclear accident in the definitions adopted by the U.S. Nuclear Regulatory Commission, they nevertheless consider the risks of accidents when regulating nuclear operators. Lévêque (2014) describes how technological standards in the U.S.

Likewise, in 2006, France adopted a legal definition of nuclear safety, as "*the set of technical and organizational dispositions aiming to prevent or to mitigate the consequences of accidents occurring during the conception, construction, operation, and decommissioning of nuclear power stations, research or fuel-cycle facilities, and nuclear material transportation activities*".

Hence, the French and American regulators define nuclear safety as a list of technical or organizational dispositions that aim to prevent, mitigate and protect; rather than as a measure of the level of prevention, mitigation or protection achieved by implementing these dispositions. In other words, these institutional definitions of nuclear safety measure how much efforts are dedicated to the protection of people and goods, rather than their chances of incurring harm. In a sense, these definitions make no difference between the efforts undertaken to improve nuclear safety and the actual safety levels achieved by these efforts.

Distinguishing safety and the efforts exerted to increase safety is a particularly important feature in the safety regulation literature, which will be reviewed in the following of this introduction. In this literature, safety refers to the likelihood that some harmful events will occur, while safety care describes how individuals, firms or policy-makers adapt their behaviours, strategies or policies to avoid accidents or to mitigate potential harm. At a more abstract level, safety can usually be modelled in economics as a probability distribution over some states of the world that correspond to harmful situations, and safety care as a choice variable that affects this probability distribution.

Distinguishing these two dimensions is paramount to study the uncertainties that characterize the risk of major nuclear accidents. Indeed, under the hypothesis of a known relation between safety care and safety, the main problem that regulators have to solve, and that has been at the center of the safety regulation literature, is the design of incentive mechanisms that reconcile cost minimization objectives with the implementation of proper levels of safety care. Yet, the failure to properly characterize the damage and probabilities associated with major nuclear accidents - and thus the failure to relate the investments required by safety regulators with quantitative improvements in nuclear safety - cast doubts on this underlying hypothesis and on the optimality of the regulatory mechanisms used in the nuclear industry.

Hence, in the following of this thesis, I refer to *nuclear safety* as the chances for the commercial use of nuclear power stations not to result in harmful consequences borne by people

nuclear safety regulations are defined based on cost-benefit analyses where the impact of a standard on the probability of major releases of nuclear materials are compared to their implementation costs. To perform this cost-benefit analysis, the NRC and nuclear operators use probabilistic risks assessments, which will be presented and discussed more extensively in the second section of this chapter.

or the environment. Additionally, I refer to the actions taken to prevent nuclear accidents, to mitigate their adverse consequences, or to protect people from these adverse consequences as nuclear *safety care* or *safety effort*.

Market failures and safety regulation

Given this definition, nuclear safety is a subject of study for economists as the harmful consequences of the commercial use of nuclear power are incurred by third parties such as the environment, or individuals otherwise uninvolved in the nuclear production process, or in the consumption of its electrical output. From a general perspective, failures to account for the negative external effects undergone by third-parties can lead to two kinds of inefficiencies. First, for a given level of risk, the output of an hazardous activity may be produced in excessive, socially-suboptimal quantities. It is for instance the case of generic pollution hazards, when a production process emits pollutants as by-products. Second, for a given level of production, economic agents undertaking the activity will under-invest in the abatement of the risks induced by their activity.

In the present dissertation, I focus on the second source of inefficiency, which is particularly salient in large projects where potential risks are not correlated with the quantity of output produced, but rather with the completion of the project itself. In the case of nuclear power, limiting the quantity of electricity produced by a nuclear station (e.g. its output) will not lead to significant safety improvements, as it is the fact that the station *can produce* electricity that creates a risk, not *how much* it produces. Thus, an important question in the case of nuclear safety is to know whether the market, a public planner or their combination are able to enforce acceptable safety levels throughout the construction, operation and decommissioning of the power station.

The rationale for underinvestment in safety care is that firms or agents exerting hazardous activities and investing in the reduction of these hazards incur all the costs of their efforts, but reap only part of the associated benefits. Examples of the private benefits of safety care can be thought of as increases in product quality, in plant reliability or in firm reputation. Yet, as safety is a non-rival good[14], third parties freely benefit from the firm's efforts, through the reduction in the risk borne or improvements in health and environmental quality. The firm, failing to internalize the positive externalities induced by its safety efforts, will under-invest in

[14] Safety is an economic good as it seems fair to assume that individual preferences for safety are locally non-satiated. It is non-rival as an individual consumption of safety does not reduce the available quantity of safety available to its peers.

safety.

The existence of external effects of an agent's actions does not directly imply that regulation is necessary. When the activity of an agent (the injurer) has an effect on third parties (the victims), and when transaction costs are null, negotiations between potential injurers and victims should yield the first-best outcome (Coase, 1960).[15]

Yet, transaction costs are rarely null and bargaining may not counterbalance these weak private incentives for safety care. A first source of transaction cost is the existence of asymmetric information between agents undertaking hazardous activities and those incurring their external effects. Examples of these information asymmetries are frequent in technological industries: a plane passenger will hardly be capable of assessing the relative safety of a plane, while the commercial airline owning the plane may hire experts capable of doing so. Likewise, nuclear operators have access to private information regarding the operation of their plants, through their past experience, and because of the limited monitoring capacities of safety authorities.

A second transaction cost that prevents victims from obtaining fair compensation from injurers is the latter's limited liability for harm. Bankruptcy laws insure firms against losses that exceed the net worth of their assets. Firms can escape liability when their responsibility is hard to prove in court: injurers may no longer exist by the time the adverse effects of their activities materialize, and harm may be shared among a potentially large number of injurers. In the example of asbestos releases in the workplace, induced lung diseases appear long after the exposure, and demonstrating causality is hardly possible. In the case of nuclear power, radio-induced cancers are not different from other cancers, and may appear decades after an accident, which hampers the compensation of victims.

These transaction costs limit the possibility for third parties to incentivize injurers to internalize the full range of consequences of their actions, and explain why safety regulations are pervasive in modern economies (see e.g. Arrow et al. (1996) for a review).

[15] For instance, Salop (1976) and Armstrong (1981) argue that a certain number of well-informed customers is sufficient to induce an optimal level of safety care. The existence of these consumers exerts pressure on manufacturers: the possibility of incurring some costs if an accident occurs, or the risk of multi-products firms to see their demand decrease if an accident involving one of their products occurs are both incentives for firms to increase their levels of safety care. On a similar note, Banerjee and Shogren (2010) show that in the presence of asymmetric information, firms subject to environmental risks and interested in their reputation will forego some of their informational advantage, loosening the constraints faced by a safety regulator. These examples show that to some extent, voluntary regulation can occur.

1.1.2 An introduction to the economics of risk regulation

The design of regulatory mechanisms inducing larger safety care from risk-taking parties has been tackled in the economics literature. I present here various incentive mechanisms that allow policy makers to foster safety care, and their various applications in the regulation of nuclear safety. I then focus on the coordination issues that arise when combining multiple regulatory instruments, and when multiple principals seeking conflicting objectives coexist.

Ex post regulation: liability limits, rules and insurance

Legal liability rules are a first instrument designed to overcome the market failures associated with the possibility for injurers to escape liability (Becker, 1974). As it may sometimes be less costly to have victims change their behaviours to avoid harm, different liability rules have been proposed in the literature to transfer the risks in an optimal manner.[16] We focus here on the specific cases where injurers and victims are well identified.

Demsetz (1972) proposes to place the liability on the party with the lowest cost of avoiding the external effects, regardless of the safety care exerted, and regardless of the cause of damage. This is known as the *strict liability rule*, whose benefits are the reduction in administrative costs, as victims no longer need to prove the responsibility of the injurer. This rule is often used when injurers and victims are clearly identified and when the cost of avoidance of the injurer is much lower than that of the victims. This is the liability regime enforced in the nuclear industry.

When firms undertaking risky projects are protected by limited liability, and financed by banks, Boyer and Laffont (1997) discuss how extending strict liability rules to lending parties may help regulators to extract more information from the firms, to foster socially desirable investment decisions, and to correct their incentives to exert due safety care. While this instrument is optimal under complete information of the lending parties, strict and full bank liability leads to suboptimal consequences when agency costs exist, such as underfunding socially desirable projects while failing to implement optimal safety care levels.

A related instrument regulators can use to enhance safety care is mandatory liability insurance. Although liability insurance first ensures fair compensation for victims, Ehrlich and Becker (1972) and Shavell (1984a) have argued that liability insurance can also convey incentives

[16] When injurers and victims have similar costs of avoiding harm, Brown (1973) proposes rules that induce both parties to adapt their behaviours. Two liability rules allow to perform such adaptation: the *negligence rule with contributory negligence*, which holds the victim liable unless she is not negligent but the injurer is. A symmetric rule which holds the injurer liable unless he is not negligent but the victim is. These rules are informationally demanding, as they require a judge to verify the injurer's and the victim's safety care.

for safety care if insurers are capable of identifying risk-reducing behaviours and link those to the premiums.

Conversely, if insurers do not have this ability, liability insurance can become counter-productive, since in its absence, at least the net worth of the firm would have been at risk. For instance, Katzman (1988) studies the determinants of the collapse of the US market for chemical liability insurance between the late sixties and the early eighties. He argues that insurers lacked skills in safety and actuarial science to identify and assess risk-reducing technologies, and determine premiums. More generally, Epstein (1996) argues that the conditions enunciated above are not met in the case of disaster insurance, and that the high correlation among individual losses caused by disasters is a further argument against mandatory liability insurance.

The private incentives for safety care provided by liability insurance may also break-down when the liability of agents is limited. This occurs for instance when governments cannot commit not to compensate victims above some level of damage. Lewis and Nickerson (1989) and Laffont (1995) describe how failures to coordinate limited liability issues with public or mandatory private insurance may give rise to the Samaritan dilemma:[17] when the uncertainty characterizing the severity of natural disasters increases, people tend to increase their risky investments and reduce their risk-reducing investments. The rationale for this behaviour comes from the existence of public relief schemes, which insures agents against the worst states of nature. Thus, any mean preserving spread in the damage distribution improves the distribution of states of nature for the individual, and worsens that of the government. Increases in wealth or reductions in risk-aversion tend to increase the gap between optimal cost-minimizing behaviours and private behaviours.

In the nuclear industry, the lack of incentives provided by limited liability regimes has been tackled by Dubin and Rothwell (1990); Heyes and Liston-Heyes (1998, 2000); Heyes and Heyes (2000) and Rothwell (2002). These authors provide assessments of the hidden subsidy provided to nuclear utilities by the limited liability regime implemented in the United-States, ranging from U.S.$ 2.3 to 22 million per reactor.year. A 2008 report by the U.S. Congressional Budget Office re-estimated this subsidy at U.S.$ 600.000 per reactor.year (CBO, 2008).

Kunreuther (1984, 1996, 1997, 2006) and Teh (2017) propose solutions to partially alleviate the limited liability issues created by non-committed public insurance. Enforcing technical

[17]The Samaritan dilemma was defined by Buchanan (1975) as the adverse effects on individual behaviours that the existence of public relief funds may induce. As an example, Kydland and Prescott (1977) cite the behaviour of groups of individuals building cities in flood-prone areas, expecting the State to build dams to protect them. The cornerstone of the existence of this effect is the impossibility for the State to commit not to help its citizens. Recent experimental evidence of such behaviours have been observed by Brunette et al. (2013).

standards (e.g. building codes in the case of flood risks), using appropriately designed partial insurance contracts, or making in-kind transfers of insurance could help to restore appropriate incentives to exert due safety care. In the case of nuclear power, implementing safety standards is the solution that has been chosen to complement strict but limited liability regimes, and is the focus of the next paragraphs.

Ex ante regulation: standards, monitoring and enforcement

Shavell (1984a) or Hansson and Skogh (1987) noted that liability rules can only induce safety care if injurers are able to correctly assess the threat embodied by the rule, and infer from it the cost-effective safety investments that should be made. Hence, when regulators are better-informed than injurers regarding cost-effective investments in safety, implementing safety standards may be more effective than relying on liability rules. Though, if the implementation of safety standards is not observable by the regulator, and if the cost of compliance can be higher than the expected cost of facing penalties for non-compliance, then regulated firms have an incentive not to comply with safety standards. Safety standards are thus socially costly, as they require enforcement resources and monitoring efforts. This gives rise to a trade-off between increasing safety and minimizing the costs of safety regulation (McKean, 1980).

A first solution to this trade-off is the coordinated use of liability rules and safety standards (Wittman, 1977). Shavell (1984b) and Kolstad et al. (1990) identify cases in which combining liability rules with safety standards allows the regulator to incentivize injurers to exert greater safety care while reducing the need for monitoring and enforcement. Strand (1994) studied the optimal governmental policy when injurers have limited liability and their safety efforts are unobservable. If profits and consumer surplus are equally weighted in the social welfare function, then a first-best level of care can only be achieved by a large subsidy, or the full confiscation of the firm's assets contingent on the occurrence of an accident. Hiriart et al. (2004) extended these models to the case of an unobservable harm with *ex post* verifiable safety effort. In this case a social optimum can be reached through a menu of incentive contracts. Extending the previous literature on the combination of standards and liability rules, Grepperud (2014) studies the case in which some workers, a firm, and a safety regulator interact in a risky environment. When the firm and its employees have an impact on the level of risk, a combination of liability rules and safety standards allow to achieve second-best outcomes, due to the large discrepancy between the worker's liability limit and the potential damage at stake.

The nuclear industry shares similarities with the abstract cases studied in these articles. Nuclear operators and their employees' liability limits are inferior to the potential damage caused, and the efforts exerted by nuclear operators are only imperfectly observable by the regulator. In addition, the regulator is not as well informed as the operators regarding the operation of nuclear power stations, and thus has to extract this private information to determine effective safety standards. This combination of factors is consistent with the observation of a combined use of liability rules and safety standards in all nuclear countries hosting commercial nuclear plants.

The enforcement of safety standards can also be optimized by performing repeated or random audits of the regulated parties by a monitoring agency. A literature on audit mechanisms shows that given a budget constraint, it is optimal not to conduct random audits among all regulated firms, but to condition audit frequencies or the penalties for non-compliance on past behaviours of the regulated firms. Macho-Stadler and Pérez-Castrillo (2006) show for instance that limited audit budgets always lead to false declarations and some level of non-compliance with existing regulations. Gilpatric et al. (2011) and Oestreich (2015) argue that shifting audit mechanisms from a random attribution of audit resources among firms to a competitive audit mechanism, in which past levels of compliance determine each firm's probability of inspection, could lead to more compliance. Experimental evidence obtained by Cason et al. (2016) confirm the former results and suggest that public-shaming procedures do not affect compliance.

It is unclear how this literature could help to improve the monitoring of nuclear power stations. In the United-States, two inspectors are permanently present within each nuclear station. This mode of monitoring first questions the assumption that regulators cannot observe the implementation of standards on site. If inspectors perfectly monitor the implementation of the standards required by the U.S. Nuclear Regulatory Commission, the question of compliance is settled. Yet, if we assume that all inspectors are equally skilled, but that two inspectors are insufficient to monitor all the stations' systems simultaneously, then this audit mechanism is equivalent to the random allocation of audit resources among all regulated parties. Yet, Feinstein (1989) showed that inspectors are not equally skilled at detecting non-compliance issues, so the takeaways from the literature on audit mechanisms would be to allocate inspectors partly[18] based on the history of compliance of nuclear operators, as well as based on the skills of each pair of inspectors. In France, there is no available data regarding the allocation of monitoring resources across the different nuclear power stations.

[18] This allocation of inspectors within power stations also has to account for other factors such as the risk of inspector capture by the inspected operator.

Safety governance: multiple principals and regulatory capture

Another coordination issue that has been addressed in the recent literature on the regulation of large risks is that of firms subject to multiple regulations. For instance, when firms subject to large environmental risks are natural monopolies, they may face both safety and economic regulation. When a monopolist is subject to both economic and safety regulations, Laffont (1995) shows that high-powered incentive schemes will diminish the incentives to exert safety care. A solution to alleviate this moral hazard issue is to use low-powered incentive schemes that do not foster cost reductions. The French regulation of the nuclear electricity does not follow this recommendation, as the sole nuclear operator, EDF, has to sell up to 75% of its nuclear electricity to electricity retailers at a fixed regulated price. As this regulated price is seldom lower than the electricity market price, there is little demand for regulated nuclear electricity, which could dilute the incentives for EDF to curb down its cost of producing nuclear energy.

The combination of economic and safety incentives has also been tackled in a dynamic procurement setting by Biais et al. (2010), who show that a public-planner can act as a safety regulator. Incentives for safety care can be provided to the private contractor by conditioning the size of the project and possibility of being awarded future contracts on the level of safety care implemented during past and present stages. This result seem not to be applicable to the market for nuclear power stations, which have to go through a licensing process and a subsequent continuous monitoring that ensure an appropriate level of safety care from the vendor. Yet, it could describe well the market for fuel cycle services, maintenance activities and construction of large replacement pieces such as steam generators, as the demand for these services is repeated throughout the life of a nuclear reactor.

A second potential failure of safety regulation due to the interaction of multiple agents driven by different objectives is regulatory capture. Judges that evaluate safety care after an accident and inspectors that assess compliance with safety standards can both be bribed for their leniency, either by firms trying to reduce the stringency of their oversight, or by policy-makers driven by other motives than the simple reduction of potential harm. An important coordination issue in the safety literature is thus the governance of nuclear safety regulation, and its robustness to regulatory capture.

Solutions to the possibility of regulatory capture by risk-taking firms has been investigated by Hiriart et al. (2004, 2010), who argue that separating the institutions responsible for setting safety-standards *ex ante* and enforcing liability rules *ex post* increases the cost for the injurer to capture its regulators. This separation of judges and regulators is observable in many countries in the case of nuclear power. In France, the penalties in case of an accident are set by law rather than by the regulator, who is only in charge of the licensing of reactors and of the monitoring of nuclear operators. In the United-States as well, regulators in charge of licensing and monitoring nuclear operators are independent from the judiciary power in charge of enforcing liability laws.

In a third paper, Hiriart and Martimort (2012) address the risk of capture of a risk-regulator by its governing body (say the Congress). A Congress driven by electoral objectives rather than purely safety-related motives, may find it optimal to only partially delegate the task of enforcing proper safety care to the regulator, by setting limits on the sanctions the regulator can levy on risk-taking firms. Institutional caps on fines, observed in many instances of risk regulators, can thus be a form of institutional capture of risk regulators by their governing bodies, and the independence of risk-regulators seems insufficient to ensure the effectiveness of their actions. Regulators also have to be delegated sufficient enforcement power to effectively incentivize agents to exert due safety care.[19]

The strategic institutional capture described in this article can be illustrated by the multiple liability laws that limit the liability of nuclear operators in almost all countries using nuclear power stations. The aforementioned Price-Anderson act in the U.S. or the French legislation on nuclear liability, which limit the liability of nuclear operators to €91 millions, are examples of such laws.

These laws are framed by multiple international treaties that were elaborated by countries using nuclear energy as well as countries which do not rely on it, in order to facilitate international compensations after nuclear accidents. In particular, the Paris Convention was ratified by OECD members in 1960 and the Vienna convention was ratified in 1963 by IAEA members. These two original treaties were then linked in 1988 by a joint protocol, and supplemented by amendments to increase the coverage of populations. These treaties advocate liability caps for nuclear operators, and justify these caps as a counterpart to the strict liability regimes to which operators are subject, as well as a way to account for the external benefits of nuclear power[20]. An instance of such external benefits is the curtailment of the environmental externalities that characterize other fossil fuel generation technologies, and which are not properly internalized.

This rationale is not totally convincing from an economic standpoint, as it consist in creating

[19]This result echoes the more general political economy framework developed by Boyer and Laffont (1999), which stipulates that flexible environmental regulations ought to be enacted when the variability of environmental issues is large and the biases in politicians' objectives are small. In this framework, allowing politicians to set caps on regulators' actions can be seen as form of flexibility in safety regulation.

[20]Source: The World Nuclear Association, Liability for Nuclear Damage: http://www.world-nuclear.org/information-library/safety-and-security/safety-of-plants/liability-for-nuclear-damage.aspx

a new market failure (limited nuclear liability) to correct a pre-existing one (global warming), and as the fact that these caps on liability undermine the ability of regulators to foster safety care is indisputable. Yet, this second rationale competes with strategic institutional regulatory capture to explain the existence of these liability caps.

1.2 Regulation under catastrophic risks and uncertainties

Although existing practices observed in the regulation of nuclear safety are well grounded in the existing economic theory of safety regulation, the following sections aim to show that some characteristics of nuclear safety raise additional questions regarding the way nuclear safety is regulated. In particular, the failures to properly quantify the risks associated with nuclear power lead to two issues. First, quantifying the impact of safety care and safety regulations on nuclear safety is hampered by the rare frequencies of nuclear accidents, and our lack of knowledge in the identification and measurement of nuclear damage. Second, the perceptions of the nuclear risks vary among the population, as well as among experts, which casts doubts on the definition of the objectives of safety regulators.

1.2.1 Nuclear safety and the failures of risk measurement

On nuclear accident probabilities and uncertainties

One of the practical reasons not to define nuclear safety as economists would - i.e. as the probability of inflicting harm to people or goods - is the fact that quantifying the risks of extremely rare events can prove to be a difficult conceptual task. Following the classical definition proposed by Savage (1954), defining safety as an *objective probability* of inflicting harm to oneself or third parties is possible when the underlying hazardous activity results in frequent harmful events, which satisfy the concept of independently and identically distributed events.

In this context, nuclear accidents are extremely rare, as twenty-four events leading to damaged nuclear pressure vessels[21] have been recorded since the beginnings of the nuclear industry (Cochran, 2012). Over this operating period of seventeen thousand reactor.years[22], safety standards have evolved differently in all countries relying on nuclear power, and some events were correlated, such as the three core-meltdowns which occurred in 2011 at the Fukushima-Daiichi station. Hence, the repetitions of operating years cannot be assumed to be an independently

[21]The nuclear pressure vessel of a nuclear reactor is the part of the reactor that hosts the nuclear fuel.
[22]Source: the IAEA PRIS database.

and identically distributed process, and inferring an objective probability distribution based on this operation history is precluded.

To overcome this measurement shortcoming, probabilistic risk assessments were developed in the late seventies in the United States (see e.g. Rasmussen (1975)). PRAs consist in bottom-up safety analyses based on event-trees and Monte-Carlo simulations of failure scenarios that may bring about nuclear accidents. They measure which systems are the most vulnerable, which combinations of factors may lead to the rupture of the nuclear confinement barriers and, in case of a release of nuclear materials in the environment, the consequences of this radioactive leakages[23]. Initially developed to pinpoint design weaknesses within nuclear reactors, PRAs have been used to provide insights regarding the future probabilities of nuclear accidents. For instance, the French Institute for Radioprotection and Nuclear Safety (IRSN) recently used probabilistic safety assessments to calculate an estimation of the expected cost of nuclear power accidents (IRSN, 2013).

More recently, statistical analyses of past nuclear accidents have either adopted a Bayesian approach (see e.g. D'Haeseleer (2013) and Rangel and Lévêque (2014)) or widened the scope of their assessments to many other events of different nature (see e.g. Hofert and Wüthrich (2011) and Wheatley et al. (2017)).

The Bayesian approach consists in defining a prior based on available data at some point in time and update it dynamically based on annual observations of accidents. This process produces a subjective posterior probability of nuclear accidents, which depends on the choice of the original prior distributions, and on additional parametric assumptions. Using a poisson exponentially weighted moving average model to update a prior probability distribution based on the first PRA published in the United states, Rangel and Lévêque (2014) show that the Fukushima-Daiichi accident should result in a significant update of the subjective probability of nuclear accidents, that would bring it back to its post-Chernobyl level, to reflect the lessons learnt from the catastrophe.

The second approach is based on datasets which gather events from different plants such as power stations or fuel cycle facilities.[24] Using these broad sets of events, Hofert and Wüthrich (2011) and Wheatley et al. (2017) fit annual loss distributions on past nuclear accidents. Hofert and Wüthrich (2011) show that nuclear accidents are characterized by an infinite mean distributions, which precludes the use of market-based insurance against this risk and calls for efforts

[23] See D'Haeseleer (2013) for further references and details or Paté-Cornell (1996, 2002) for further examples of the use of PRAs in non-nuclear related contexts

[24] See e.g. Hirschberg et al. (2004); Sovacool (2008) for further details on these aggregated datasets

in the direction of limiting the potential consequences of nuclear accidents. Wheatley et al. (2017) stress the runaway-disaster (or *dragon king*) nature of nuclear accidents: passed a certain damage threshold, accidents get out of control and their associated damage increase drastically. In other words, their associated distribution of losses is very heavy-tailed. However, by relying on heterogeneous events taking place in every step of the nuclear fuel cycle, the distributions obtained by these studies do not characterize the risk of undergoing a major disaster in a nuclear plant, but rather in the nuclear industry in general. It is thus less informative for a decision-maker facing, for instance, a specific investment decision regarding a power station, a waste recycling facility, or a fuel processing plant.

The existence of a multiplicity of models used to measure the risks of nuclear accidents raises questions as to the applicability of the prescriptions of the literature reviewed in the previous sections, as these prescriptions usually rely on the existence of a probability distribution over the potential damage caused by accidents. This multiplicity of models motivates chapter 2, in which I provide a framework for the estimation of the expected cost of nuclear accidents that accounts for the attitude of individuals toward model uncertainty.

Assessing the damage caused by nuclear accidents

The lack of observations of past nuclear accidents is not the only obstacle that hampers the measurement of a probability distribution over potential levels of damage caused by nuclear accidents. A second obstacle to the measurement of the risk of nuclear accidents are the uncertainties that characterize the damage associated with nuclear accidents themselves. Two kinds of assessments can be distinguished. Some studies have tried to assess the extent of the consequences caused by specific past events such as Chernobyl (Hohmeyer, 1990; Ottinger et al., 1990; Ewers and Rennings, 1991, 1992). Others have tried to use available information from past events to estimate the expected cost of future nuclear accidents (Eeckhoudt et al., 2000; Rabl and Rabl, 2013; IRSN, 2013).

These two types of studies differ in their motivations. The former type of studies are legally-driven. Their aim is to assess the consequences of a disaster to determine fair compensation for victims. The latter type of studies are economically-driven: they aim to provide guidelines for decision-makers faced with choices regarding the future use of nuclear power.

Large discrepancies appear in both types of assessments of the damage caused by nuclear accidents. Three main sources of uncertainty can be identified: the definition of the perimeter of the consequences of the accident, the assessment of the health effects induced by a nuclear accident, and the monetary valuation of some particular losses.

First, the scope of consequences assessed vary from one study to another. Studies can focus on different types of costs: while the studies dedicated to the Chernobyl catastrophe focused on health and agricultural costs, more recent studies have focused on larger panels of consequences. Hayashi and Hughes (2013) shows for instance that the Fukushima-Daiichi accident has impacted the electricity bills of households in gas-intensive countries such as the United-Kingdom or South-Korea.

Different studies can also focus on different geographical areas as is shown by two reports on the consequences of the Chernobyl accidents that were published in 2006. A first report commissioned by the IAEA and the World Health Organization (IAEA, 2006) focused on the consequences of the accident in Belarus, Ukraine and Russia. The second report accounted for all consequences across Western Europe (Fairlie and Summer, 2006). Their assessments of the number of radiation-induced cancers differ by a factor ten, which illustrates the local nature of damage assessments, and their sensitivity to the definition of their boundaries.

A second source of uncertainty is the assessment of the effects of radioactivity on health, which are subject to various controversies, reviewed by Kathren (1996). A compelling example of these controversies is the so-called *linear-no-threshold model.* Until the bombings of Hiroshima and Nagazaki, the state-of-the-art knowledge about radiation safety assumed that the human body would recover from any exposure to radiations below some acceptable threshold. After the second world war and the beginning of nuclear weapon try outs, the idea that low levels of radiations could be harmless was challenged. The linear-no-threshold model was developed, and claims that the excess rate of cancers observed in a population is proportional to the dose absorbed by the population. For instance, Ottinger et al. (1990) calculates the damage associated with the Chernobyl accident using an absorbed dose of 2.4 million person.Sievert and a 7.7% carcinogenic coefficient, and concludes that 185.000 cancers should be attributed to the nuclear accident. This model has been widely debated in the dose-response literature[25], as it stands on the controversial hypothesis that the effects of radioactivity on health can be drawn from the population-aggregated absorbed dose, thus considering as equivalent a small number of highly irradiated persons and a large number of weakly irradiated ones.

Finally, the monetary value of unusual monetary losses such as food-bans have been derived from empirical assessments which have tenuous similarities with nuclear accidents. For instance, the French institute for nuclear safety and radio-protection evaluates the potential cost of a food

[25] See e.g. Cohen (1990, 1995); Hooker et al. (2004); Tubiana et al. (2009) or Doss (2013).

ban associated with a nuclear accident using data from the Spanish cucumber ban which occurred in 2011 after an outbreak of E.coli bacteria. The monetary value of non-monetary losses such as lost lives or polluted environmental amenities also require parametric assumptions which may differ from one study to another. For instance, Hohmeyer (1990) and Ottinger et al. (1990) both use the human-capital method to assess the health cost of nuclear accidents. Yet they respectively adopt $ 1 million and $ 4 million as the statistical value of life.

Empirical assessment of nuclear safety

The uncertainties regarding the assessment of the consequences of nuclear accidents add up to the uncertainties associated with the measurement of their probabilities of occurrence. This raises the question of the measurement of the effect of safety efforts on safety. This measurement is essential if one wants to focus regulatory efforts on the design of effective incentive mechanisms.

Indeed, by requiring mandatory investments in safety, or by distorting production in order to limit the risks borne by third parties, safety regulation have significant effects on firms' performance and the economy. In the United-States for instance, Gray (1987) estimates that 30% of the productivity decline observed in the U.S. during the seventies could be attributed to environmental and safety regulations, while Arrow et al. (1996) report an annual total direct cost of enforcing environmental, health and safety regulations of approximately U.S.$ 200 billion. Thus, implementing regulations that effectively reduce risks and promote the internalization of hazardous externalities is a leading economic question. To answer this question, ex post econometric analyses can assess the effect of existing regulations - or the absence of regulations - on safety levels, and provide guidelines for policy-makers to improve safety-related policies.

For instance, several examples of ineffective safety regulations have been reported in the fields of food safety, occupational safety or road safety. Grabowski and Vernon (1978) criticized the efficiency of the allocations induced by a reform of the Food and Drug Administration. Viscusi (1979) argued that a reform of the Occupational Safety and Health Administration was not stringent enough to impact safety investments. Lewis-Beck and Alford (1980) and Neumann and Nelson (1982) criticized the impact of occupational safety reforms in the mining industry. In the road safety sector, although Robertson (1981) and Michener and Tighe (1992) found significant positive impacts of safety standards such as mandatory lap seat-belts, Michener and Tighe (1992) and Keeler (1994) found that speed limits implemented in the eighties in the U.S. had no significant impacts on fatal accidents. Finally, in a review of cost-effectiveness-analysis, Tengs et al. (1995) questions the rationality of the allocation of public resources, as the costs of life-saving interventions reviewed range from negative values (i.e. strictly cost-beneficial interventions) to up to 10 billion dollars per year-of-life saved.[26]

Likewise, a large number of empirical studies have shown that, to some extent, the market can provide incentives for safety care. Filer and Golbe (2003) find that larger debt or higher operating margins significantly improve the level of private investments in workplace safety. The introduction of competition in the airline industry during the eighties has also been investigated by Borenstein and Zimmerman (1988); Moses and Savage (1990); Rose (1990); Dionne et al. (1997) or Adrangi et al. (1997), who find that larger profitability are correlated with higher safety levels. These studies suggested, among other policy recommendations, to tailor safety regulations and focus monitoring efforts on the firms characterized by the lowest performance.[27]

Overall, the findings from these various empirical studies stress the need to design safety regulations that effectively convey incentives for safety, in order to avoid bearing undue economic costs attached to weak, let alone null, safety improvements. Second, these studies embody the need to conduct ex post analyses of existing regulations, to ensure that public resources are allocated effectively.

These needs have only been partially met in the nuclear safety literature, as the econometric analyses of the effect of market or regulatory incentives on nuclear safety are relatively scarce. To the best of our knowledge, three exceptions may be mentioned. First, Feinstein (1989) used data on inspections of nuclear stations to characterize the propensity of nuclear operators not to comply with safety standards, and to describe the heterogeneity in detection abilities that exists among the inspectors of the U.S. Nuclear Regulatory Commission. Davis and Wolfram (2012) and Hausman (2014) used counts of safety incidents to show that the divestiture of some U.S. nuclear power stations in the late nineties had had a positive impact on the economic performance, production reliability and safety levels of the divested nuclear stations.[28]

The two empirical chapters of this thesis contribute to bridging this gap by performing two ex post analyses of the French nuclear safety regulation. We first measure variations of safety - measured as occurrences of certain types of significant safety events - over time and across reactors of different ages and technologies. Second, we measure the effect of an existing regulation

[26] An in-depth discussion of the normative rationale (and shortcomings) of cost-benefit analysis for health and safety regulation has been provided by Hammitt (2013).

[27] On the other hand, Raghavan and Rhoades (2005) find adverse effects of commercial airlines profitability on safety levels, which cast doubts on these previous recommendations.

[28] A fourth exception that does not focus on nuclear power stations is the aforementioned paper by Wheatley et al. (2017), which studies the historical evolution of occurrences of nuclear accidents in the whole nuclear fuel cycle. The authors measure a continuous decrease in the frequency of nuclear accidents, as well as a statistically-significant structural decrease in this frequency after the Chernobyl disaster, which they interpret as an increase in safety due to the upgrades that were implemented after the 1986 catastrophe.

on the behaviour of nuclear plant managers, and in particular on their propensity to exert safety care and to comply with existing regulations. Both studies aim to provide statistical tools and guidance to better tailor enforcement and monitoring resources to a potentially heterogeneous fleet.

1.2.2 Safety regulation under catastrophic risks, misperceptions and uncertainty

The uncertainties listed above question the applicability of the prescriptions of the safety regulation literature to cases such as nuclear safety. The problems these uncertainties raise for environmental or safety regulation have been stressed by Laffont (1995) or Boyer and Laffont (1997), and some descriptive or normative solutions have been proposed. The following of this section reviews these contributions.

Safety regulation under biased risk-perceptions

Numerous examples observed in the experimental psychology literature have shown that the perceptions of the risks associated to nuclear power can largely differ from one population to another. First, these perceptions have been found to be driven in part by the amount of information and knowledge available regarding the risk. Sjöberg and Drottz-Sjöberg (1991) survey employees from Swedish nuclear power stations and find that workers hired by subcontractors report larger risk than other workers. Kivimäki and Kalimo (1993) survey 1000 employees in Finnish nuclear power stations and find that nuclear employees perceive the risk of nuclear accidents as less likely than the general population.[29]

Nuclear risk perceptions were also found to be largely driven by psychological motives, such as the feeling of security or the public legitimacy of hazardous activities (Bastide et al., 1989). Peters and Slovic (1996) and Sjöberg (1998) show that world-views and emotional factors such as worry and affect significantly drive individual perceptions of the nuclear risk.

In the domain of nuclear waste management, risk perception were also shown to be determined by various other factors (Sjöberg, 2004; Sjoberg and Drottz-Sjoberg, 2009), such as attitude towards nuclear power, past military failures to handle nuclear wastes, and distrust regarding either the authorities in charge of managing nuclear wastes (Slovic et al., 1991) or the

[29]Interestingly, Wahlberg and Sjoberg (2000) reject the hypothesis according to which the media would be a strong driver of personal risk perceptions. This result contradicts some of the conclusions of Rothman and Lichter (1987), who argued that the ideology of journalists could explain the difference between public and experts' risk perceptions.

science describing the physics and technology of nuclear waste storage (Sjöberg, 2009).

Another issue raised in this literature is the conflict between the risk perceptions of the public and experts. A wide body of evidence (see e.g. Slovic et al. (1980, 1981); Fischhoff et al. (1983); Rothman and Lichter (1987); Kasperson et al. (1988); Pidgeon et al. (2003) and references therein) suggests that perceptions of risks by laypeople differ from the perceptions by experts, and that experts' perceptions are driven by objective evidence such as annual casualties or technical risk assessments. Yet, Sjöberg (2002) only finds empirical evidence for the first part of the previous results: the risk perceptions expressed by experts do differ from laypeople perceptions, but the psychological determinants of their risk perceptions are similar, contrarily to their attitude towards nuclear power, which the author proposes as a main driver of the discrepancy between experts' and the public's risk perceptions.

Related to the issue of conflicting risk-perceptions is the disagreement among experts themselves. Barke and Jenkins-Smith (1993) find significant variations in the risk perceptions expressed by scientists, across their fields of research, and across types of institutions of affiliation. Sjöberg (1999) argues that this disagreement among experts is a driver of public scepticism and distrust regarding risk management and risk managers, and that this scepticism and distrust in turn require an adaptation of public policy-making. This is consistent with Slovic (1993) who argues that public perceptions of major risks such as nuclear power are driven in part by the amount of trust populations have in the people managing the risks. These findings were empirically measured by Whitfield et al. (2009) who found that increases in public trust led to lower perceived risks.

Finally, the effect of major accidents on risk perceptions has been investigated. Drottz-Sjöberg and Sjoberg (1990) measures the effect of the Chernobyl accident on nuclear risk perceptions in Sweden, and find that people from one of the most affected area are twice as worried by nuclear power than people from other surveyed areas. Huang et al. (2013) and Prati and Zani (2013) measure the effect of the Fukushima-Daiichi nuclear accident in China and Italy, and find a significant decrease in trust towards nuclear issues. Quantitatively, Huhtala and Remes (2017) provide an assessment of the social cost of nuclear risk perceptions. This study suggests that policy-makers should consider a social cost of risk-perceptions as high as 4-7€/MWh[30]. This social cost of risk perceptions is defined as the value of the electricity lost due to the opposition of voters to the construction of nuclear reactors, and based on the results of an opinion survey conducted in Finland.

[30] As a comparison, this figure is four to seven times as high as the expected external cost of nuclear accidents considered in the study by Huhtala and Remes (2017).

A question raised by these discrepancies in risk perceptions is the basis on which the nuclear risk should be regulated. In particular, an open question is whether policy-makers should rely on the perceptions held by populations or on the views of their experts. In his Happyville parable, Portney (1992) questions how society should regulate risks when regulators and populations do not share a common prior regarding the damage at stake. In particular, what should a public planner do when he believes that a given investment in safety is useless, whereas the entire population would agree to make the investment anyway?

To answer this question, Salanié and Treich (2009) oppose populist regulators, who regulate risks based on the population's perceptions, and paternalistic regulators, who use their own prior of the risk at hand. Whereas populist regulators will increase their investments in safety when the population is too pessimistic, and reduce these investments when the population is too optimistic, the authors provide a rationale for paternalistic regulators to increase safety investments in both cases. Increasing safety investments is intuitive when populations are too optimistic regarding the risk they face. On the other hand, if people over-estimate the risk they face, and if the regulator follows his own prior to determine the level of safety investments, people will reduce their consumption of the risky asset. Hence there is a trade-off between cutting down ineffective safety investments and fostering consumption of the not-so-risky asset. Under some assumptions, the authors show that the equilibrium of this trade-off always results in an over-investment in safety, from the perspective of the regulator.

This result can be seen as a positive explanation of the continuous upgrades in safety that are implemented in nuclear power stations. Given that the perceptions of the population of the risk associated with the use of nuclear power are larger than those of experts, safety regulators might voluntarily increase their safety requirements in order to convince the population that the use of nuclear power is safe, although, in their subjective perception of the risk, these investments may be cost-ineffective.

In a similar spirit, Brandt (2014) considers extremely infrequent risks and a populist form of regulation, and argues that this form of regulation leads to welfare losses and over-investments in safety. Drakopoulos and Theodossiou (2016) show that regulatory intervention is needed to increase workplace safety when workers misperceive the risks they face. Finally, when risk-perceptions are affected by safety regulation, Grüne-Yanoff and Rosencrantz (2011) argue that a regulator may find it optimal to voluntarily deteriorate safety. The rationale for this behaviour is a form of rebound effect: if more stringent safety regulations induce agents to be more biased into thinking they are safe, they will make riskier decisions. Hence, voluntary reductions of safety may reduce the bias of agents exposed to the risk, and foster more safety care.

An important feature of the three references mentioned in the previous paragraph is the common assumption that policy-makers or experts hold an objective assessment of a social risk, while the population holds a biased assessment of it. This allows the authors to derive normative general welfare assessments, i.e. to qualify the optimality of some given levels of investment in safety. An important characteristic of the nuclear risk mentioned above is that experts themselves disagree on the assessment of the risk. Claiming that nuclear regulators know a true objective probability distribution is thus unwarranted. In this sense, Salanié and Treich (2009) only define subjective notions of utility that differ between the population and the regulator and show that over-regulation is the best course of action for a regulator trying to further its interests.

An interesting empirical result related to this question of the use of public perceptions of risks in policy-making is provided by Carlsson et al. (2012) who, using a stated-preference approach, find no strong evidence of divergence in the opinion of policy-makers and individuals regarding what policy-makers risk-reduction priorities should be. An exception is made for the case of large accidents, for which individuals prefer the investments in the mitigation of many small accidents, while policy-makers are empirically split between mitigating small and large risks.

Social choice, catastrophic risk and uncertainty

In the literature described before, regulators were assumed to be expected-welfare maximizers. Yet, little attention was paid to the aggregation of individual utilities into the social welfare function used as the regulator's objective. In particular, the optimal mitigation policies obtained when computing welfare as a sum of individual utilities may lead to unfair outcomes, as catastrophes and risky activities can harm specific subgroups of the population, such as future generations, or population located in particular areas.

A second open question is whether catastrophes and more frequent risks should be accounted for in the same way in public policy-making. For instance, in the respective fields of technological and climate risks, Starr (1969) and Chichilnisky (2000) advocate the use of social welfare-functions that exhibit *catastrophe-averse* preferences, i.e. that emphasize the costs of rare disasters when compared with more frequent but less harmful events. Conversely, Rheinberger and Treich (2016) perform a review of experimental studies in which individuals were proposed to make social decisions involving potential catastrophes (including, for instance, the Asian disease experiment by Tversky and Kahneman (1985)). Within this literature, they find no evidence for catastrophe aversion. However, they provide evidence of stated-preferences expressed by risk-regulators such as the U.S. Nuclear Regulatory Commission that can be assimilated to catastrophe aversion.

Adopting a theoretical perspective on these issues, Bommier and Zuber (2008); Fleurbaey (2010); Bovens and Fleurbaey (2012) and Fleurbaey and Zuber (2017) provide solutions for the aggregation of individual preferences into social welfare functions when ex-post equity is a concern and some form of risk is present. Fleurbaey (2010) proposes an aggregation criterion when a population of expected utility maximizers shares a common prior regarding a known risk. In order to satisfy dominance, inequality-aversion and a weak version of the Pareto principle, the proposed welfare function is computed as the weighted sum of individual utilities, where weights are defined according to the rank of an individual's utility level in the population. This criterion differs from Harsanyi's weighted utilitarianism in which individual weights are constant across states of the world (Harsanyi, 1955).[31] If weights increase when the rank of an individual in the population decreases, this criterion favours ex-post equity. This paper solves the equity issue that may be raised by rare disasters. Yet, it does not address the fact that social preferences may differ according to the value of the common prior (e.g. exhibit catastrophe aversion), or that the existence of a common prior can be challenged by biased risk-perceptions.

Other welfare measures based on equity concerns and severe risks have been derived in the literature. Bommier and Zuber (2008) consider catastrophe risks under catastrophe aversion, e.g. differentiated attitudes regarding repeated small risks and rare but large risks. Bovens and Fleurbaey (2012) consider the case in which social decisions entail mortality risks. When individual preferences need not be represented by utility functions, Fleurbaey and Zuber (2017) propose an aggregation criterion that encompasses uncertainty-averse preferences. According to this criterion, the social preferences are defined as the maxmin-expected utility derived by the individual worst-off in the population, computed according to the worst-case prior selected over the set of all individual beliefs. This particular criterion is further described and applied to the case of the assessment of the expected social cost of nuclear power accidents in chapter 2.

Mechanism design under uncertainty

If catastrophes raise concerns over the ex-post equity of the outcomes induced by classical welfare functions, the uncertainties that characterize catastrophes also hinder the normative adequacy

[31] Fleurbaey's criterion is equivalent to applying Harsanyi's weighted utilitarianism to a virtual society composed of virtual individuals obtained by re-ranking prospects within each state of nature so that virtual individuals keep the same ranking in the population across each state.

of classical incentive mechanisms with the market failures they aim to solve. In other words, if the effect of an agent's action on the cost, or quality, or safety associated with the production of a good is uncertain, or if the signals regarding the performance of agents are ambiguous, then the validity of the results presented previously are somehow at doubt. This question has been tackled in a literature dedicated to mechanism-design under *ambiguity*, that we will define in the following paragraphs as well as in chapter 2 as the fuzziness of the probabilities describing the likelihood of occurrence of some events.[32]

First, allowing for ambiguous beliefs in a mechanism design environment can provide positive results, e.g. alternatives to classical asymmetric information models for the explanation of observed behaviours. Ambiguous beliefs have been used to explain the incompleteness of contracts (Mukerji, 1998), as an alternative theory explaining the predominance of cost-plus contracts in R&D procurement (Mukerji, 2003), or the relative use of delegation (or flexibility) in procurement contracts (Gottardi et al., 2017). In contexts in which firms or regulators have no past experience, this explanation may be more accurate than competing ones, such as asymmetric information leading to classical rent-efficiency trade-offs, or risk-averse agents seeking for insurance against poor *ex ante* previsions.

From a normative perspective, Ghirardato (1994) showed that non-additive preferences should not radically change the structure of incentive mechanisms, which only become sensitive to the quality of the information available to the principal. In a moral hazard setting with incomplete preferences, Rigotti (1998) derived an optimal incentive mechanism, in which the principal uses simple two-part contracts that are robust to a set of priors. Mukerji (2003) shows that the power of incentive mechanisms is negatively related to the relative ambiguity-aversion of agents with respect to the principal. Karni (2009) adapts Gilboa and Schmeidler (1989)'s Maxmin-Expected-Utility (MEU) theory to a principal-agent setting, and shows that implementing high levels of effort is more costly to the regulator when agents are ambiguity averse. In equilibrium, the effort implemented can be shifted towards lower levels when both the agent and the principal are ambiguity-averse. Kellner (2015) proposes the use of tournaments, e.g. individual pay-offs that depend on the outcomes obtained by other agents, as a way to alleviate these issues. In an adverse-selection environment, Giraud and Thomas (2017) show that ambi-

[32]In this review, we focus on the literature on mechanism design under ambiguity in which ambiguity characterizes the effect of an agent's action on the underlying outcome of the contractual relation, or on the ability of the principal to elicit the agents' types. This literature extends to more general cases in which the mechanism itself can be ambiguous, for instance when principals can communicate with agents using ambiguous communication devices (see e.g. Bose and Renou (2014); Kellner and Le Quement (2015) or Tillio et al. (2016)). This literature is not reviewed here as it did not seem to bear interesting policy implications regarding nuclear safety regulation.

guity hampers the ability of a principal to extract the private information of agents, and thus profits to inefficient agents.

To the best of my knowledge, only Melkonyan and Schubert (2009) propose a model of safety regulation under ambiguity. Their model is dedicated to food safety regulation, and captures the uncertainty that characterizes food-borne hazards. The authors assume that both regulators and firms behave according to the α-maxmin expected utility maximization proposed by Ghirardato et al. (2004).[33] The regulator's actions are determined based on the realization of an ambiguous safety signal. As a result, the regulator should induce actions characterized by lower levels of ambiguity, and the power of the incentive schemes used should increase when the ambiguity characterizing actions increases.

These results have several implications for nuclear safety regulation. A first alarming take-away is that when ambiguity affects the adverse selection problem facing the regulator, e.g. the elicitation of the underlying safety of nuclear power plants, ambiguity may profit to the least safe operators. The second set of results described above affects the moral hazard problem of the regulator. The levels of safety care implemented, and the power of the associated incentive mechanisms, depend on both the ambiguity characterizing the effect of the agent's actions, and the attitude of the regulator and of the agent towards ambiguity. This calls for efforts in the direction of reducing the ambiguity that characterizes the effect of nuclear safety care on nuclear safety, which can take two forms: developing methods that better measure the link between efforts and safety, or by identifying safety efforts that are less ambiguous. In chapter 4, I engage in the former type of effort by proposing an empirical strategy that allows to measure the effect of a nuclear safety-related policy on nuclear safety.

1.3 Objectives and summary of this dissertation

1.3.1 Research question

To sum up, nuclear accidents in commercial power stations can hardly be represented using probabilistic tools. Their objective probabilities of occurrence, defined in the sense of Savage (1954), are not conceptually well grounded. The measurement of their associated damage is hampered, among several reasons, by our lack of knowledge of the long-term effects of radia-

[33] Mukerji (2003) and Melkonyan and Schubert (2009) are both subject to the comment of Eichberger et al. (2011), who showed that α-MEU preferences over acts defined on a finite state-space have to satisfy $\alpha = 0$ or $\alpha = 1$ to be consistent with their original axiomatic definition proposed by (Ghirardato et al., 2004). This comment casts doubts on the validity of the comparative statics derived by these two studies.

tions on health. Public perceptions of these risks are not in line with experts perceptions - let alone calculations - all of which may be biased by psychological or emotional factors. Finally, the regulation of the risk of nuclear accidents has been constructed on the basis of a continuous upgrade of safety care, e.g. through the multiplication of safety systems and organizational improvements. Yet, measuring the impact of safety care on the risks borne by either the population or the environment is precluded by the characteristics of nuclear accidents.

In addition, the existing economic theory of safety regulation is based on the assumption that the level of safety associated with any given amount of safety care is known by individuals, voters, policy-makers, plant operators and safety regulators. For instance, this literature agrees that the first-best or socially-optimal level of investment in safety care should equate its marginal cost with the marginal benefits derived from the reduction of accidents damage or frequencies. But as these marginal benefits are hardly measurable, it seems important to wonder whether these prescriptions remain accurate.

Although an objective probability distribution driving the occurrence of nuclear accidents may not exist, decision-makers may nevertheless form a subjective prior regarding this risk, based on all available information, however scarce it may be. In this case, this subjective prior can be used by the decision-maker to derive optimal levels of safety investments, or to implement the other prescriptions from the existing economic theory of safety regulation described above. Yet, the normative appeal of these prescriptions can be questioned, considering the fact that the prior used to define these prescriptions may differ significantly from the priors held by laypeople or by other experts.

For instance, consider the simplest prescription of the economics of safety, i.e. that the first-best level of safety care equates the marginal social costs of care with its marginal social benefits. As the marginal social benefits of safety care are defined as a function of the change in the distribution over damage induced by safety care, any expert able to form a subjective prior over potential damage can come up with a subjective assessment of the first-best level of safety care. As the prior distribution over damage used to derive this first-best level of care is necessarily subjective, and may vary across experts and the population, then, how appealing is the first-best level of care determined by any particular expert? The same reasoning holds for other prescriptions of the literature on safety which require the use of a distribution over potential damage, such as the optimal trade-off between safety standards and liability-rules, or the optimal use of monitoring and enforcement resources, or the optimal level of delegation between congress and safety regulators.

The four chapters of this thesis deal with two different but related questions regarding the uncertainties that characterize nuclear accidents. This thesis is thus divided in two parts. The first part of this thesis tackles ex ante safety measurement when accidents are rare. The second part of this thesis engages in the analysis of the implementation of ex ante safety regulation and ex post disaster management strategies when risks are rare and catastrophic. Both parts are composed of two chapters.

In the first part of the thesis, chapter 2 presents a theoretical framework for the assessment of the expected social cost of nuclear accidents, that accounts for the diversity of scientific models describing the probabilities associated with major nuclear disasters. Using this framework, I estimate an expected cost of a nuclear accident that accounts for the aversion of individuals towards model uncertainty, and discuss the policy implications that this cost conveys.

Second, in chapter 3, I present a new dataset of minor but significant safety incidents reported in the French fleet. I argue that these events can be used to derive new insights regarding nuclear safety. In particular, I describe various statistically significant variations of nuclear safety over time and across reactors of different ages an technological designs.

In the second part of this thesis, chapter 4 assesses the causal effect of a French policy on safety and compliance within nuclear plants using data on reported safety incidents. Since 2006, this policy requires French Departments hosting nuclear plants to constitute commissions that monitor local operators and communicate information to the public. I show that this policy has a significant positive effect on the transparency of plant managers, but no significant effect on safety care. I discuss how this policy could be fine-tuned to increase the amount of significant information regarding nuclear safety that is retrieved by the safety regulator.

In chapter 5, I propose a game-theoretic analysis of ex post disaster management. I propose a cheap-talk model that captures the incentives for an informed government or safety regulator to distort the information communicated to a population at risk in order to avoid panic, or to foster pro-social behaviours. This model sheds light on disaster communication strategies and on their credibility. I also study how mitigation policies interact with the ability of governments to convey credible information to the population. Multiple policy implications are derived.

The following paragraphs describe in more details the specific contributions of each chapter of this thesis.

1.3.2 Content of the thesis

Model uncertainty and the social cost of taking the nuclear lottery

In chapter 2, I question how the existing assessments of the risks of major nuclear accidents can be used to provide accurate guidelines for decision-makers faced with choices involving the use of nuclear power stations. This question was revived after the Fukushima-Daiichi accident, when it appeared that most existing assessments of both the probabilities and damage associated with major nuclear accidents were unable to provide a satisfactory range of values to guide future energy policy-making, or private investment decisions related to nuclear power technologies. Existing reviews of nuclear accident costs and probabilities of occurrence (see e.g. D'Haeseleer (2013) and Matsuo (2016)) show that the results obtained by industry and regulators using probabilistic safety assessments are significantly different from the results obtained by performing statistical analyses of past events.

The existence of conflicting assessments is not in itself very puzzling. Some experts calculate legally-driven, *ex post* assessments of the consequences of past nuclear accidents, in order to determine appropriate compensations for the victims of the accident. Others compute economically-driven, *ex ante* assessments, which aim to provide decision-makers with accurate information on which to base robust decisions. Thus, experts may rely on different data, account for various ranges of physical consequences, and the valuation of these consequences may be driven by various hypotheses.

Nevertheless, these conflicts question how good decisions can be made under such uncertainties. For instance, on which assessment should a policy-maker rely to motivate his decisions regarding the use of different electricity generation technologies? or for an investor to choose whether to participate to a nuclear power station project? In both of these examples, the conclusions drawn after relying on a given assessment of the nuclear risk will not be robust to a change in expert reference, as there can be up to six orders of magnitude between the smallest and largest assessments of nuclear accident damage or probabilities.

To shed light on this question, I use a Bayesian decision criterion developed by Klibanoff et al. (2005) and propose a framework for the calculation of the expected cost of nuclear accidents in which individual attitudes toward both risk and model uncertainty are accounted for. In this framework, model uncertainty is classically defined as the existence of several stochastic models describing the likelihood of occurrence of major nuclear disasters. I consider two main models: probabilistic risk assessments and statistical analyses of past events.

This chapter generalizes the framework developed by Eeckhoudt et al. (2000) in order to include individual attitudes towards risk in the assessment of the expected social cost of nuclear accident. I extend Eeckhoudt's work by adding one layer in the analysis: the existence of multiple models describing how probable a major nuclear accident can be. I also discuss how my framework can be reduced to Eeckhoudt's, as well as some of the implication of using this framework as a social welfare function.

Using results from experimental studies and recent damage assessments, I propose a simple numerical application of this framework, which yields an expected cost of 1.8 €/MWh. This number is consistent with recent estimations, and remains small when compared with the levelized cost of electricity (LCOE) of most modern electricity generation technologies. This numerical application contributes to the literature dedicated to the assessments of the cost of nuclear accidents by providing a figure which accounts for and combines the multiplicity of information sources available regarding nuclear accident frequencies.

For instance, to derive their estimations of the expected cost of nuclear accidents, Rabl and Rabl (2013) choose to rely on the past observed frequency of nuclear disasters, while the IRSN (2013) study relies on probabilistic safety assessments. Our assessment is more general, as it provides an assessment of the expected social cost of nuclear accidents as a function of a prior belief over available models.

Another takeaway from the numerical application is that aversion to model uncertainty has a limited impact on the cost of the accident. I show that the expected social cost of nuclear accidents can be decomposed in two distinct components: a compound-expected cost and an uncertainty premium. The compound expected cost is the expected value of the damage caused by the accident measured by a compound probability distributions, defined as the combination of possible models, weighted by a prior belief. The uncertainty premium is associated with the attitude of individuals toward the existence of model uncertainty regarding the chances of occurrence of nuclear disasters. When all models associate major disasters with very small probabilities of occurrence, the uncertainty premium is small when compared to the compound expected cost.

Finally, this chapter participates to a growing number of applications of the literature on decision-making under ambiguity and model uncertainty to uncertain decision problems, such as climate change mitigation, robust control, financial puzzles, or food safety regulation. To the best of our knowledge, this chapter is the first application of this literature to the case of nuclear power accidents.

Are older reactors less safe?

In chapter 3, I use a novel dataset that contains all the reports of significant safety events reported in all French reactors. Using these reported events, I investigate the variations of safety over time and across reactors of different ages and technological designs.

The question of the existence of a relation between the age of nuclear reactors and their safety levels has gained a lot of visibility in Europe since Belgium, Germany and Switzerland announced their wish to phase-out of nuclear power after the Fukushima-Daiichi accident, while other countries such as France, the United-Kingdom, the United-States, or China, are considering nuclear policies that would either entail new reactors with long lifespans (for China and the UK), or the extension of the lifespan of existing nuclear power stations (for France and the United-States).

This paper relates to the literature on the empirical analysis of safety performance based on voluntary reported data. One paper close to this chapter is Rose (1990), who study the relation between commercial airline safety and economic incentives based on reports of flight incidents reported by pilots. The main identification issue in both Rose's and the present empirical studies is the fact that the observation of a change in reporting behaviours can be due to changes in event frequencies, in the ability of plant managers to detect events, or in their level of compliance with reporting guidelines.

To overcome this identification issue, I develop an identification issue close in spirit with the work of Rose (1990) and Hausman (2014). Using a subcategory of events characterized by perfect detection and declaration, I first observe a general decrease over time of the frequency of these events. This suggests an overall increase in nuclear safety, as measured by these events. Subsequently, I observe that age has a positive and significant effect on annual counts of events reported in French reactors. This effect could be due to numerous causes, ranging from technical rationales, such as the deterioration of material equipments over time, to organizational ones, such as a selective matching of teams of better quality to the youngest reactors of the fleet.

An additional result is that the tendencies observed on the subset of perfectly detected and declared events are not observable when considering all significant safety events. This may be due to a selection bias: operators may exert specific efforts dedicated to the types of events I chose to study. This rationale can be mitigated by the observation that two distinct types of perfectly detected and declared events were used to perform our analysis and yield very similar results. Another possible explanation is that the reporting process is not fully transparent, as plant managers may fail to detect some safety events, or deem them not significant enough to require reporting.

The effect of local monitoring on safety care and compliance

In chapter 4, I estimate the causal effect of a French policy that sets a mandatory local monitoring scheme for nuclear plant managers. Since 2006, in each department hosting a nuclear power plant, a commission composed of competent citizens and elected officials meets with the plant management several times a year, and communicates to the local population the main information obtained regarding the safety and operation of their local nuclear station.

In this chapter, I overcome two identification issues. The first one is the endogeneity of monitoring intensity and reporting behaviours described in chapter 3. Indeed, the compliance of plant managers with reporting guidelines may be affected by the intensity of the monitoring performed by local commissions, but local commissions may also exert a more careful monitoring when they expect their operator to enforce poor levels of safety and compliance in his power station. In addition, an omitted variable bias adds to the endogeneity of monitoring intensity, as I do not measure the stringency of the national safety regulator, which may be correlated with the intensity of the monitoring performed by local commissions.

The second identification issue faced in this chapter is a channel identification issue. Indeed, if the counts of events reported change significantly when the intensity of their local monitoring commission varies, this change can be attributed to several channels: the safety of the station, the ability of plant managers to detect events and their compliance with declaration guidelines can all explain variations in observed counts of events.

In this chapter, I disentangle these channels by first presenting a principal-agent model that sheds light on the incentives a plant manager has to exert safety care and to comply with declarations guidelines. This model is adapted from Evans et al. (2009) and Gilpatric et al. (2011), and differs slightly from these original models by allowing both the firm and the regulator not to detect significant events. I derive from this model testable predictions which I then test in the data.

It arises from the empirical estimations that plant managers do not seem to improve safety care due to increased monitoring intensity. On the other hand, increased monitoring intensity leads plant managers to significantly increase their transparency, defined as the combination of their ability to detect safety events and their compliance with safety declarations guidelines.

This paper first relates to the literature on the effect of economic incentives on nuclear safety and on compliance with regulatory requirements. The research question tackled is closely related to the study of Feinstein (1989). This chapter differs from Feinstein (1989) in the sense that I do not have data on detected violations of safety guidelines, which requires us to come up with our identification strategy. Our empirical estimations are close in spirit from the estimations carried out in Davis and Wolfram (2012) and Hausman (2014), which use similar data from U.S. power stations to assess the effect of the deregulation of some nuclear power stations on their economic performance and safety levels. Contrarily to these two studies, I account for potential non-compliance in the declaration process, as well as for failures to detect some events by the firm and by the regulator.

This chapter also participates to the literature on the enforcement of and compliance with environmental regulations under a voluntary reporting scheme and an audit mechanism. First, this chapter is closely related to the work of Duflo et al. (2013) and Telle (2013), who use randomized controlled trials to assess the effectiveness of monitoring programs on self-reporting behaviours. Telle (2013) studies the effect of deterrence on self-reporting, and shows that whereas specific deterrence (e.g. fines) does increase self-reporting, an increased frequency of audit does not. Duflo et al. (2013) show that preventing conflicts of interests in audit mechanisms leads to less non-compliance and more mitigation efforts. Another related paper is Lin (2013), who assesses the effect of increases in the probability of being monitored on emissions reduction efforts and truthful self-reporting of emissions levels, using rainfalls as an instrument for the monitoring probability. Compared to these three papers, I investigate similar questions in a different industry, and use an instrumental variable different from the one used in Lin (2013), and instead of a randomized experiments. I contribute to this literature by introducing possible non-detection of events by the firm and imperfect audit results for the regulator.

Communicating disasters

Finally, in chapter 5, I tackle the question of disaster communication. More precisely, I propose a game-theoretic analysis of the communication strategies adopted by governments or regulators in the wake of major catastrophes, and when the actions that would be chosen by informed citizens differ from those the government would like them to take. Examples of such situations are numerous. Preventing bank runs during financial crises, preventing people from trampling on each other during fire evacuation procedures or advising people to stay home to avoid a nuclear fallout are instances during which the informed party tries to convince the population to take an action they may not have taken otherwise.

Yet, as the population may anticipate that an informed government may try to downplay the risks associated with a catastrophic situation to avoid panic reactions, it is interesting to wonder how credible strategies can be designed by governments or safety regulator. In other words, if the communication of an informed government is not credible, then the population will loose trust in the government, and may in turn take ill-advised actions.

To study this issue, I propose a cheap-talk model in which a sender informed of an incoming catastrophe can communicate in a non-committed way with heterogeneous receivers. These receivers are equally exposed to harm, but have different private costs of taking a protective action. In addition, the protective action entails a social cost, which can be either negative (e.g. panic), or positive (e.g. killing irradiated livestocks to prevent human exposure).

The equilibria of this game are structurally equivalent to the equilibria characterized by Crawford and Sobel (1982). I first use this structure to discuss the classical trade-off of the crisis communication literature (Pljansek et al., 2017) between the trust of a population in the recommendations made by its government and the opportunity for the government to derive short term benefits by downplaying risks. In particular, I confirm several findings of this literature. I show that considering the response of a population to the information disclosed is paramount to the design of credible communication strategies, that fostering transparency increases welfare, and that private communication improves the response of the population.

In addition, I use this model to study how the actions that can be taken by a government to prevent, prepare for or respond to a disaster can interact with this communication game. More specifically, I show that the mechanics associated with the negative and positive external effects cases are different. Preparedness actions that reduce the negative external effects of a population's response to a disaster - for instance through flood evacuation guidelines - tend to increase welfare while reducing the conflict of interest of the government, allowing a more transparent communication. On the contrary, fostering pro-social behaviours always increases welfare, but mechanically shifts away individual preferences from those of the government, thus hindering its credibility.

I also use this model to study the effect of the use of disaster relief funds on the incentives of the government and the population to respond to a catastrophe. I show that when the government can use transfers, and when protective actions are characterized by negative externalities, governments should provide partial public insurance to the individuals that do not engage in the protective action. On the other hand, when protective actions are charaterized by positive externalities, transfers should be used to subsidize the pro-social action. The first result goes against the usual results of the Samaritan dilemma literature which argues that public insurance may create an ex ante moral hazard problem by fostering careless behaviours from agents exposed to publicly insured risks (see e.g. Coate (1995); Kunreuther (1996); Epstein (1996); Kunreuther (2006); Kunreuther et al. (2013) or Teh (2017)). Our results show that partial public insurance can also provide positive incentives during the communication phase, as it allows the government to foster socially optimal actions by communicating in a more transparent way. In the case of negative externalities, I also show that the *state of natural catastrophe* is a simple policy instrument that allows to avoid squandering public resources during mild adverse events.

Part I:

Measuring Safety when Accidents are Rare

With models of increasing complexity we have shown the type of problems that moral hazard, limited liability, risk aversion and multiprincipals create in the regulation of firms inducing significant hazards for the public. [...] To the difficulties listed and studied we should add the further difficulties related to the usual fuzziness of the probabilities involved and the fact that their very low values may cast some doubt on the expected utility paradigm.

J.-J. Laffont, in Laffont (1995)

Résumé du chapitre 2

Ce chapitre présente une analyse du coût espéré des accidents nucléaires majeurs. En particulier, la méthode proposée permet de différencier la valeur actuarielle de l'accident de la propension des victimes à payer afin d'éviter d'avoir à faire face à la possibilité de cet accident. Pour ce faire, le chapitre utilise le critère de décision proposé par Klibanoff, Marinacci et Mukerji (2005) afin de modéliser des individus averses non seulement au risque, mais également à l'incertitude caractérisant les probabilités associées aux accidents nucléaires. Ces incertitudes sont dépeintes ici par l'existence de multiples modèles permettant de décrire le risque d'accidents nucléaires futurs. Le coût espéré de l'accident est ainsi défini comme la somme des équivalents certains individuels de loteries décrivant l'éventualité d'un accident nucléaire majeur. Une application numérique pour un cas français stylisé est proposée, et suggère un coût espéré de l'ordre de 3€/MWh. Le choix des différents modèles probabilistes et la prise en compte de cette source d'incertitudes jouent un rôle important dans la détermination de cette application numérique, contrairement au niveau d'aversion à l'incertitude des individus constituant la population.

CHAPTER 2

Model Uncertainty and the Social Cost of Facing the Nuclear Lottery

2.1 Introduction

When assessing the risk of large-scale nuclear disasters, two different sources of information regarding their probabilities of occurrence exist. On the one hand, probabilistic risk and reliability assessments (PRAs in the following) have been conducted by the nuclear industry and nuclear regulators since the WASH-1400 report produced by the U.S. Nuclear Regulatory Commission in 1975. These assessments are based on simulations and event trees, and aim to identify and correct safety weaknesses in the designs of nuclear reactors. They are still widely used today, by nuclear vendors, safety authorities, or policy-makers. This adoption of PRAs in the nuclear policy-making process has largely been described in the political-science literature (see e.g. Downer (2014) and references therein). Yet, the Fukushima-Daiichi accident raised concerns regarding the accuracy of the policy guidelines derived from their results (see e.g. Downer (2014) and Ramana (2011a,b)). In summary, these studies stress that PRAs overlook significant risk factors, such as human errors or beyond-design-basis[1] events.

On the other hand, numerous statistical analyses based on past nuclear events propose an alternative view on these events. Two recent reviews of the literature on the nuclear risk (e.g.

[0]We are particularly indebted to Professor Jean-Marc Tallon (Centre d'Économie de la Sorbonne) for his numerous comments and insights on this paper. We would also like to thank Professor Einar Hope (Norwegian School of Economics) for selecting an early version of this paper for short publication in *The Energy Forum - Bergen Special Issue*, July 2016. We would like to thank and all the participants to the seminars held in Cambridge University (EPRG), Kyoto University, Tokyo University, Waseda University, Tsinghua University, The Beijing Institute of Technology (CEEPR), Tianjin University, the Mines ParisTech Research seminar in energy economics, the 2016 IAEE conference held in the Norwegian School of Economics and the 2017 SBCA conference held in George Washington University for their very helpful remarks, questions and comments regarding this paper. We finally thank two anonymous reviewers for their helpful remarks and suggestions regarding a previous version of this paper. All errors, approximations or limitations are entirely ours.

[1]A beyond design-basis event is an extreme event whose consequences are not designed to be withstood by the plant. An example of such an event is the 20-meter wave that hit the Fukushima-Daiichi reactor, which design was only planned to withstand waves as high as 10 meter.

D'Haeseleer (2013) for the European Commission, and Matsuo (2016) for the Japanese Institute for Energy Economics) show that the results obtained by these statistical studies are in sharp contradiction with results obtained by PRAs. They show that four orders of magnitude may exist between the assessed expected frequencies of future nuclear catastrophes.

Despite this lack of precision in the determination of these future probabilities, governments have to choose whether to rely on nuclear energy, investors have to decide which technology to finance, and utilities have to determine when to shut down old nuclear plants, or where to locate new ones. To better inform these decisions, this chapter questions the way we assess the risks associated with the use of nuclear power, and the basis on which choices among these alternatives are made.

When facing multiple and conflicting information regarding a risky prospect, comparing expected costs or benefits based on either of these conflicting sources of information may seem like an *ad hoc* choice, rather than a rational ground for making sound decisions. Indeed, Ellsberg (1961) first showed that uncertainty[2], or the absence of knowledge regarding the probabilities of some events, had an effect, distinct from the effect of risk, on individual behaviours. Since then, a wide body of evidence has been accumulated on the aversion of individual decision-makers towards the fuzziness of the information describing the stochastic processes governing the outcomes their decisions may bring about. See e.g. Barham et al. (2014) or Berger and Bosetti (2016) for more recent experiments.

This chapter proposes a methodology for the assessment of the social cost of nuclear accidents, which accounts for the fact that these events cannot be properly described by a single probability distribution over monetary outcomes. To do so, we use the theoretical literature dedicated to decision-making under uncertainty. More precisely, scholars proposed various decision criteria that explicitly account for uncertainty, and for the attitude of individual decision-makers towards uncertainty. Examples of such criteria can be found in Schmeidler (1989); Gilboa and Schmeidler (1989); Bewley (2003); Epstein and Schneider (2003); Ghirardato et al. (2004) or Klibanoff et al. (2005, 2009). These criteria depart from Savage's subjective expected (SEU) utility framework, and allow for more general classes of preferences regarding risk and uncertainty. Using the smooth model of Klibanoff et al. (2005), we assess the expected social cost of nuclear power accidents, adjusted for individual attitudes towards risk and uncertainty. This expected social cost is defined as the certainty equivalent of a nuclear lottery characterized by

[2]In the present chapter, following the terminology defined by Knight (1921), risk will refer to situations that can be represented as lotteries associated with known probabilities. Uncertainty and ambiguity will be used equivalently throughout the chapter to refer to situations in which probabilities are vague or unknown.

model uncertainty - i.e. the existence of two distinct models regarding nuclear accident probabilities: PRAs and statistics based on past nuclear accidents. We then apply this method to the French case, based on the recent assessment of the damage caused by nuclear accidents performed by Rabl and Rabl (2013).

This chapter contributes to the literature dedicated to the assessment of the nuclear risk. First, by introducing model-uncertainty in our analysis, we further the efforts of Eeckhoudt et al. (2000), who proposed a method to account for risk-aversion in the assessment of the external cost of nuclear power, but based their analysis on the information provided by probabilistic risk assessments alone.[3] Second, our analysis can be compared to the recent paper of Rangel and Lévêque (2014), who proposed a Bayesian-revision framework to derive a subjective assessment of the probability of the next Fukushima-like nuclear accident. These probabilities are derived by constructing a theoretical prior based on PRAs, and updating it using the historical observations of nuclear accidents. This approach is set in a classical subjective expected-utility framework, in which all available information is aggregated into a single probability distribution. We differ on the interpretation of the nature of these two sources of information, as we consider PRAs and statistical evidence as two competing models describing future possible occurrences of nuclear accidents, and propose a Bayesian[4] decision-theoretic framework, which accounts for the attitude of individuals towards the fuzziness of the information available regarding the likelihood of nuclear accidents.[5]

Second, we contribute to the aforementioned literature dedicated to nuclear probabilistic risk and reliability assessments (PRAs), and to their use in the policy-making process. Our model provides an assessment of the risk associated with the use of nuclear power which is not only based on PRAs, but also on recent statistical analyses of past events. Thus, our figures acknowledge the existence of conflicting information regarding the risks of nuclear accidents, as well as the attitude of individuals regarding this uncertainty. Second, comparing our results to previous assessments of the nuclear risks shows that a significant part of the cost of nuclear

[3] Earlier references can be found in Gressman (1988) and Markandya (1995).

[4] A discussion of the Bayesian and non-Bayesian approaches to decision-making under uncertainty can be found in Gilboa (2004) and Marinacci (2015). The main three characteristics of Bayesian decision-making are that the probabilities associated with any state of the world are known, at least subjectively; that decision-makers use Bayes rule when they can; and that decision-makers make their decisions according to a decision rule that consist in maximizing an expected utility with respect to known probabilities. The smooth-model proposed by Klibanoff et al. (2005) is considered as a Bayesian model, in which two types of uncertainties are distinguished: physical uncertainty regarding the likelihood of occurrence of each state of the world and epistemic uncertainty regarding the adequate probabilistic model over the state space.

[5] Another branch of the literature dedicated to the analysis of the nuclear risk is based on the use of extended sets of past nuclear accidents, which include smaller events occurring at both power stations or fuel cycle facilities, in order to circumvent the limits associated with the analysis of extremely short statistical series. Hofert and Wüthrich (2011) or Wheatley et al. (2017), for instance, derived such estimations of the risk of nuclear accidents.

accidents was overlooked in past PRA-based decisions. Policy implications associated with the quantitative assessment performed in this study are also derived.

Finally, this chapter participates to a growing number of applications of recent advances in decision theory. In the finance literature, the theory of decision under uncertainty has been used to study asset pricing and portfolio selection (see e.g. Dow and Werlang (1992); Epstein and Wang (1994); Chateauneuf et al. (1996) or Epstein and Schneider (2008)). Hansen and Sargent (2001) developed applications of this theory for robust control in macroeconomics. Additional applications to the evaluation of climate policies have been proposed by Gonzalez (2008); Athanassoglou and Xepapadeas (2012); Lemoine and Traeger (2012); Millner et al. (2013), or Berger et al. (2016).[6] To the best of our knowledge, this chapter is the first attempt to apply this theoretical literature to the analysis of the risks associated with the use of nuclear power. Nevertheless, the framework we develop is similar to the frameworks used by Treich (2010) to study the effect of ambiguity aversion on the value of a statistical life, by Barham et al. (2014) to analyse the result of an experiment on farmers' behaviours aiming to elicit their attitude towards uncertainty, or by Alary et al. (2013) to study the effect of ambiguity-aversion on self-insurance and self-protection.

This chapter is organised as follows. Section 2.2 will present our theoretical framework for the calculation of the expected social cost of rare nuclear disasters. Section 2.3 will present our numerical application of this method to the French case of nuclear new builds. Section 2.4 will discuss some policy implications and conclude.

2.2 Model uncertainty and the social cost of nuclear accidents

2.2.1 Evidence of model uncertainty

This section aims to present some evidence regarding the relevance of the notion of model uncertainty in the analysis of nuclear safety. To do so, and as mentioned in the introduction, table 2.1 summarizes the results of two literature reviews conducted by D'Haeseleer (2013) for the European Commission and by Matsuo (2016) for the Japanese Institute for Energy Economics. These reviews were both performed after the Fukushima-Daiichi accident, to present a description of state-of-the-art knowledge regarding the analysis of the risks and costs of major

[6] Additional applications of theoretical decision criteria can be found in Paté-Cornell (1996) and Henry and Henry (2002), who advocated for the use of decision processes that acknowledge uncertainty and uncertainty-aversion in the study of epistemic risks, and in Gajdos et al. (2008) and Crès et al. (2011) who proposed methods for policy-makers to aggregate conflicting opinions of experts.

nuclear catastrophes, and to provide guidance for future investments in electricity generation technologies.

Both studies review the existing assessments of nuclear accident probabilities. We gather their results here and specify how these figures were derived. As we focus on the damage associated with large releases of radioactive materials in the environment, we focus on the studies reviewed that assess these probabilities. For instance, we omit the work of Hofert and Wüthrich (2011); Rangel and Lévêque (2014) and Wheatley et al. (2017) as they more generally tackle the issue of core-meltdowns rather than that of large releases of radioactivity outside the containment vessel of a nuclear reactor.

Probabilities of nuclear accidents are here expressed per reactor.year. For a 400-reactor fleet, these figures can be interpreted in terms of expected frequency of occurrences of events. For instance, a probability of occurrence of 10^{-4} per reactor.year is equivalent to witnessing one event every 25 years.

From the observation of table 2.1, it first appears that three to four orders of magnitude can separate the most optimistic estimations from the most pessimistic ones. Second, two main models describing the occurrence of nuclear accidents emerge from these reviews: probabilistic risk assessments and statistical analyses of past events. PRA-based results range between 10^{-6} and 10^{-7} accident per reactor.year, while statistical studies yield results in between 10^{-3} and 10^{-4} accident per reactor.year. Lévêque (2015b) or Downer (2014) explain these differences by noting that PRAs fail to account for human errors, regulatory capture or *beyond-design-basis* events[7], while statistical analyses of past accidents cannot account for local specificities and safety upgrades that are continuously implemented in nuclear stations.

These observations constitute the main motivation of the following of this chapter. Given the existence of these competing models, and given the range of values they provide, it is unclear which information should be relied upon to determine the costs and benefits associated with the use of nuclear energy.

The existence of multiple models providing large ranges of possible values regarding some important decision parameter has been tackled by Millner et al. (2013) in the case of climate sensitivity, where a survey of the opinions of multiple scientists shows large discrepancies. Similarly, we argue that the existence of conflicting models ought to be taken into account when

[7] A beyond-design-basis event is an event that has not been planned for during the design of the plant, and whose likelihood and potential consequences are not accounted for in risk assessments. The simultaneous flood of the emergency coolant system and shut-down of the national electricity network which led to the Fukushima-Daiichi accident was a beyond-design-basis event.

Table 2.1: Various assessments of nuclear accident probabilities

Review	Study	Frequency of large releases	Method used
D'Haeseleer (2013)	NEA (2003)	1,9 10^{-6}	ExternE (i.e. PRAs)
	Rabl and Rabl (2013)	1 10^{-4}	1 accident every 25 year (Chernobyl-Fukushima)
	IRSN (2013)	1 10^{-5} to 1 10^{-6}	AIEA targets (based on PRAs)
	IER (2013)	1 10^{-7}	PRAs
	D'Haeseleer (2013)	1.7 10^{-5}	Bayesian update of PRAs using observations
	ExternE (Dreicer et al., 1995)	1 10^{-5}	PRAs
Matsuo (2016)	Japanese Cost Analysis Committee (2015)	1 10^{-5}	AIEA targets (based on PRAs)
		2.1 10^{-4}	3 accidents and world nuclear experience
		3.5 10^{-4}	5 accidents and world nuclear experience
		6.7 10^{-4}	1 accident and japanese experience
		2 10^{-3}	3 accidents and japanese experience
	Cour des Comptes (2014)	4.3 10^{-4}	1 accident in 40 years in the French fleet
	Eeckhoudt et al. (2000)	1 10^{-6}	ExternE (based on PRAs)

comparing energy production technologies. The following sections propose a methodology that does so.

2.2.2 A generalized framework for the study of the cost of nuclear accidents

The model

Consider N individuals living in a society facing the possible use of nuclear power. Let $\mathbb{S}$ be a measurable state space, and X be the set of outcomes. Canonically, we define lotteries as mappings from the state space into the space of outcomes. Let $\mathcal{L} = X^{\mathbb{S}}$ be the set of lotteries. In the following, l generally refers to an element of $\mathcal{L}$. Individual preferences among lotteries are assumed to be homogeneous across all individuals, and to be well represented by the smooth model of decision-making proposed by Klibanoff et al. (2005) (KMM in the following). According to this non-expected-utility criterion, individual preferences are no longer represented by a utility function and a subjective probability distribution over states of the world, but by a set of probability distributions, and by two functions that respectively capture the attitude of the decision-maker towards risk and ambiguity, where ambiguity captures the decision-maker's lack of knowledge regarding the probabilities describing the outcomes of his decision. This criterion allows to account for the extended body of evidence showing that people behave in ways that cannot be explained by classical expected-utility frameworks when facing ambiguous risks. The most well-known type of behaviours was first described by Ellsberg (1961) in his seminal urn experiments.

According to the KMM framework, individual preferences can be represented by a set M of probability distributions over $\mathcal{S}$, a probability distribution μ over M, and two functions u and ϕ respectively defined over X and $\mathbb{R}$. Then, for any individual i, and any two lotteries l_1 and l_2, lottery l_1 is strictly preferred to lottery l_2 if and only if $V_i(l_1) > V_i(l_2)$, where the functional V_i is defined by:

$$\forall l \in \mathcal{L},\ V_i(l) = \sum_{m \in M} \mu(m)\ \phi\left(\sum_{s \in S} m_s\ u(l(s))\right). \tag{2.1}$$

In the following, we assume that M, μ, u and ϕ are common to all individuals.

Each probability distribution m in M can be thought of as an objective[8] model representing the stochastic process governing the result of the lotteries. μ represents a common subjective belief regarding the plausibility of each model. u is a utility function that captures the attitude

[8] A model is said to be objective in the sense that it is known to everyone, and based on scientific evidence. Though, an objective model may not be accurate in describing the realization of some events, which is the source of model uncertainty.

of the decision maker with respect to risk, whereas ϕ captures his attitude regarding ambiguity, e.g. how the likelihood of obtaining each outcome of the lottery varies across the different possible models.

In the following, we interpret lotteries as describing uncertain prospects faced by individuals. In particular, lotteries can be used to describe the risks borne by citizens due to the use of nuclear power. For instance, if various states of the world describe various scenarios leading to nuclear accidents, then a lottery l_i can associate these states with the loss of wealth incurred by individual i due to the associated releases of radioactive materials. If each individual is assumed to hold initial wealth W, and if $l_i(s)$ describes the loss of wealth incurred in state s by individual i due to the operation of a nuclear reactor, then the ex-ante utility derived by i from this lottery can be noted:

$$V_i(l_i) = \sum_{m \in M} \mu(m)\ \phi\left(\sum_{s \in S} m_s\ u(W - l_i(s))\right) \tag{2.2}$$

Before defining more clearly the cost of a nuclear accident, we first define the certainty equivalent $C_{A,i}(l_i)$ of lottery l_i as the quantity that verifies $\phi(u(W - C_{A,i}(l_i))) = V_i(l_i)$. Equivalently, we have:

$$C_{A,i}(l_i) = W - u^{-1}\left[\phi^{-1}\left(\sum_{m \in M} \mu(m)\ \phi\left(\sum_{s \in S} m_s\ u(W - l_i(s))\right)\right)\right] \tag{2.3}$$

In other words, $C_{A,i}(l_i)$ is the maximum amount of money individual i would be ready to forego in order to avoid facing lottery l_i. This definition is identical to the willingness-to-pay for risk elimination defined in the literature on the effect of ambiguity and ambiguity aversion on the demand for self-insurance and self-protection (see e.g. Alary et al. (2013) and Berger (2015)).[9]

The adverse consequences of using nuclear power on a population can be multiple. Nuclear accidents may lead to the relocation of people living near the power station, it may also have a global impact on the economy which will affect every individuals. Therefore, in the most general case, a given use of nuclear power will lead each individual to face a different lottery describing the potential harm faced when a nuclear accident occurs. To capture this variability across individuals, we define a *nuclear lottery* as an N-tuple of lotteries $L = (l_i)_{1 \leq i \leq N} \in \mathcal{L}^N$, in which l_i is the lottery faced by individual i. Thus, a nuclear lottery can be thought of as a given policy involving nuclear power stations. For instance, when deciding where to locate a new nuclear power station, a public planner has to choose between a set of nuclear lotteries, in which each possible location can be associated with a nuclear lottery.

[9] It is also close in spirit to the equivalent certain baseline mortality risk defined by Treich (2010) as the pure risk that equates the utility derived by an individual from an ambiguous mortality risk.

Following Eeckhoudt et al. (2000), we define the expected social cost of facing a nuclear lottery L, $SC(L)$, as the sum of the certainty equivalents $C_{A,i}(l_i)$ defined in equation (2.3). In other words, we have:

$$\forall L \in \mathcal{L}^N, \; SC(L) = \sum_{i=1}^{N} C_{A,i}(l_i) \tag{2.4}$$

$SC(L)$ captures the sum of the individual willingness-to-pay to avoid the nuclear risk. This is an ex ante measure of the cost of facing the possibility of a future nuclear accident, that accounts for the fact that the sum of individual willingness to pay to avoid these events may exceed the sum of their monetary consequences.

The expected social cost of nuclear lotteries

By introducing the collection M and the common belief μ over models in M, our setting generalizes the setting of Eeckhoudt et al. (2000), in which individual preferences were represented using a classical expected-utility approach. To see this more clearly, and using the notation $v = \phi \circ u$, equation (2.1) can be rewritten:

$$\forall l_i \in \mathcal{L}, \; V_i(l_i) = \sum_{m \in M} \mu(m) v\left(W - C_{R,i,m}(l_i)\right) \tag{2.5}$$

$$\text{with } C_{R,i,m}(l_i) = W - u^{-1}\left(\sum_{s \in S} m_s \, u(W - l_i(s))\right) \tag{2.6}$$

In equation (2.5), the argument of function v, noted $W - C_{R,i,m}(l_i)$, can be interpreted as the certainty equivalent of lottery l_i for an expected-utility-maximizer who would hold model m as his subjective belief over the state space. Therefore, the interpretation of functions v and ϕ are different. First, v is defined over outcomes, while ϕ is defined over utility levels. Second, function v captures the aversion of individuals to the variations across models of the certainty equivalents of a given lottery, whereas ϕ captures more generally their aversion towards the variations across models of their expected utility. Canonically, v is said to capture aversion to model-uncertainty, whereas ϕ more generally captures aversion to ambiguity.

As a remark, one can notice that by summing the quantity $C_{R,i,m}(l_i)$ over individuals, we obtain exactly the definition of the cost of a nuclear accident proposed by Eeckhoudt et al. (2000), who suggest to calculate these certainty equivalents using a unique probabilistic model derived from PRAs.

Our framework generalizes this definition to the cases in which several models are available, as the unambiguous certainty equivalent $C_{A,i}(l_i)$ of lottery l_i can be rewritten as a function of the ambiguous certainty equivalents $C_{R,i,m}(l_i)$:

$$C_{A,i}(l_i) = W - v^{-1}\left[\sum_{m \in M} \mu(m) v\left(W - C_{R,i,m}(l_i)\right)\right] \tag{2.7}$$

Equation (2.7) generalizes the definition of the cost of a nuclear accident proposed by Eeckhoudt et al. (2000), as it accounts for both the physical uncertainty contained in each lottery l_i (i.e. the risk described by any given model), but also accounts for the epistemic uncertainty characterizing l_i (i.e. the existence of model uncertainty).

Our framework can be formally reduced to the framework of Eeckhoudt et al. (2000) in four cases, which have distinct interpretations. First, and perhaps most convincingly, if the support of μ in equation (2.1) is a singleton, then our framework is equivalent to a SEU framework à la Savage, in which decision-makers dogmatically believe one particular model to be true. This is a way to describe the original paper of Eeckhoudt et al. (2000), who only rely on PRAs in their description of the likelihood of nuclear accidents.

Second, if ϕ is the identity function, then utility $V_i(l_i)$ reduces to $\sum_s m^\star(s)u(l_i(s))$ with $m^\star$ denoting the compound probability distribution obtained by computing the weighted average of the various models in M. For each state s, we have $m^\star(s) = \sum_m \mu(m)m(s)$. In this case, our unambiguous certainty equivalent $C_{A,i}(l_i)$ formally reduces to the definition proposed by Eeckhoudt. Nevertheless, it uses the compound probability distribution $m^\star$, whereas Eeckhoudt's framework relied on PRAs to elicit the probabilities associated with the various individual outcomes of the nuclear lottery.

Third, if M were a singleton, then we would be back to a classical risky model à la Von Neumann-Morgenstern, in which there exists an objectively unique model describing the likelihood of each state of the world. In this case, as ϕ is increasing, individual preferences can be represented by the argument of function ϕ in equation (2.1). In this case, our notations match the original notations used in Eeckhoudt et al. (2000), but this reduction seems colloquial as the discrepancy between PRAs and statistical analyses of past events was acknowledged as early as after the Chernobyl accident (Downer, 2014).

Finally, it can be noted that if M was a collection of Dirac distributions over particular states of the world, we would be in a situation of epistemic uncertainty devoid of physical uncertainty (see e.g. Marinacci (2015) for the origin of this terminology). In this case, let's note $s(m)$ the support of each $m \in M$. Preferences would then be represented by $V_i = \sum_{m \in M} \mu(m)\phi \circ u(l(s(m)))$. This representation is identical to that of Eeckhoudt et al. (2000) once we identify their utility function with our functional $\phi \circ u$, and their probabilities of occurrence of nuclear events with our belief function $\mu(m)$. Although our model indeed reduces to Eeckhoudt's under this hypothesis, claiming that nuclear accidents are devoid of any physical uncertainty is not a convincing assumption, as the various models at our disposal have non-singleton support. Indeed, both statistical analyses and PRAs consider that the operation of a nuclear reactor can lead to several outcomes corresponding to various types of accidents.

Relation to other frameworks

In most other papers tackling the assessment of the cost of nuclear power accidents, the authors either try to provide an estimation of the total damage caused by the accidents, or to estimate an expected cost defined as the product of the accidents monetary consequences by their respective probabilities of occurrence. The former approach is not directly related to our question, as we take nuclear lotteries as exogenously determined. On the other hand, the latter approach, used for instance in Rabl and Rabl (2013), IRSN (2013), D'Haeseleer (2013) and Matsuo (2016), can be generalized within our framework.

Several sets of conditions are sufficient for our definition of $SC(L)$ to boil down to the classic expected sum of monetary consequences of the accident. These conditions nevertheless differ on the probabilities that are then associated with each outcome of the lottery. First, if X is a subset of $\mathbb{R}$, and u and ϕ are the identity function, then $C_{A,i}$ is equal to the expected sum of the consequences of the lottery l_i, where each outcome $l_i(s)$ is associated with its compound probability of occurrence $m^\star(s)$. Second, if u is the identity function and either the support of μ is a singleton or M is itself a singleton, then $C_{A,i}$ is equal to the expected sum of the $l_i(s)$, where each outcome is associated with the probability m_s defined by the single element in either M or the support of function μ.

Our definition of the cost of a nuclear accident is broader than the ones used in the literature, as it accommodates a larger set of possible behaviours. Indeed, as was discussed in the previous paragraphs, using Eeckhoudt's sum of risk-corrected certainty equivalents as a guideline for nuclear policy-making is equivalent to assuming that all individuals in the population dogmatically believe in PRAs, although this belief has been extensively criticized after the Fukushima-Daiichi accident. Likewise, using the sum of the expected monetary consequences of nuclear accidents as a guideline for nuclear policy-making would be equivalent to making the assumption that all individuals in the population are either risk-neutral and model-uncertainty-neutral, or that individuals are risk-neutral and that nuclear accidents are not characterized by model uncertainty, or that the whole population dogmatically believes in PRAs. As our framework does not require these assumptions, we believe it to be descriptively more accurate than the ones that preceded it.

The uncertainty premium

In their experimentation dedicated to the elicitation of ambiguity-averse behaviours in farmers' decisions to adopt new genetically modified seeds, Barham et al. (2014) define the uncertainty premium associated with any given prospect as the difference between the expected value of the prospect under the compound probability distribution $m^\star$ and its certainty equivalent. Following their definitions, the ambiguity premium associated with a lottery l_i would be $C_{A,i}(l_i) - \sum_s m^\star l_i(s)$.

We propose to break down this uncertainty premium in two parts. To do so, we first define the *compound expected cost* $C_{i,m^\star}(l_i)$ of lottery l_i as the sum of its monetary consequences according to the compound prior $m^\star$:

$$C_{i,m^\star}(l_i) = \sum_{s\in S} m^\star(s)\ l_i(s) \tag{2.8}$$

Then, we define the *compound certainty equivalent* of l_i as:

$$C_{R,i,m^\star}(l_i) = W - u^{-1}\left(\sum_{s\in S} m^\star(s) u(W - l_i(s))\right) \tag{2.9}$$

Then, following the definition of Barham et al. (2014), we define the individual uncertainty-premium P_i associated with lottery l_i as:

$$P_i(l_i) = P_{MU,i}(l_i) + P_{R,i}(l_i) = (C_{A,i}(l_i) - C_{R,i,m^\star}(l_i)) + (C_{R,i,m^\star}(l_i) - C_{i,m^\star}(l_i)). \tag{2.10}$$

$P_{R,i}$, the second term of the right-hand side of equation (2.10), captures the risk-premium associated with lottery l_i, when assessed by a classical expected-utility maximizers holding the compound distribution $m^\star$ as a prior over states of the world. To obtain the uncertainty premium P_i, one should add to $P_{R,i}$ another premium $P_{MU,i}$, which captures how the various models in M vary around the compound distribution $m^\star$. The interesting feature of this breakdown of the premium P_i is that $P_{R,i}$ only depends on the compound distribution, while the level of model uncertainty present in M is fully captured by $P_{MU,i}$. Summing P_i over individuals, we can define the same premiums P, P_{MU} and P_R at the level of society.

Finally, given the expression of the certainty equivalent $C_{A,i}(l_i)$ expressed in equation (2.7) and provided that u and ϕ are concave functions, e.g. if individuals exhibit both risk and uncertainty aversion, then we can argue that:

$$\forall l_i \in \mathcal{L},\ C_{i,m^\star}(l_i) \leq C_{R,i,m^\star}(l_i) \leq C_{A,i}(l_i) \tag{2.11}$$

From equation (2.11), we directly deduce that under risk and uncertainty aversion, all the premiums defined above will be positive. Moreover, one can simply show that, if we define increases in risk or uncertainty aversion as respective concave transformations of functions u and ϕ, then the compound-risk premia will be increasing in risk aversion, while uncertainty premiums will be increasing in uncertainty-aversion. Therefore, to a public decision-maker, the size of P_i can be understood as a measure of the importance of risk and uncertainty in individual preferences. In other words, the greater P, the more a decision based solely on the compound expected sum of the monetary consequences will neglect individual preferences towards risk and uncertainty.

2.2.3 Social welfare and normative scope of the cost of nuclear accidents

As noted by Eeckhoudt et al. (2000), our definition of the expected social cost of nuclear accidents corresponds to a compensatory approach, in which we acknowledge the gap between the sum of the monetary consequences of an accident, and the individual willingness-to-pay to avoid it. An interesting feature of this definition is that it boils down to traditional cost-benefit analysis (in the form of classical utilitarianism) when considering certain or risky prospects devoid of any uncertainty. This means that the cost of nuclear projects or nuclear policies measured using our definition of the expected social cost can be compared with the usual costs of competing technologies, as long as they are not characterized by model uncertainty.

When making such comparisons, an important question is whether this definition of the expected social cost of a nuclear accident is a normatively appealing welfare measure. In other words, it is not clear whether minimizing the expected social cost provided by our method is a sound objective for a policy-maker. To tackle this question, we focus on the choices a policy-maker would make according to our criterion, and compare them to other possible welfare measures.

Assume that a decision-maker (DM) faces a choice among several nuclear lotteries - for instance when deciding where to locate a new plant - and that the objective of this decision maker is to minimize the total willingness-to-pay of individuals to avoid facing nuclear accidents. In other words, this DM minimizes the expected social cost of nuclear accidents. Formally, this objective can be written:

$$\min_{L \in \mathcal{L}^N} \sum_i C_{A,i}(l_i) \tag{2.12}$$

Using the notations introduced in the former paragraphs, this program is equivalent to:

$$\max_{L \in \mathcal{L}^N} W_S(L) = \sum_i v^{-1}(V_i(l_i)) \tag{2.13}$$

In the following we refer to the objective $W_S(L)$ as the social choice function.

If individuals are assumed to be both risk averse and uncertainty averse, then v^{-1} is a convex, non-decreasing functional. Hence, the social choice function used by our DM will lead him to make decisions that are incoherent with *ex ante* egalitarian principles.[10] Indeed, imagine a society constituted of two individuals, Ann (A) and Bob (B), facing to nuclear lotteries L_1 and L_2, respectively associated with individual lotteries $(l_{1,i})$ and $(l_{2,i})$, $\forall i \in \{A, B\}$. Assume that L_2 is a mean-preserving spread of L_1 in terms of individual expected utilities V_i, e.g. that $V_A(l_{2,A}) < V_A(l_{1,A}) < V_B(l_{1,B}) < V_B(l_{2,B})$ and $V_A(l_{1,A}) + V_B(l_{1,B}) = V_A(l_{2,A}) + V_B(l_{2,B})$. By the convexity of v^{-1}, we have that $W(L_2) > W(L_1)$. This example shows that a decision-maker that would minimize the expected social cost of nuclear accidents would prefer, ex ante, to transfer the burden of nuclear accidents to a single agent or group of agents. This remark also holds for the social cost function defined by Eeckhoudt et al. (2000).[11]

The fact that minimizing the expected social cost of nuclear accidents fails to satisfy egalitarian principles is conform to the intuition that nuclear stations should be located in areas characterized by low population densities. However, this result also undermines the normative scope of our assessment, as minimizing the social cost of nuclear accidents could lead social-planners to choose unfair policies, for instance by bringing the costs associated with present decisions onto future generations.

[10] According to Diamond (1967), an ex-ante egalitarian social planner maximizes a social choice function that can be written, for instance, $\sum_i \psi(\mathbb{E}(u_i))$, in which $\mathbb{E}(u_i)$ is the expected utility of individual i and ψ is a concave function. Under this criterion, the social planner will exhibit aversion to ex-ante inequality, that is to variations of expected utilities across individuals. Adler and Sanchirico (2006) also define *ex post* egalitarianism as the action of a social planner that would maximize the expected value of a welfare function defined as $\sum_i \psi(u_i(s))$. As this criterion first aggregates individual utilities in each state and then computes the expectation of this state-wise welfare function, it cannot be compared with our criterion, which computes expectations at the individual levels before aggregating it into a welfare function.

[11] Minimizing the social cost of nuclear accidents defined by Eeckhoudt et al. (2000) would define a welfare function $W_E(L) = \sum_i u^{-1}\left(\sum_s m_{PRA}(s)u(W - l_i(s))\right)$, in which m_{PRA} is the probability distribution they consider (based on PRAs), and u^{-1} is a convex non-decreasing function, as individuals are assumed to be risk-averse.

In order to address the possible unequal allocation of resources caused by policies characterized by uncertain prospects, Fleurbaey and Zuber (2017) propose an aggregation criterion that accommodates ambiguity-averse preferences while satisfying inequality aversion. Assuming that all individuals are maxmin expected utility (MEU in the following) maximizers à la Gilboa and Schmeidler (1989), expected welfare is computed as the expected utility that would be derived by a virtual individual facing the worst possible outcome in every states of the world, characterized by the most risk-averse preferences present in the population, and under the worst-case prior considered among the joint beliefs of the whole population.

This egalitarian welfare criterion - referred to as W_{eq} in the following - can be adapted to our setting. Assuming that ϕ is such that individuals are maxmin-expected-utility maximizers, we have:

$$\forall L \in \mathcal{L}^N, \; W_{eq}(L) = \min_{m \in M} \left(\sum_{s \in \mathcal{S}} \min_{i \leq N} m_s u(W - l_i(s)) \right) \tag{2.14}$$

Hence, using our assumption that all individuals have the same beliefs and the same attitude towards risk (i.e. the same utility function), we can define the egalitarian certainty equivalent of any nuclear lottery L as the quantity $C_{eq}(L)$ that satisfies:

$$C_{eq}(L) = W - u^{-1} \left(\min_{m \in M} \sum_{s \in \mathcal{S}} m_s \min_{i \leq N} u(W - l_i(s)) \right) \tag{2.15}$$

The welfare measure defined in (2.14) is based on the computation of the utility derived by a virtual individual, rather than by aggregating the utilities derived by each individual in the society. In addition, when considering prospects characterized by no uncertainty, the welfare measure defined in (2.14) does not boil down to classical cost-benefit analysis. Therefore, if one wants to compare competing prospects using this equalitarian welfare measure, it is necessary to compute the certainty equivalents of each prospects using the rule presented above. In particular, when calculating the certainty equivalent C_{eq} of a particular nuclear prospect, this certainty equivalent cannot be compared with the usual estimations of the cost of alternative sources of energy, if these estimations are not also based on the estimation of the utility of the individual worse-off according to the worse prior in the whole population.

This discussion clarifies the role of accident cost estimations in the making of public policies. If egalitarian policy-makers ought not to base their nuclear policies on the expected social cost defined in the previous sections, acknowledging the existence of a gap between the expected cost and the willingness-to-pay of individuals not to adopt a technology could be a useful information.

In this sense, the cost defined in this framework is a descriptive indicator of individual willingness to avoid some technology, or to adopt others. A piece of anecdotal evidence is the case of the French "Superphenix" reactor, a fast-breeder reactor, which was shut down in 1996 after only ten years of operation due to intense social protests. This waste of public resources may have been avoided if policy-makers had been able to better assess the opposition that this project would later encounter.

2.3 A numerical application to the French case

In the following section, we propose to estimate the expected social cost of a nuclear accident in a new-build for the French case. To do so, we make several parametric assumptions for the tractability of the framework presented above. First, we elicit M on the basis of the existing literature on the assessment of the risk of nuclear disasters. Then, invoking the principle of insufficient reason (à la Millner et al. (2013)), we assume that μ is the uniform distribution over M. Finally, we specify two parametric forms for functions u and v, and use the damage estimation provided in the most recent academic evaluation of the cost of nuclear accidents, e.g. Rabl and Rabl (2013), and recent French demographic data to define a general nuclear lottery L. We then use these assumptions to derive an estimation of the expected social cost of nuclear accidents $SC(L)$, and of its egalitarian counterpart $SC_{eq}(L)$.

2.3.1 Elicitation of beliefs

In order to elicit the models constituting M, we rely on the variations in scientific assessments of the likelihood of nuclear accidents. Given this, two options are available: either consider that each study that ever produced an estimation of the probability of a nuclear event can constitute a model in M, or consider that the methodologies used to obtain these assessments should be used to derive the different models in M. The former is the option chosen by Millner et al. (2013) to study the impact of the uncertainty characterizing the climate sensitivity parameter on the optimal climate change mitigation efforts. The authors consider each experts' opinion as a different model describing the possible value of the climate sensitivity. We argue in favour of the latter, as different assessments of the nuclear risk that use the same methodology may differ for reasons that are not related to model uncertainty, but rather to the specifics of the question tackled. As an example, the PRA-based studies listed in table 2.1 on page 44 differ in their estimations of the probabilities of nuclear accidents. This is not particularly striking, as these studies focus on different reactor technologies and different countries subject to different regulations.

As PRAs and statistical analyses are the main two sources of information regarding nuclear safety, M contains only two models in this application. We rely on PRAs to establish a first model describing the probability of occurrence of a major nuclear accident, referred to as m_{PRA}. Then, we use the statistical studies presented above to derive a second model, referred to as m_{SA}.

The PRA-based model is elicited using the results of AREVA's PRA studies for the British nuclear safety authority and the EPR design specifications. According to this source, the target set by the British nuclear safety authority regarding the probability of an accident leading to more than 100 fatalities is 10^{-7} per reactor.year. According to this same source, the firm achieved an even higher result, as it claimed that the probability of a core-meltdown associated with more than 100 fatalities for the current design of the EPR is 6.10^{-8} per reactor.year (HSE, 2011). In order to provide a conservative assessment, we will use the British objective of 10^{-7} as the probability of a major nuclear according to the PRA model.

The second model is based on the approximation performed in Rabl and Rabl (2013): we use a probability of a major accident of 10^{-4} per reactor.year, which corresponds to observing one major accident every 25 year. This assessment is based on the observation of Chernobyl and Fukushima-Daiichi. A similar assessment is made in Matsuo (2016), in which the frequency of 2.1×10^{-4} per reactor.year is calculated based on the observation of three major accidents[12] and the total experience of the world fleet.

The elicitation of m_{SA} is based on a simple identification of the past frequency of nuclear accidents. The rationale for this identification is twofold. First, it is coherent with our aim to provide a method that allows social planners to combine the engineering knowledge captured in PRAs with the information learnt from nuclear accidents. In particular, this model captures the likelihood of occurrence of events that fail to be accounted for in PRAs, such as human failures or beyond-design-basis events. Second, although more sophisticated statistical analyses have been carried out (see e.g. Hofert and Wüthrich (2011) and Wheatley et al. (2017)), these studies are based on events of much lower magnitude and occurring not only in nuclear power stations, but also in nuclear enrichment or recycling facilities. These statistical analyses tackle

[12] These three accidents include the Three Mile Island event, the Chernobyl accident and the Fukushima-Daiichi accident. This calculation is based on the assumption that the Fukushima-Daiichi accident counts as one event, even though three reactors were destroyed. We dismiss the other estimations reported in table 2.1 in Matsuo (2016) as they only consider the Japanese nuclear operating experience.

a different question, e.g. the assessment of the risks of nuclear catastrophes over the whole industry. Likewise, we do not consider the results obtained by D'Haeseleer (2013) and Rangel and Lévêque (2014) as they use Bayesian methods to combine PRAs with observations of nuclear accidents.

2.3.2 Attitude towards risk and model-uncertainty

In order to capture the attitude of individuals regarding risk and model-uncertainty, and to derive a numerical assessment of the expected social costs of nuclear accidents, we parametrically specify functions u and v.

Following the empirical evidence proposed by Berger and Bosetti (2016), we propose to specify both functions u and v as constant relative risk aversion (CRRA) functions. We then use the respective coefficients of relative aversion to risk and model uncertainty experimentally derived by these authors, e.g. $r_u = \frac{-xu''}{u'} = 0.28$ and $r_v = \frac{-xv''}{v'} = 0.72$. To remain coherent with the empirical finding that people exhibit stronger aversion towards model uncertainty than towards risk, we study the sensitity of our results with respect to the values of r_u and r_v while keeping r_v larger than r_u.[13]

Formally, we can write:

$$\forall x \in \mathbb{R}_+,\ u(x) = \frac{x^{1-\beta}}{1-\beta} \text{ and } v(x) = \frac{x^{1-\eta}}{1-\eta}. \tag{2.16}$$

This parametric specification is in line with some of the related literature. Eeckhoudt et al. (2000) use a similar CRRA utility function in his study of the external cost of nuclear accidents, and uses a coefficient of relative risk aversion of 2. Treich (2010) uses the same utility function for a numerical application regarding the effect of ambiguity and ambiguity aversion on the value of a statistical life. In an application of their asset-pricing model, Ju and Miao (2012) use the same parametric specification of the utility and model-uncertainty functions. Moreover, when η is superior to β, our specification is consistent with the empirical finding that ϕ exhibits decreasing absolute ambiguity aversion (Berger and Bosetti, 2016).

Given this specification, the individual willingness to pay to avoid any lottery l_i is given by:

$$C_{A,i}(l_i) = W - \left[\sum_{m\in M} \mu(m) \left(\sum_s m(s)(W - l_i(s))^{1-\beta}\right)^{\frac{1-\eta}{1-\beta}}\right]^{\frac{1}{1-\eta}} \tag{2.17}$$

[13] Another way of using the experimental results of Berger and Bosetti (2016) would be to specify the function ϕ as a constant relative ambiguity aversion function (CRAA) with its constant relative aversion coefficient r_ϕ set at 0.53.

Table 2.2: Total costs of a nuclear accident according to Rabl and Rabl (2013)

Types of costs	Costs (in b€$_{2013}$)
Relocation costs	250
Agricultural costs	7,5
Cancer costs	19
Cost of clean up	30
Cost of lost power	18
Cost of lost reactors	30
Total cost	354,5

2.3.3 The nuclear lottery

As was presented in the first section above, a nuclear lottery is an N-tuple of individual lotteries, each of which describing the consequences faced by a given individual due to the likelihood of occurrence of nuclear accidents. For tractability, and in order to avoid the complexities of having to consider multiple reactors, we further assume that the nuclear lottery faced here consists in the potential consequences of operating one nuclear power reactor in a country like France. In the following, we assume that 65 million French live in a territory similar to France, but only endowed with one nuclear reactor.

Using the central damage estimation performed by Rabl and Rabl (2013), we account for the following six types of costs induced by a nuclear accident: relocation cots, agricultural costs, health costs, clean-up costs, the cost of lost electricity production, and the cost of lost reactors. The total cost of the nuclear accident considered is €354.5 billion. The details of each cost item are provided in table 2.2.

We separate the population in three distinct groups defined according to their exposure to the nuclear accident. All members of a given group face the same lottery. The first group is constituted of local residents that require relocation in case of a nuclear accident. The second group gathers the local residents that do not need relocation in the wake of the accident. Both the members of the first and second group bear the agricultural and health costs of the accidents. The third group is composed of individuals living far enough from the power station. The global economic consequences of the accident, e.g. the clean-up cost and the costs of lost reactor and lost electricity generation, are homogeneously sorted among the members of all three groups.

Among these costs, health costs are borne by locals, e.g. both individuals from the first and second group. Rabl and Rabl (2013) assume that a nuclear accident will lead to 10.000 deaths due to radio-induced cancers. In order to associate health costs with individuals from group 1 and 2, we assume that these deaths will occur homogeneously within groups 1 and 2. Thus, for each model m in M, if the probability of a nuclear accident is m_{acc}, then the associated probability of contracting cancer is given by $m_{cancer} = m_{acc} * \frac{1}{N_1+N_2}$ in which N_1 and N_2 refer to the respective population within groups 1 and 2. This way of distributing health costs across the population is very schematic, and neglects the complexity of the analysis of the health effects caused by the exposure to radioactivity. Yet, this source of uncertainty is beyond the scope of this chapter. Agricultural costs are distributed equally across members of groups 1 and 2.

Following Rabl and Rabl (2013), we assume that approximately 500.000 people will be relocated after a nuclear accident. This is approximately equivalent to saying that the population living within 50 kilometres from the plant will be relocated. Indeed, 11 million French people live within 50 kilometres of the 19 French nuclear stations, which roughly means that on average, 578.000 people live within 50 kilometres of each nuclear station. Next, we assume that all the population living within 100 kilometres of the nuclear power plant will be affected by the local consequences of the accident (health and agricultural costs). In France, approximately 40 million people live within 100 km of the 19 power station, which means that on average, 2.1 million people live within 100 km of each nuclear power station. Rounding this figure to 2 millions, and accounting for the members of group 1, we thus assume that group 2 counts 1.5 million individuals. Group 3 is constituted of the rest of the French population, e.g. 63 million individuals. These estimations were performed using publicly available French demographic data from INSEE.

On the other hand, clean-up costs and the costs associated with the loss of nuclear reactors and their future generation of electricity are assumed to be borne by all individuals. Indeed, these three cost items are not necessarily external costs, as they may be paid for by the nuclear operator, if they do not exceed its limited liability. As we aim to calculate an expected social cost, we will assume that the nuclear operator is state-owned, and that these costs will be financed by the entire society through either taxation or future sales of electricity. This is coherent with the French organization of the nuclear industry.[14] For all of these cost items, we use the valuation proposed by Rabl and Rabl (2013), which is summarized in detail in table 2.2.

[14] A caveat of this assumption is that these costs would be asymmetrically paid for by large and small electricity consumers, but as we have assumed that our society is constituted of homogeneous individuals, our assumption remains somehow coherent.

Table 2.3: Individual nuclear lotteries by population group

Group	**Population** (million)	**State of the world**	**Damage** (€)	m_{PRA}	m_{SA}
Group 1 Relocated	0,5	No accident	0	$1-1.10^{-7}$	$1-1.10^{-4}$
		Relocation	504950	$9{,}95.10^{-8}$	$9{,}95.10^{-5}$
		Relocation, cancer	2404950	$5{,}00.10^{-10}$	$5{,}00.10^{-7}$
Group 2 Other locals	1,5	No accident	0	$1-1.10^{-7}$	$1-1.10^{-4}$
		No cancer	4950	$9{,}95.10^{-8}$	$9{,}95.10^{-5}$
		Cancer	1904950	$5{,}00.10^{-10}$	$5{,}00.10^{-7}$
Group 3 Remote	63	No accident	0	$1-1.10^{-7}$	$1-1.10^{-4}$
		Accident	1200	$1{,}00.10^{-7}$	$1{,}00.10^{-4}$

2.3.4 Results

The expected social cost of nuclear accidents

We can now derive the various lotteries faced by individuals from groups 1 to 3. These lotteries and their associated probabilities for each probability distribution in M are summarized in table 2.3.

Then, using the definitions presented in section 2, we can first compute the expected costs associated with this nuclear lotteries. In order to present this cost in a manner coherent with the literature, we present it in euros per megaWatt-hours (€/MWh). To do so, we divide the results obtained from equation (2.8) by the total production of electricity expected from a new build reactor. This estimated production is rounded to 10 TWh per year, after assuming a nominal power of 1500 MW and a load factor of 75%. In the following, all costs are reported in €/MWh, using this normalizing factor.

Following equation (2.8), we find a compound expected cost of 1.35€/MWh. This compound expected cost is the exact mean of the two expected costs that would have been obtained by computing the expected sum of the monetary consequences of the lottery presented on table 2.3 using respectively model m_{PRA} and m_{SA}. According to model m_{SA}, the "worst-case" expected cost amounts to 2,70€/MWh. This figure is coherent with the findings of Rabl and Rabl (2013). Likewise, according to model m_{PRA}, the "best-case" expected cost amounts to $2.70.10^{-3}$ €/MWh.

Social costs and the uncertainty premium

Second, we turn to the calculation of the certainty equivalents of each individual lottery for group 1, 2 and 3. Results are gathered in table 2.3.4, in which we present both the social

Table 2.4: The expected social cost of facing the nuclear lottery.

r_u	r_v	Compound social cost (€/MWh)	Social cost (€/MWh)
0,28	0,72	1,800106	1,800108
2	5	1,98753	1,98754
3	8	2,12856	2,12858
4	10	2,30665	2,30669

cost of nuclear accident obtained using our framework, and the compound social cost of nuclear accident, obtained under ambiguity neutrality. Using equations (2.7) and (2.9), the expected social cost of the nuclear lottery is defined as the sum over individuals of the $C_{A,i}(l_i)$, whereas the compound social cost is defined as the sum of the $C_{R,i,m^\star}(l_i)$. The compound social cost matches the definition proposed by Eeckhoudt et al. (2000), but uses the compound distribution $m^\star$, rather than the PRA model.

We present our results for several values of the parameters r_u and r_v. For each result, we keep the ratio between r_u and r_v experimentally determined by Berger and Bosetti (2016), rounding both parameters to the nearest integer.

The main takeaway from table 2.3.4 is the numerical value obtained for the expected social cost of an accident. Our values vary between 1.8 and 2.3 €/MWh, which is larger but consistent with most of the recent estimations listed in D'Haeseleer (2013) and Matsuo (2016). This figure can be compared with the social costs of other means of production of electricity. For instance, when comparing our figures with those presented by Rabl and Rabl (2013) concerning conventional fossil fuels and wind technologies, it appears that the expected social cost of nuclear accidents remains lower than these other social costs.

Another takeaway from table 2.3.4 is that accounting for aversion to model uncertainty does not significantly modify the estimation of the expected social cost of a nuclear accident from the value calculated using Eeckhoudt et al. (2000), provided one uses the compound distribution $m^\star$ obtained by mixing probabilistic risk assessments with statistical analyses of past events. This result is somehow satisfying, as it provides an additional rationale for the method designed by Eeckhoudt and coauthors, but using the compound probability distribution $m^\star$ instead of the PRA-based model. Thus, we provide a simple heuristic for the calculation of the social cost of nuclear accidents: their cost should be calculated as if these accidents were classical risky lotteries, associated with a compound probability distribution capturing the various models that can be used to describe their potential consequences.

The egalitarian cost of nuclear accidents

Finally, using the definition of the egalitarian social cost of a nuclear lottery presented in equation (2.14), and noting that individuals from group 1 are the individuals worse-off, in each state of the world, the expression of $C_{eq}(L)$ boils down to:

$$C_{eq}(L) = W - u^{-1}\left(\sum_{s\in\mathcal{S}} m_{SA} u(W - l_1(s))\right) \tag{2.18}$$

which, by equation (2.3), is equal to the certainty equivalent of lottery l_1 for any individual in group 1, and when preferences presented in equation (2.1) are represented by MEU preferences.

In this case, and using a coefficient of relative risk aversion of 0.28, the egalitarian certainty equivalent of lottery L is as high as €52.3, which can be compared to the certainty equivalents of the same nuclear lottery for members of group 2 and 3, which respectively amount to €1.5 and €0.12.

For a policy-maker abiding by the criterion proposed by Fleurbaey and Zuber (2017) and presented in equation (2.14), the meaning of a €52.3 egalitarian certainty equivalent is the following. The policy-maker should be willing to require every individuals to forego up to €52.3 in order to avoid facing a nuclear accident. Equivalently, this would amount to spending up to €3.4 billion per reactor and per year in a country such as France to avoid facing nuclear accidents, or to an expected social cost of 340€/MWh.

This very high figure cannot be compared with the other figures derived in previous paragraphs and previous references, as this calculation aims to capture both the specificities associated with the uncertainties characterizing nuclear accidents, but also the very high inequalities that exists between the prospects faced by individuals in group 1, 2 and 3. Hence, this number can only be compared with costs obtained using a similar aggregation rule, i.e. by always considering the outcome borne by the individual worse-off in all states of the world. In addition, as our simple set up presents nuclear accidents as extremely non-egalitarian events, it is normal that this welfare criterion associates them with very high costs.

2.4 Discussion and policy implications

This chapter develops a method for the calculation of the expected-costs of rare and catastrophic nuclear accidents, that takes into account the uncertainty that characterizes their probabilities of occurrence, and the aversion of individual preferences regarding these uncertainties. To do so, our method proposes a theoretical framework adapted from the literature dedicated to decision-making under model-uncertainty, which generalizes the methods previously used to assess the cost of nuclear accidents. The expected-cost provided by this method is no longer the expected sum of the monetary valuations of the damage undergone by society in the aftermath of an accident, but the sum of the individual willingness-to-pay to avoid facing this type of lottery. This definition is relevant as it allows social-planners to account for individual preferences towards both risk and model-uncertainty. We apply this method in order to derive the expected-cost of nuclear accidents in France, in new-builds. This expected-cost is evaluated at approximately 2 €/MWh when accounting for relocation, agricultural and health costs at the local level; and for lost reactors, lost electricity production and clean-up costs at the level of the whole country.

We can now focus on the policy implications of these figures. First, from a methodological standpoint, the figures we derive are well suited to be used in order to compare different energy alternatives involving nuclear power. For instance, using our framework, one could compare the social acceptability of various future nuclear energy scenarios in countries that use this technology. Likewise, our results can be compared to the social costs of other energy-generation technologies, such as those derived by Rabl and Rabl (2013). It appears that when accounting for uncertainty-aversion, the social cost of nuclear power remains lower than the social cost of other conventional fuels.

Moreover, when comparing nuclear power with technologies that may also be subject to catastrophic risks, such as dams or fossil fuels, it could be interesting to use the method developed in this chapter to compare the expected social costs of nuclear accidents with the expected social costs associated with these other rare energy disasters. Global warming, oil spills or dam failures are good examples of such rare disasters, that may lead to an increase in the social cost of other conventional technologies when accounting for individual attitudes towards risk and uncertainties.

Likewise, when setting safety standards for firms subject to ambiguous risks, our method provides a tractable way to measure the marginal expected-costs and benefits of modifying existing safety standards, as long as these modifications can be associated with changes in the multiple probabilistic models characterizing accidents, or with their associated damage.

Our approach provides information regarding the individual willingness to pay to avoid facing the risk of a nuclear accident. We discuss the implications of considering this as a measure of welfare. We find that the social choices implied by this measure are consistently inegalitarian, which may at first be coherent with intuition, as catastrophic risks are usually located in areas characterized by low population densities. Nevertheless, this result casts doubts on the normative scope of the measurement of the cost of nuclear accidents, and suggests that social choices involving nuclear power may have to rely on other measurements of welfare.

Another key takeaway of this chapter is the necessity to combine technical engineering expertise with the available information derived from past failures and accidents when providing guidelines for policies characterized by rare but catastrophic outcomes. Indeed, the Fukushima-Daiichi and the Chernobyl accidents have shown that nuclear accidents can be caused by human errors or "beyond-design-basis" failures, such as major geological events. As these risks are not accounted for in PRAs, learning from these past failures is essential if we want to make good decisions for the future. This implies to include this alternative source of information when providing policy guidelines in the form of risk and cost assessments.

To conclude this chapter, it is possible to point out some shortcomings of this framework. First, the numerous parametric assumptions required to perform our numerical application may cast some doubt on the plausibility of our numerical results, and undermine even more their normative scope. Therefore, it may be important in the future to try and assess the extent of uncertainty-aversion shown by individuals when making energy-related decisions. Efforts in this direction have already been exerted by Barham et al. (2014), who showed that farmers exhibit ambiguity-averse behaviours when adopting new seed technologies. It could be interesting in the future to assess whether these uncertainty-averse preferences can also be observed in energy-related decisions.

Finally, in our framework, decisions are assessed using an indicator constructed based on information available *ex ante*. Hence, we do not consider the possibility of learning new relevant information in the future. With a dynamic setting in which information could evolve over time, as is for instance the case regarding nuclear waste management and storage technologies, one could argue that it might be better *ex ante* to choose energy alternatives that allow to adapt decisions after the acquisition of new information in the future. Such preferences have been studied by Kreps (1979), who modelled the behaviour of individuals characterized by preferences for flexibility: being uncertain about their future preferences, they would rather keep a combination of options than choose among them right away. Other dynamic frameworks have been proposed by Epstein and Schneider (2003) and Klibanoff et al. (2009). This refinement is left for future research.

Résumé du chapitre 3

Ce chapitre présente une première analyse statistique d'une nouvelle base de données contenant l'ensemble des événements significatifs de sûreté déclarés par les opérateurs des centrales nucléaires françaises entre 1997 et 2015. L'étude statistique vise à identifier l'effet de l'âge des réacteurs français sur leur sûreté, en utilisant ces événements significatifs comme indicateurs de sûreté. Deux problèmes d'identification sont écartés. D'abord, les déclarations des événements significatifs de sûreté étant volontaires, des variations dans ces déclarations pourraient à la fois traduire des variations dans la transparence des opérateurs. Ce facteur confondant est écarté en sélectionnant un sous-ensemble d'événements systématiquement détectés et déclarés par les opérateurs. Ensuite, afin d'écarter les facteurs confondants spécifiques à des cohortes de réacteurs ou à des périodes d'observations spécifiques, des effets fixes années et sites sont spécifiés dans les régressions proposées, et les hypothèses nécessaires à l'identification de l'effet de l'âge sur la sûreté sont clairement énoncées. Plusieurs tests de ces hypothèses sont proposés en annexes. Les résultats de cette analyse suggère que dans une majorité des centrales françaises, l'âge a un effet en U sur la sûreté : si la sûreté augmente significativement avec les premières années d'opération des réacteurs, elle décroît ensuite avec leur âge. L'analyse montre aussi une hétérogénéité forte des résultats entre les centrales, et l'ampleur des biais pouvant être causés par la non prise en compte des effets effets spécifiques aux périodes d'observations ou aux cohortes de réacteurs.

CHAPTER 3

Are Older Nuclear Reactors Less Safe? Evidence from Incident Reports in the French Fleet

3.1 Introduction

In this chapter, we empirically investigate the causal effect of the age of nuclear reactors on their safety, using new safety incidents data reported between 1997 and 2015 in the 58 French nuclear reactors.

The question of reactor ageing has been at the center of a political and scientific debate for nearly two decades[1], since U.S. reactors started to apply for and obtain operating licence renewals, extending their lifespans from 40 to 60 years. Opposing views of experts have recently been gathered by Tollefson (2016). Advocates of licence extensions argue that nuclear reactors do not age, as most of their parts are replaced during their lifespans. Sceptics reply that the uncertainties associated with surveillance, replacement and repair techniques could prevent the observation of some ageing patterns. Despite this high political and social importance, this question has not been answered in a satisfactory way by the existing literature. The available data on nuclear safety is often of limited amount or quality, since major safety accidents are rare. The existing literature has therefore relied on datasets composed of accidents from both nuclear reactors and fuel cycle facilities (Hirschberg et al., 2004; Sovacool, 2008; Hofert and Wüthrich, 2011; Wheatley et al., 2017). Due to the scarcity of these accidents, these datasets contained at most 216 observations, which is a crucial limitation for statistical inference.

In addition, the statistical estimation of causal effects on nuclear safety faces two methodological challenges. First, when the risks of catastrophic accidents are measured by safety indicators

[0]We thank the audience of the Young Energy Economist and Engineers seminar in Edinburgh University, as well as four anonymous reviewers for their remarks and comments. All remaining errors are entirely ours.

[1]See e.g. the following wide-audience media article published in 2009 by Scientific American or in 2015 by the MIT Technology Review.

based on information reported by regulated firms,[2] safety may be prone to measurement errors. For instance, safety is measured in this chapter using incident data declared by nuclear operators to the safety regulator. When there is a positive probability that these events remain undetected by operators, or when plant managers have incentives not to report some events[3], then a spurious effect in the data could bias the estimates due to unobservability of both detection abilities and compliance with declaration criteria.[4] In the following of this chapter, we refer to these potential measurement errors due to detection failures or non-compliant behaviours as *transparency* issues. An instance of this spurious effect was described by Rose (1990) within the airline industry, in which safety incidents are also used as a measure of safety.

The second identification challenge in the estimation of the relation between age and nuclear safety is the classical Age-Period-Cohort issue (APC in the following). This problem stems from the existence of a linear relation between the period of observation, the age of a reactor during this period, and the cohort of this reactor, defined for instance as its commissioning year. Due to this multiple collinearity, no statistical regression model featuring these three factors can be identified, and it is impossible to disentangle the effect of age from period-specific or cohort-specific effects, without making further assumptions. This identification problem and its potential solutions have been largely studied in the social science literature (see e.g. Bell and Jones (2013, 2014) or Keyes et al. (2008, 2010) for recent illustrations, or Glenn (2005) for a review).

This chapter contributes to the empirical literature on the relation between age and nuclear safety in several ways. First, we introduce a novel dataset which contains all nuclear safety incidents reported between 1997 and 2015 in French nuclear reactors. This dataset contains detailed information on more than 13000 so-called *significant safety incidents*. These events represent the most significant deviations from the general standards of operation of nuclear reactors.[5] These events are analysed on a case-by-case basis both by plant managers from EDF and by experts from the safety regulator, to identify organizational or technical weaknesses in

[2]As an example, the World Association of Nuclear Operators defines annual counts of automatic and manual reactor shut-downs as a key nuclear safety indicator.

[3]Incentives not to report can be conveyed by the fear of regulatory sanctions or of public backlashes. Anecdotal examples of this situation in the nuclear industry is the recent request (28th of September 2017) by the French nuclear safety authority to shut-down the Tricastin power station due to the declaration by EDF of a significant safety event, or the occupation of the French Fessenheim power plant in 2016 by Greenpeace after a German newspaper claimed that an incident had been understated by the French nuclear safety authority in 2014.

[4]The ability to detect an event might vary with the design of nuclear reactors, while the propensity of firms to comply with declaration criteria might vary over time with the evolution of the stringency of safety regulation.

[5]The French Nuclear Safety Authority (ASN) defines ten declaration criteria that characterize which events ought to be reported. These events are not deemed significant by the regulator on the basis of their real consequences of damage but rather on the basis of the information they carry regarding the management of plant safety, and on the basis of their causes and potential consequences.

nuclear stations and foster the exchange of best practices across the nuclear fleet. To the best of our knowledge, no statistical analysis of these events has ever been conducted. We exploit descriptive statistics of this dataset to document some important initial characteristics of the distribution of accidents. In particular, we show that the annual number of reports per reactor is unevenly distributed across reactors of different ages and technological designs.

Second, we use this dataset to estimate the effect of the age of reactors on nuclear safety. In order to avoid bias due to missing observations, we use a subset of events that exhibit perfect detection and declaration rates (PDD events in the following). The PDD subset contains two types of events: automatic shut-downs and so-called safeguard events. By considering PDD events, we ensure that observed variations in event frequencies reflect variations in safety rather than variations in transparency.[6] Next, we deal with the APC problem by noting that multiple reactors within power stations were commissioned during different years. Our major identification assumption is that, conditional on covariates and time dummies, unobserved heterogeneity is at the site level and do not vary over time. Therefore, including site dummies effectively accounts for unobserved heterogeneity. This assumption is related to the Hierarchical model of Yang and Land (2006), and is based on the observation that reactors within a site share not only technology and management, but are also exposed to identical infrastructure and geographical conditions. We challenge this assumption with a variety of robustness checks, which produce results consistent with those of the main specification.

Third, we analyse whether transparency affects the declaration of significant safety events. To do so, we compare the results obtained on the subset of perfectly detected and declared events with the results of a similar regression run on the whole set of significant safety events. We argue that if transparency has no effect on the declaration process, the results obtained on both sets ought to be similar. This is related to the approach in Rose (1990), who analysed the effect of a deregulation of the U.S. commercial airline market on airline safety.[7]

Our results indicate that safety evolves with the age of reactors consistently with the usual *bathtub* trend from the reliability literature (see e.g. Aarset (1987); Mudholkar and Srivastava (1993); Xie et al. (2002) or Chen et al. (2011)). This literature describes the bathtub trend as a three-step process composed of an initial increase in reliability due to learning effects, a

[6]This novel approach for the analysis of nuclear safety is related to the strategy used in Hausman (2014), who uses automatic shut-downs to evaluate the effect of economic incentives on nuclear safety in the U.S.

[7]Feinstein (1989) also carries out a study of reporting behaviours in U.S. power stations based on violation data. The detection-controlled estimator proposed in this study cannot be replicated here, as we do not have data on detected non-compliance. Indeed, Feinstein's model is based on a mixture model that requires "inspector-specific" data that would specifically affect detection of non-compliance situation to determine his coefficient β_2.

subsequent phase of constant reliability, and a final phase during which reliability decreases as the system wears out. In a majority of French nuclear stations, the effect of age identified in this chapter is non-linear and consistent with an initial decrease of occurrences of perfectly detected and declared events, followed by a subsequent increase in these occurrences. However, we also show that this effect of age is heterogeneous across reactors, and that measuring an average effect across the whole fleet (i.e. imposing homogeneous treatment effect) or across large cohorts of reactors leads to insignificant results. Thus, the erroneous assumption of homogeneity leads to a loss of important causal insights. Finally, we show that failing to account for cohort effects leads to a considerable negative bias.

The chapter is organized as follows. Section 3.2 describes the French declaration process and conducts a descriptive analysis of our dataset. Section 3.3 presents the identification strategy and the empirical specifications. Section 3.4 exposes our results and section 3.5 concludes.

3.2 Exposition: significant safety events

3.2.1 Institutional set-up and data

The French nuclear fleet is constituted of 58 pressurized water reactors (PWR), located in 19 nuclear power stations (referred to as sites in the following), and owned by a single utility (EDF). These reactors were built in separate phases from the late 1970s to the late 1990s. The technological design of reactors evolved over the construction of successive reactors. For instance, reactors differ in their nominal capacity, the nature of their fuel, or in their ability to perform load-following. French reactors can be split within three cohorts of reactors according to their electrical production capacity. Each of these cohorts is constituted of one or more sub-cohorts that capture more minor design features. These groups of reactors are summarized in table 3.1. The first column of this table lists the three capacity cohorts of reactors. The second column lists their respective sub-cohorts. The following columns describe the number of nuclear power stations and reactors that belong to each cohort, as well as their construction period. An important remark is that all reactors within a given power station share the same technological design (i.e. they share the same capacity and belong to the same sub-cohort).

In France, nuclear safety is regulated by the Nuclear Safety Authority (ASN in the following), who defines regulatory standards for the operation of nuclear reactors. In particular, regulatory standards include mandatory reporting criteria that define a list of particular situations, or events, that have to be reported to the safety authority, as they are deemed significant for

Table 3.1: The French fleet, by conception and nominal capacity levels.

Capacity	Conception	Power stations	Reactors	Construction
900 MW	CP0	2	6	1971-1979
	CP1	4	18	1974-1985
	CP2	3	10	1976-1988
1300 MW	P4	3	8	1977-1986
	P'4	5	12	1980-1992
1450 MW	N4	2	4	1984-2000

Note: The 900 MW (MegaWatt) cohort contains reactors of three different conceptions (CP0, CP1 and CP2 reactors). The 1300 MW cohort contains reactors of two different conceptions (P4 and P'4 reactors). The 1450 MW cohort contains only one conception cohort (N4). Construction phases span from the beginning of the construction of the first reactor and until the connection of the last one.

safety. These events are referred to in the following of the chapter as *significant safety events* . This terminology is also the one used by both EDF and the ASN. The aim of these reports is to aggregate information and share experience and best practices among reactors, and to detect generic defaults in the reactors' designs.

To comply with these reporting criteria, plant operators gather information, on a daily basis, on a broad set of situations which depart from normal operation. All events are analysed by plant managers. The subset of events that match the reporting criteria have to be reported to the safety authority. Upon the detection of a significant safety event, plant managers have two weeks to provide the regulator with a detailed summary of the event and an analysis of its causes and consequences. Plant managers also keep a record of all detected situations, regardless of there significance, in case of an inspection by the regulator.

An important feature of this declaration process is that plant managers analyse a large number of situations, and decide which ones should be reported to the safety authority. A common issue associated with these voluntary reporting processes is the existence of missing observations, in the sense of detection failures (if an event goes undetected by a plant manager) or reporting failures (if an event is not reported despite its significance).[8] In particular, reporting failures could occur for multiple reasons: an operator could misinterpret the significance of a situation and fail to report it, or deliberately decide not to report the event.[9]

The ASN enforces this reporting procedure through periodic and random inspections, during which inspectors get access to the events filed by plant managers as too insignificant to be

[8]This issue has been thoroughly documented, from an empirical and theoretical perspective, in other industries. See e.g. Gray and Shimshack (2011) or Shimshack (2014) for recent reviews of this literature.

[9]We stress the fact that we do not necessarily describe detection and reporting failure as *moral hazard*, as the failure to detect or report an event need not be malignant nor intentional. This point is discussed further in the identification section.

reported. If inspectors come across a file describing an event which should have been reported as significant for safety, the regulator can engage in punitive actions against the operator, such as lawsuits or temporary shut-downs. Although inspectors get access to all the situations considered by plant managers, situations leading to no declaration are too numerous for inspectors to review them all thoroughly. As a result, inspectors review only a subset of events detected by operators but not reported to the authority.

Our dataset, obtained from the ASN, contains all 13482 significant safety events reported from 1997 to 2015 in the French nuclear fleet.[10] Each event is characterized by a set of variables, describing the location and date of declaration of the event, the nature of the components, materials and systems of the reactor affected by the event, its level on the International Nuclear Event Scale[11] (INES), the reporting criterion associated with the event, the state of the reactor at the time of detection (e.g. production, refuelling or maintenance) and a description of its causes and consequences. Details regarding the deletion or multiple count of some particular types of significant safety events are provided in appendix B.1.

3.2.2 Descriptive evidence

Figure 3.1: Annual declarations of four categories of events in the French fleet

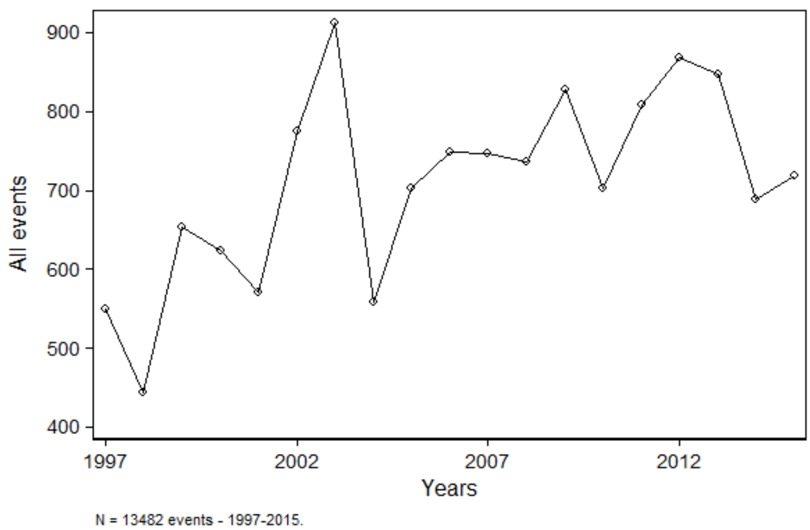


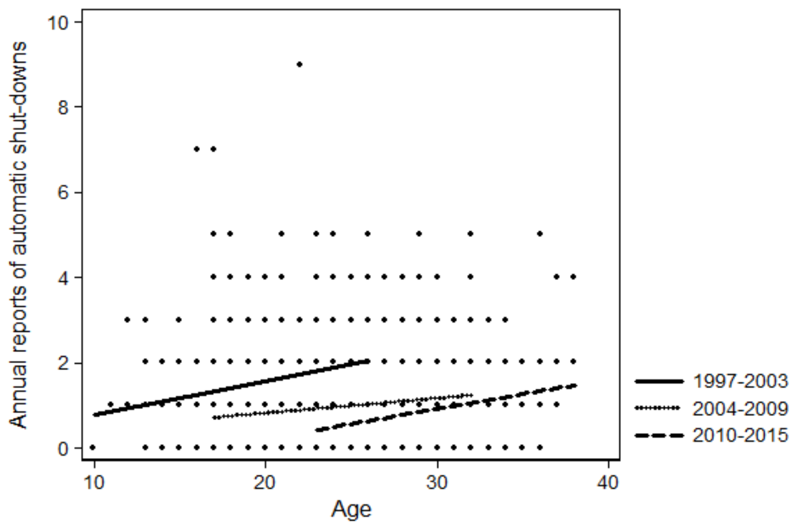
Figure 3.1 presents the evolution of the total number of reports over time in the French fleet from 1997 to 2015. On average, there are 700 events reported each year in the fleet, or one

[10] In other words, among all the situations detected by plant managers, our dataset is constituted of the events which were reported to the safety authority. Situations reported during the lifetime of permanently shut-down reactors are not included in our dataset.

[11] The International Nuclear Event Scale is a severity scale for nuclear events, defined by the International Atomic Energy Agency.

significant safety event per reactor and per month. Figure 3.1 shows that reports of significant safety events apparently follow an increasing time trend. This could be due to increases in the stringency of the safety standards, leading the regulator to consider new types of events as significant for safety. An increase in the ability of the operator to detect significant safety events, or a deterioration of nuclear safety with the ageing of the fleet, could also explain this increase. The distribution of reports of events across groups of reactors of similar age, technologies and location are presented and discussed in appendix B.2.

Figure 3.2: Reports of automatic shut-downs as a function of reactor age



As another piece of descriptive evidence that safety may vary with the age of nuclear reactors, we provide in figure 3.2 multiple linear fits of the reports of automatic shut-downs on the age of the reactors at the time of reporting. Each point on the graph corresponds to a given reactor-year, and each regression line is obtained by performing a simple linear regression of the annual counts of reports on the age of the reactor during the year observed. Three linear fits are presented, based on the reactor.years observed during three periods of roughly equal length: 1997-2003, 2004-2009 and 2010-2015. For instance, the solid line corresponds to the reactor.years observed between 1997 and 2003. This graphical representation technique is used for example by Keyes et al. (2010). Its purpose is to give initial evidence upon which plausible estimation constraints can be based. In our case, the division into three periods amounts to implicitly grouping reactors.years into three aggregated cohorts. The graph reveals two patterns. First, the regression fit of a more recent cohort lies below the regression fit of a more ancient cohort. This could be due, for instance, to learning effects. Second, within each cohort, the older the reactor, the more numerous the reports of automatic shut-downs.

3.3 Identification of the effect of age on safety

3.3.1 Endogeneity issues

Assume Y is a binary random variable such that $Y = 1$ when a safety event occurs. Our objective is to measure the evolutions of *safety*, which we define here as $\mathbb{P}\{Y = 1\}$. The rationale for this definition is twofold. First, it is consistent with the definition of safety as a probability of occurrence of events which may harm people or goods, which has become usual in the economics literature.[12] Second, even though the event $\{Y = 1\}$ is minor in terms of real consequences, reducing its probability of occurrence has a direct impact on the likelihood of major accidents, as major accidents are often composed of a combination of individually minor events.

Consider the model

$$\mathbb{P}\{Y = 1|W, C\} = \mathbb{E}\{Y|W, C\}, \tag{3.1}$$

where the random vector W contains observed factors of safety such as age and technology, and the random variable C is unobserved. Examples for factors in C are the unobserved (reactor-specific) ability to detect an event, or the propensity of an operator to report a significant event when detected. We are interested in the marginal effect of a covariate W_j (e.g. age) on the likelihood that an event occurs: $\frac{\partial \mathbb{E}\{Y|W,C\}}{\partial W_j}$. If, however, C is related to W, then the observed regression function $\mathbb{E}\{Y|W\}$ will capture this effect, which will lead to a bias, $\mathbb{E}\{Y|W\} - \mathbb{E}\{Y|W, C\} \neq 0$.

We are faced with three major endogeneity concerns. First, there might be a secular time trend in the overall nuclear safety. Potential drivers of such a trend are unobserved changes in regulatory standards, learning effects, as well as changes in the safety care exerted by plant managers. These changes might be correlated with both age and technological change. An instance of regulatory change is the evolution of the reporting criteria defining significant safety events. Whereas only three criteria existed before 1996, in later years the number of these criteria was increased to ten, directly influencing the quantity of information reported by managers to the safety authority. If all observations were pooled together and such regulatory changes ignored, different reactors could reveal different observed frequencies of events simply because the measurements were taken at different points in time. Thus, ignoring secular changes in the regulatory framework could lead to a spurious effect of age in the data.

Second, there might be other unobserved factors that are related to the age and technology

[12] See for instance Shavell (1984b); Hansson and Skogh (1987); Faure and Skogh (1992) or Laffont (1995).

of a nuclear reactor. An important category of such factors are the so-called cohort effects, which reflect the existence of specific common conditions characterizing the time of construction or commissioning of a set of reactors, such as a power station or a group of reactors sharing a common design.[13] Cohort effects may include internal conditions such as infrastructure or management culture within the particular set of reactors, as well as external conditions, such as geographical aspects or common exposure to regulatory inspections and norms at the time of the construction (see e.g. Glenn (2005) and Suzuki (2012) for a more in-depth definition of cohort effects). The same logic applies more broadly to any time-fixed reactor-specific effects that might be related to age and/or technology. Omitting cohort (or more generally, age-related individual) effects from the list of controls would bias the estimators of the independent variables.

Third and most importantly, the statistical analysis of safety is potentially hampered by measurement errors due to missing observations. There are two channels through which missing observations might occur: plant managers might fail to detect a safety event, or they might not report an observed event to the safety authority.[14] To formalize this, assume O is a binary random variable such that $O = 1$ if a safety event is observed, or detected, by the plant manager. $\mathbb{P}\{O = 1|Y = 1\}$ represents the plant manager's *ability to detect* events. Assume D is a binary random variable such that $D = 1$ if a safety event is reported - or declared - by the plant manager to the regulator. $\mathbb{P}\{D = 1|O = 1\}$ captures the plant manager's *propensity to declare* events. With this notation, the only data observable to the econometrician are the reports, which occur only when the realized state is $\{Y = 1, O = 1, D = 1\}$. Detection of events by plant managers might be related to management practices and technology, and incentives not to report events may vary across reactors. Thus, both reasons for missing observations are potentially related to age and technology, and may hence bias the estimates. Whereas the compliance channel is common to most industries subject to environmental regulation and self-reporting rules (the main example being the CO_2-emitting industries, such as the pulp and paper industry or the coal industry), the detection channel is mainly endemic to the complexity of the nuclear energy industry.

[13] We are thankful to an anonymous referee for pointing this out.

[14] Several reasons may lead to failures to report a significant event. First, there exist sanctions associated with the reporting of significant safety events. For instance, the ASN requested the shut-down of the Tricastin nuclear station on the 28th of september 2017 due to the report of a significant safety event. Subjective misconceptions of the reporting criteria are a second rationale for the non-reporting of an event. For instance, the following administrative subtlety of the reporting process is consistent with the existence of reporting failures. Managers actually report two types of safety events to the ASN: *significant* safety events and *interesting* safety events. The former constitute our source of data, and correspond to the reporting criteria presented above. The latter are much more numerous, and much less significant. Missing observation could occur if a *significant* safety event was reported (intentionally or not) as only *interesting* for safety. Such misreports can occur unintentionally, for instance if a plant manager underestimates the potential consequences of a particular situation.

3.3.2 Identification strategy

A standard approach to solve endogeneity problems 1 (secular time trend) and 2 (cohort or individual effects) is to add year dummy variables and individual fixed effects, respectively. Time dummies capture period-specific factors that are common for all reactors, such as unobserved changes in regulatory stringency and in safety efforts. Individual fixed effects, on the other hand, account for time-constant factors that are specific to a particular reactor or to a cohort of reactors. One major problem in our set-up, however, is that we are mainly interested in the effect of age on safety. Including time dummies, individual fixed effects and an age variable leads to perfect multicollinearity. To see this, note that the age of a reactor is defined in this chapter as the time elapsed between the year of observation and the date of the first nuclear activity of its core.[15] Then, the age variable is defined as a linear combination of year and reactor dummies.

This is a fundamental identification problem referred to as the Age-Period-Cohort (APC) identification problem by the literature, see Bell and Jones (2013, 2014, 2016) or Keyes et al. (2008, 2010).[16] In our case, using reactor fixed effects instead of a cohort variable leads to an equivalent problem, since the additional degree of freedom is sufficient to trigger a perfect multicollinearity in the data. There are several methods to tackle the APC problem developed in the literature, the most important ones being the constraint-based method by Mason et al. (1973), the Holford approach by Holford (1983, 1991), the median polish approach (Tukey, 1977; Selvin, 2004), and the hierarchical APC model by Yang and Land (2006).[17]

Our identification approach explores age variation of reactors *within* site cohorts of reactors. A site (or power station) consists of up to six reactors, which are built in a pre-specified order within the common geographic area defined by the perimeter of the power station. Our strategy amounts to assuming that all individual-specific time-fixed endogeneity (including cohort effects) can be captured with a site fixed-effect. This assumption is based on the observation that all reactors within a site share a very large number of characteristics. First and foremost, they share a common technological design. In addition, in most of the cases, reactors within a site were even built by the same set of subcontractors. Second, due to their geographical proximity, reactors within a site are exposed to the same climatic and seismic (and other geography related)

[15] Age could alternatively be measured with respect to other reference dates, such as the beginning of construction, the first connection to the grid, or the beginning of commercial operation. The date of the first divergence is chosen as it best captures the amount of radiations received by the reactor's different systems.

[16] Again, we thank to an anonymous referee for pointing out this literature to our knowledge.

[17] Exploring all of these methods is well beyond the scope of the chapter. Moreover, as the paper by Bell and Jones (2013) beautifully points out, the validity of the results produced by these methods hinges crucially on the theoretical foundation of their assumptions. Black box application of the methods may lead to arbitrary results with no causal interpretation.

conditions. Furthermore, for the same reasons, these reactors share common infrastructures, such as their cooling source. Third, reactors within a site share operational management and staff. At the same time, reactors within a site have some (even if not very large) variation in age, as they were typically not built and commissioned simultaneously. Thus, we can treat reactors within a site as a cohort and explore the age variation as a source of identification. Based on these considerations, we spell our first identification assumption:

Once controlled for observed site characteristics, time and site fixed effects, all endogenous variation is driven by transparency, i.e. by the ability to detect an event (O) and the manager's propensity to report safety events (D). Formally, we assume that

$$C = (O, D, \varepsilon) \quad \text{with}\, W \perp\!\!\!\perp \varepsilon, \tag{A1}$$

where ϵ is an idiosyncratic error term, and W contains now all observed factors, as well as time dummies and individual fixed effects.

Assumption A1 is closely related to the identification approach of Yang and Land (2006): the endogenous parts (i.e. cohort effects) are treated as common to a whole cohort, whereas the age is an individual variable. This allows to break the perfect multicollinearity between Age, Period and Cohort effects. As with most exogeneity assumptions, assumption A1 cannot be tested directly without a valid instrument. There are two main deviations from A1 that one has to worry about. First, the order of building the reactors within a site might have a long-term impact on the performance of the reactors. This would be the case if learning effects during building the first reactor significantly contributed to the safety of the reactors that were built later. Such an effect would be related to age and therefore violate assumption A1. The violation would lead to a negative bias in the estimate: older reactors would appear less safe solely due to age (instead of due to learning effects). We deal with this possibility by introducing a binary variables indicating whether a reactor was the first one to be built in a site, or the first of its particular design (see below the empirical specifications section). A second possibility of a violation is that there are reactor-specific time-varying factors which are related to age but not captured by the time dummies.[18] We deal with this possibility in a comprehensive way in section 3.4 and in appendix B.4.

We now turn to the crucial problem of imperfect observability of events (problem 3). First, given the description of the declaration process and its regulatory oversight, we assume that

[18] Although it is difficult to think of such factors, the simulation of Bell and Jones (2014) reveals the necessity to consider such a possibility.

type I errors cannot occur:

$$\mathbb{P}\{O = 1|Y = 0\} = \mathbb{P}\{D = 1|O = 0\} = 0. \tag{A2}$$

Intuitively, assumption (A2) means that the plant manager cannot detect an event which did not occur, nor can he declare an event which has not been previously detected.

Our identification strategy consists in finding a subsample in which detection and reporting failures are precluded. We define an event to be subject to *perfect detection and declaration* (PDD events in the following) if the following assumption is fulfilled:

$$\mathbb{P}\{O = 1|Y = 1\} = \mathbb{P}\{D = 1|O = 1\} = 1. \tag{A3}$$

According to the definition presented in (A3), an event is said to be perfectly detectable and declared if it is certain that an operator will observe the event conditionally on its occurrence, and declare it conditionally on its observation. We denote the subset of all PDD events as Θ_{PDD}. We obtain the following result:

Lemma 3.3.1. *Assume that assumptions* (A1), (A2) *and* (A3) *hold. Further, assume that the idiosyncratic noise* ε *has an expectation equal to* 0*. Then, for the model*

$$Y = g(W, C) = g(W, O, D) + \varepsilon \tag{3.2}$$

it holds

$$\mathbb{E}\{Y|W, C\} = \mathbb{E}\{Y|W\}. \tag{3.3}$$

The proof can be found in appendix A.1. The assumption $\mathbb{E}\{\epsilon\} = 0$ is a trivial assumption and can be achieved through a corresponding normalization of the model. Additivity of ε is implicitly assumed in the model presented in equation (3.2). This assumption is common for most econometric specifications used in empirical work.

Thus, according to lemma (3.3.1), the causal effect of age on safety can be uncovered from the observed data $\mathbb{E}\{Y|W\}$ for the subset of PDD events. We describe the set Θ_{PDD} in the following section. A discussion of external validity is provided in section 3.4.2.

Table 3.2: Descriptive statistics for three dependent count variables

Variable	Definition	Mean	Std. Dev.
ASD_{it}	Automatic shut-downs	1.122	1.233
SFG_{it}	Events requiring the use of safeguard systems	0.377	0.693
ALL_{it}	All events	12.256	5.105

ASN data. This table provides summary statistics of three count variables. Variable ALL presents the total number of significant safety events reported in year t in a reactor i. Variable ASD and SFG focus on specific types of significant safety events. ASD provides the specific number of automatic shut-downs. Variable SFG provides the counts per reactor and per year of safeguard events. These variables are defined over a panel of 58 French reactors and 19 years (1997-2015). As one reactor enters the panel in 1999, the panel contains 1100 observations.

3.3.3 Perfectly detected and declared events

For any particular type of safety incidents, there are two conditions that guarantee perfect detection and declaration (i.e. A3). First, events that have a direct effect on the electrical output of a power station cannot be undetected nor hidden as the Transportation System Operator[19] monitors the electric production of each power station. Second, events which are subject to particular auditing efforts during inspections led by the safety authority ought to be declared truthfully, as it can be argued that such events are (nearly) impossible to remain unnoticed, which eliminates the incentives for plant managers not to report them.

Two categories of safety events satisfy one of these conditions. First, automatic reactor shut-downs (or scrams in U.S. terminology, ASD in the following) stop the electrical production of nuclear reactors.[20] These events have also been used by Hausman (2014) as a proxy for nuclear safety. Second, events that require the unplanned use of safeguard systems (safeguard events or simply SFG in the following) are subject to specific auditing efforts during the inspection of the power stations by the regulator. The definition of these events[21] and their relative severity makes them easy and natural targets for the ASN inspectors during the routine inspections of nuclear stations.[22]

Table 3.2 and figure 3.3 provide descriptive statistics associated with the annual counts of automatic shut-downs and safeguard events. In particular, figure 3.3 shows that both automatic shut-downs and safeguard events are characterized, at the fleet level, by decreasing time trends.

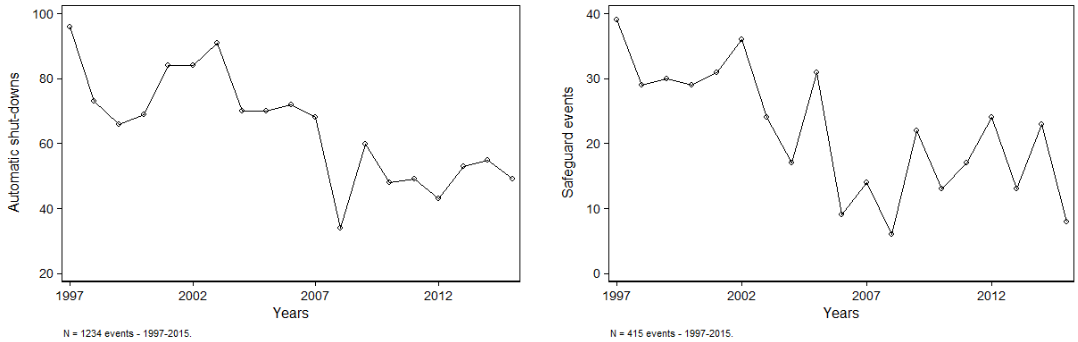
[19] In France, until 2000, the electricity transportation network was managed by EDF. Since 2000, transmission and production have been unbundled, and the transmission network has been handled by a single operator (RTE), which remains a subsidiary of EDF.

[20] Using automatic shut-downs as a proxy for nuclear safety is also supported by the fact that the annual number of automatic shut-downs is retained by the World Association of Nuclear Operators as one of their safety performance indicators. See for instance WANO's yearly performance reports on their website.

[21] Any situation leading to the unplanned use of a safeguard system is considered as significant.

[22] Interviews conducted with both the ASN and EDF seem to suggest that making the assumption that safeguard events are perfectly detected and declared is reasonable.

Figure 3.3: Historical occurrences of automatic shut-downs (left) and safeguard events (right) in the French fleet between 1997 and 2015.



3.4 Empirical investigation

3.4.1 Main results

Our empirical analysis consists of two main parts. First, we present results based on the assumption that age has a homogeneous effect on the safety in all nuclear reactors. To fix ideas, let Y_{it} denote the counts of safety events declared during year t in reactor i, $t \in \{1997, \ldots, 2015\}$, $i \in \{1, \ldots, 58\}$. Further, let AGE_{it} be the age of reactor i in year of observation t, and let X_{it} denote a set of reactor and year specific control variables, such as whether the reactor is a first-of-a-kind ($FOAK$ in the following) or a first-of-a-site ($FOAS$ in the following), and contains also a constant (intercept). We impose the following exponential specification of the conditional mean:

$$\mathbb{E}(Y_{it}|W_{it}) = \exp\left(\alpha_1 AGE_{it} + \alpha_2 AGE_{it}^2 + X_{it}\beta + \sum_{Year} \beta_{Year} \cdot \mathbb{1}_{Year} + \sum_{Site} \beta_{Site} \cdot \mathbb{1}_{Site}\right), \quad (3.4)$$

where exp denotes the exponential function and W_{it} denotes the regressors included in the model. Time dummies $\mathbb{1}_{Year}$ take the value of 1 when $t = Year$, and capture possible time trends or shocks associated with particular years, such as post-Fukushima-Daiichi safety upgrades. Site dummies $\mathbb{1}_{Site}$ take the value of 1 when reactor i belongs to $Site$. These fixed effects capture time-constant, site-specific unobserved sources of heterogeneity. Coefficients α_1 and α_2 measure the (potentially non-linear) impact of age on safety and are the main objects of interest. Note that the main independent variables, AGE and AGE^2, appear in equation (3.4) as an additive argument of the exponential function. As a result, their (multiplicative) effects on safety are not allowed to depend on the specific values of the remaining variables (homogeneous treatment effect). The reasons for choosing the specification presented in equation (3.4) over a standard

Table 3.3: Preliminary regression results: homogeneous non-linear treatment effects

VARIABLES	(OLS) ASD	(1) ASD	(2) SFG	(3) ALL
AGE	-0.0045	-0.0066	-0.037	0.013
	(0.081)	(0.073)	(0.067)	(0.019)
AGE^2	0.00091	0.0012	0.00019	0.000077
	(0.0012)	(0.00081)	(0.0017)	(0.00038)
First of a Site	-0.059	-0.063	-0.0052	0.033
	(0.12)	(0.11)	(0.13)	(0.040)
First of a Kind	0.086	0.068	-0.068	-0.080**
	(0.14)	(0.12)	(0.27)	(0.034)
Year FE	Y	Y	Y	Y
Site FE	Y	Y	Y	Y

ASN data. This table presents the result of four estimations of the effect of age on safety, using a panel of 1100 observed reactor.years across 58 reactors and 19 years (1997-2015). In all specifications, the age and squared-age variables are included, allowing for non-linear but homogeneous treatment effects. All specifications include site fixed effects and year dummies. Column (OLS) presents the result of an OLS estimation using automatic shut-downs (ASDs) as a dependent variables. Columns (1) to (3) present the results of a NB regression with quadratic over-dispersion. Column (1) uses counts of ASDs as the dependent variable. Column (2) uses counts of safeguard events as the dependent variable, to test whether the trends observed on ASDs can be generalized to other PDD events. Column (3) uses the counts of significant safety events as the dependent variable, to test whether the trends observed on PDD events can be generalized to non-PDD events. We provide site-clustered standard errors in parentheses. Intercepts are omitted. Significance: ***1%; **5%; *10%.

linear model are thoroughly discussed in appendix B.3. As a robustness check, however, we also estimate a linear fixed effects (FE) model.

The results are displayed in table 3.3. Column 2 contains results obtained under a linear FE specification on the subsample of ASD events (i.e. Y is defined as the count of ASD events), while columns 3 and 4 contain the results obtained using the exponential specification presented in equation (3.4) on the subsamples of ASD and SFG events, respectively. The results in columns 3 and 4 are produced with a Negative Binomial estimator (see appendix B.3 for a comprehensive discussion). In all three specifications, the estimates of the effect of AGE are negative and the estimates of the effect of AGE^2 are positive. This pattern is consistent with a bathtub effect, which we discuss in detail below. These estimates, however, are of small magnitude and insignificant in all three specifications. In addition, the order of commissioning of a reactor (FOAK and FOAS) is also shown to have no significant effect on safety.

The key implication of these estimates is the surprising result that ageing of a reactor has no detectable impact on the safety of a nuclear reactor. It defies the intuition that technical wearing reduces the reliability of components and thus also safety. The major issue with this analysis, however, is that neglected heterogeneity might lead to erroneous conclusions.[23] One

[23] Heterogeneity in treatment effects has received a lot of attention in the econometric literature in recent years,

obvious reason to allow for heterogeneity in our setup is the interaction between technology and ageing. In particular, technical differences in the design of reactors, as well different effectiveness of maintenance and replacement measures might lead to different effects of ageing on the safety of nuclear reactors. In addition, as pointed out in section 3.3, climatic and seismic conditions, as well as management organisation differ between reactors, and these differences could also interact with the effect of age.

A second part of our analysis is therefore to fully utilize the panel structure of our data and allow for heterogeneous effect of age on safety. Since all of the aforementioned reasons for a heterogeneous effect are on a site level (see also section 3.3.1 for a discussion), we implement a heterogeneous effect estimator by interacting the AGE and AGE^2 variables with the site FE. Thus, our main results are based on the specification

$$\mathbb{E}(Y_{it}|W_{it}) = \exp\Bigg(\beta \cdot X_{it} + \sum_{Year} \beta_{Year} \cdot \mathbb{1}_{Year} + \sum_{Site} \beta_{Site} \cdot \mathbb{1}_{Site} + \sum_{Site} \alpha_{AGE,site} \cdot \mathbb{1}_{Site} \times AGE_{it} + \sum_{Site} \alpha_{AGE,site} \cdot \mathbb{1}_{Site} \times AGE_{it}^2\Bigg). \quad (3.5)$$

The main difference with equation (3.4) is that we now do not have an additive AGE variable, but a sequence of interacted terms $AGE \times Site$. Since the sites constitute a partition of the full set, no separate additive variable AGE is included.[24]

The results are shown in table 3.4 which has an analogous structure as table (3.3). The majority of the estimates of the interacted AGE variable now are negative and significant. In addition, precisely in those cases, the AGE^2 interacted variable has a positive and significant coefficient, which is of smaller magnitude than the corresponding coefficient of AGE. More precisely, this effect is significant in 12 out of 19 cases for the SFG subsample, and in 10 out of 19 cases for the ASD subsample. In the SFG subsample, 16 out of 19 sites have their AGE coefficient negative and their AGE^2 coefficient positive.

Since the estimates in all three columns are of comparable sign and significance, we focus in our discussion on the OLS results, whose interpretation is easiest. Consider the Chooz site as a representative example. The coefficient of AGE is -0.29, which implies a reduction of -0.29 in the total number of events for each additional year. This effect is offset for high values of AGE by the effect of AGE^2, whose estimated effect is 0.014. Figure 3.4 shows the estimated pattern

see for example Heckman et al. (2006) and the references therein.

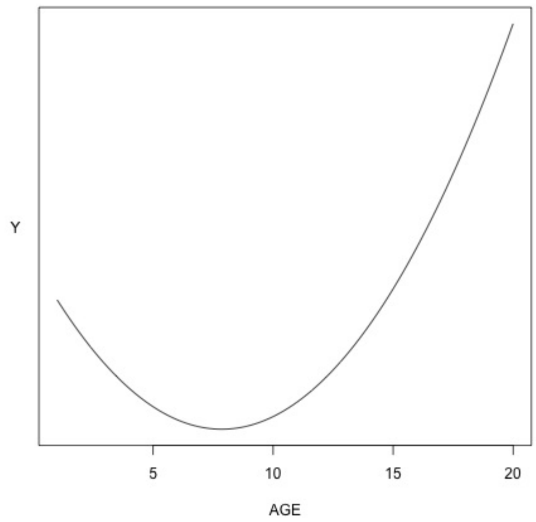
[24] Including AGE as a separate additive variable would lead to an equivalent specification. The coefficient of AGE in this case would simply measure the effect of the non-included dummy variable (the reference category).

Table 3.4: Main regression results: non-linear, heterogeneous treatment effects

VARIABLES	(OLS) ASD	(1) ASD	(2) SFG	(3) ALL
First of a Site	-0.046	-0.051	-0.044	0.028
First of a Kind	0.12	0.090	0.017	-0.069**
AGExBelleville	-0.35***	-0.37***	-0.36**	-0.045
AGExBlayais	0.11	0.10	-0.96***	0.048
AGExBugey	-0.12	-0.065	-0.79***	-0.23***
AGExCattenom	0.036	0.060	-0.0063	0.10***
AGExChinon	-0.12*	-0.14***	-0.22*	0.025
AGExChooz	-0.29***	-0.22***	-0.39***	0.021
AGExCivaux	-0.29***	-0.13***	0.19*	0.10***
AGExCruas	-0.086	-0.093	-0.71***	-0.15***
AGExDampierre	-0.47***	-0.50***	-0.35	-0.20***
AGExFessenheim	-1.16***	-0.70***	-0.24	0.45***
AGExFlamanville	-0.013	0.046	-0.58***	0.13**
AGExGolfech	-0.11**	-0.14***	-0.70***	-0.037
AGExGravelines	-0.18***	-0.20***	0.13	-0.12***
AGExNogent	-0.20**	-0.15*	-0.29	-0.046
AGExPaluel	0.33***	0.23***	-0.87***	-0.011
AGExPenly	0.052	0.051	0.32**	0.071**
AGExSt-Alban	0.12	0.086	-0.94***	-0.045
AGExSt-Laurent	0.054	0.0011	-1.66***	-0.080
AGExTricastin	-0.27**	-0.24**	-0.78***	-0.028
AGE^2xBelleville	0.0096***	0.010***	0.012**	0.0029**
AGE^2xBlayais	-0.0027	-0.0021	0.021***	-0.00081
AGE^2xBugey	0.0013	0.0012	0.013***	0.0047***
AGE^2xCattenom	0.00067	0.00048	0.000032	-0.0024***
AGE^2xChinon	0.0035***	0.0043***	0.0059**	0.000077
AGE^2xChooz	0.014***	0.011***	0.020***	-0.00063
AGE^2xCivaux	0.011***	0.0049**	-0.0045	-0.0031**
AGE^2xCruas	0.0033	0.0038	0.016***	0.0041***
AGE^2xDampierre	0.0096***	0.010***	0.0066	0.0038***
AGE^2xFessenheim	0.022***	0.013***	0.0072	-0.0068***
AGE^2xFlamanville	0.00037	-0.0012	0.016***	-0.0021
AGE^2xGolfech	0.0051***	0.0063***	0.025***	0.0013
AGE^2xGravelines	0.0044***	0.0050***	-0.0049*	0.0027***
AGE^2xNogent	0.0052**	0.0040	0.011**	0.0017
AGE^2xPaluel	-0.0087***	-0.0057**	0.021***	0.00045
AGE^2xPenly	0.00030	0.00081	-0.012**	-0.0021*
AGE^2xSt-Alban	-0.0023	-0.0011	0.024***	0.0020
AGE^2xSt-Laurent	-0.00098	0.00064	0.032***	0.0015
AGE^2xTricastin	0.0057**	0.0056**	0.016***	0.0012
Site FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

ASN data. This table presents the result of four estimations of the effect of age on safety, and using a panel of 1100 observed reactor.years across 58 reactors and 19 years (1997-2015). In all specifications, the age and squared-age variables have been interacted with site dummies to allow for non-linear and heterogeneous treatment effects. All specifications include site fixed effects and year dummies. Column (OLS) presents the result of an OLS estimation using automatic shut-downs (ASDs) as a dependent variables. Columns (1) to (3) present the results of a NB regression with quadratic over-dispersion. Column (1) uses counts of ASDs as the dependent variable. Column (2) uses counts of safeguard events as the dependent variable, to test whether the trends observed on ASDs can be generalized to other PDD events. Column (3) uses the counts of significant safety events as the dependent variable, to test whether the trends observed on PDD events can be generalized to non-PDD events. Standard-errors are clustered at the site level. Intercepts are omitted. Significance: ***1%; **5%; *10%.

Figure 3.4: Age and safety: a bathtub effect



of the causal effect for that reactor. On the x-axis, the AGE is measured with values between 0 and 20. This span approximately equals the time span of observation of the reactors of this station in our dataset. On the y axis, we measure safety as captured by Y. The plotted graph is generated by the function $\hat{\alpha}_1 AGE + \hat{\alpha}_2 AGE^2$ with $(\hat{\alpha}_1, \hat{\alpha}_2)$ here equal to $(-0.29, 0.014)$.[25] The number of accidents goes at first down with increasing time, driven by the coefficient of AGE. Roughly at the age of 7, the relationship becomes reversed and safety worsens with passing time. The initial level of safety is achieved at approximately age of 15 years. This pattern is compatible with the so-called "bathtub effect". The engineering reliability literature describes the bathtub effect as a three-state process, see e.g. (Chen et al., 2011). The first state is characterized by increases in reliability due to learning effects. It is followed by a steady state in which reliability remains relatively constant. In the final state, reliability decreases as the system wears out.[26] As our results show, the reactors with significant AGE variable all exhibit this non-linear pattern.

A further observation is that for the SFG (ASD) subsample, there are 3 (6) sites whose AGE effect estimates are positive, two of them being significant. Analogous observation with a reversed sign can be made for the AGE^2 variable.[27] These mixed results suggest an explanation for the insignificant bathtub pattern in table 3.3. Assuming homogeneity of the treatment effect

[25] The intercept and hence the values of the dependent variable Y are suppressed.

[26] The reliability literature identifies bathtub trends by estimating hazard rates and their variations across time (Aarset, 1987; Mudholkar and Srivastava, 1993; Xie et al., 2002).

[27] In the Penly power station, occurrences of safeguard events exhibit a *reverse bathtub* trend. Safeguard events in this station are characterized by a significant positive age trend and a significant negative quadratic age trend. Safeguard events do not exhibit this particular trend in this station.

essentially leads to "pooling" the above coefficients together, directly reducing the significance of the bathtub effect in the pooled results. This loss of information due to aggregation underlines how severe the consequences of neglecting heterogeneity can be. How much we lose from "pooling" (i.e. from imposing homogeneity) depends on the level of pooling. As a further illustration, table B.2 in appendix B.5 presents the regression results obtained with heterogeneity allowed at the capacity level. The capacity level partition is coarser than partitioning at the site level, but is finer than full aggregation (homogeneity). As the first and third specifications presented in table B.2 reveal, allowing for heterogeneity at the capacity level (i.e. interacting AGE and AGE^2 with capacity dummies) still yields insignificant results.

Next, we turn to our assumption that we solve the APC identification problem by defining a cohort at the site level. Assumption A1 consists of two components: we assume first that all unobserved heterogeneity is at the site level, and second that this unobserved heterogeneity is time constant. In appendices E and F, we perform additional robustness checks that challenge both parts of this assumption. We now describe these robustness checks.

First, in principle, allowing for reactor-specific fixed effects is desirable since there still can be some unobserved heterogeneity at the reactor level (an issue closely related to the preceding analysis of heterogeneity). However, as discussed in the preceding section, allowing for reactor-specific FE is not possible due to perfect multicollinearity. We tackle this problem in an alternative way: instead of using more coarse definition of cohorts (as in the main results), we now aggregate on the time scale. In particular, instead of adding time dummies for each year to the regression, we add three period dummies which respectively take the value of 1 when the year of observation belongs to the periods 1997-2003, 2004-2009, 2010-2015. This strategy allows us to include reactor-specific FE. The results are shown in table B.3 in appendix B.5. They are in line with our main results.

Second, even if unobserved heterogeneity is at the site level, it could nevertheless vary over time. In this situation, site fixed effects and year dummies would not be sufficient to capture this unobserved heterogeneity, and the estimated coefficient of the age variable could be biased. To challenge the time-constant nature of unobserved heterogeneity, we compare the results of the linear within estimation presented in table 3.4 with the results of a first difference (FD) estimation, formally described in appendix B.4. As the within and FD estimators are asymptotically equivalent under A1, one should not expect major differences in their results if A1 holds. On the contrary, if A1 failed to hold, omitting a time-varying first differenced variable in the FD approach could provide a different estimate than when omitting an averaged variable as in the within approach.[28] Results of the FD estimation are presented in table B.4, and are also in line with our main results.

Thus, all four strategies adopted in this chapter - descriptive graphical evidence, within estimation with cohorts at the site level, within estimation with aggregated time spans, first difference estimation with cohorts at the site level - document a negative effect of age on safety, at least after an initial period of time. The last three strategies provide a finer picture and point at the presence of a bathtub effect.

Finally, we focus on the estimates of the effect of $FOAS$ and $FOAK$. The literature on the costs of construction of the French nuclear reactors provides evidence that FOAS/FOAK reactors in general cost more. This suggests the presence of learning-by-doing effects in terms of lead-time and construction costs within sites and technological designs (see e.g. Berthélemy and Rangel (2015); Rangel and Lévêque (2015)). This is the main motivation for including these two variables into our regressions: if these learning-by-doing effects have an effect on the safety of the reactors, then our site cohorts assumption (assumption A1) would be otherwise violated. Surprisingly, the estimates of $FOAS$ and $FOAK$ are not significant. This finding is consistent throughout the various regressions in the main results and the robustness checks. One possible explanation is that although initial differences in safety across reactors within a power station may have existed shortly after the beginning of their operation, knowledge spillovers in management probably unified safety across reactors within sites. To verify this conjecture, we run a regression equivalent to the main specification but without site FE. The results are displayed in table B.2 in appendix B.5. The estimate of the coefficient of $FOAS$ in this regression becomes negative and statistically significant, thus supporting our conjecture. Moreover, this finding provides an additional evidence that all unobserved reactor-specific endogeneity is at the site level and therefore captured by site FE.

Remark. To round up our discussion on the APC problem, it would be useful to see in which direction the APC bias would be in our case. This would help learn more about the mechanics of the separate components. To do so, we run a regression with capacity FE instead of site FE. This exercise is closely related to our study of heterogeneity above. The results are shown in table B.2 in appendix B.5. In column (2), age is shown to have a significant linear effect on occurrences of safeguard events in all capacity cohorts, contrarily to the results of the estimation presented in column (1). These results suggest that the frequency of safeguard events increases when these reactors get older. In addition, nearly all year fixed effects become statistically

[28] A detailed description of this novel approach is provided in appendix B.4.

significant, which suggests an increase in safety over time, whereas year fixed effects were not significant when site fixed effects were included. Similar variations are observed when comparing the estimations presented in columns (3) and (4), when allowing for non-linear treatment effects. This is an illustration of the APC problem: omitting the cohort variable C (i.e. site fixed effects) can bias the results of the estimations of the effects of age A and period P on the dependent variable. In this case, the effect of age is biased upwards, while the coefficients associated with the time dummies are biased downwards.

It is hard to find a pattern that explains why some power stations do not exhibit a (significant) bathtub effect. For the ASD sample, the insignificant results can be found across power groups, conception groups and sizes, see table B.5 in appendix B.5. An exception is the P4 type, in which no bathtub effect is found on the ASD subsample. Although our tentative finding suggests more research on this topic, it is a limitation of this chapter not to uncover the precise technical reasons for this result.

3.4.2 Transparency

We now turn to the interpretation of the last columns of tables tables 3.3 and 3.4. In this specification, the dependent count variable is defined as the total number of events declared by nuclear reactors per year, without any restriction to specific types of events. The coefficients estimated in these two regressions largely differ (in both signs and statistical significance) from the coefficients obtained with the ASD and SFG subsamples. We now propose possible interpretations of this result.

Let the random variable T denote the type of an event, with $T = PDD$ being one of the possible types. Assume first that the effect of age on the probability of occurrence of an event is the same for the full sample and the restricted set Θ_{PDD}:

$$\frac{\partial \mathbb{E}\{Y|W,C\}}{\partial Age} = \frac{\partial \mathbb{E}\{Y|W,C,T=PDD\}}{\partial Age}. \tag{A4}$$

Assumption (A4) is an external validity assumption. Then, if transparency plays no role in the declaration process, we should observe similar results in the regressions run on the subsamples of ASDs or SFGs and in the regression run on the complete sample of significant safety events (ALL). This, however, is not the case. For instance, in table 3.4, seven plants which were subject to the bathtub effect in the regressions run on the ASD subsample are characterized in the last regression by no significant age trends or trends of opposite signs. A second example is the estimated coefficient of the $FOAK$ variable, which is negative and significant in the last regression in both table 3.3 and 3.4, whereas it was not statistically significant when considering only PDD events. Thus, either transparency does indeed bias the results, or assumption (A4) fails to hold.

However, assumption (A4) can be defended. First, comparison of regression (1) and (2) revealed no particular differences on the subsets of automatic shut-downs and safeguard events. Second, it can be argued that numerous investments in safety will have positive spillovers for all types of safety events. For instance, hiring skilled employees, investing in the training of safety engineers, or enhancing organizational practices are safety investments which will decrease the probabilities of occurrence of safety events, regardless of their types.

Provided assumption (A4) holds, our finding provides evidence that transparency can bias the analysis of the reports of nuclear safety significant events.[29] Thus, neglecting the unobserved changes in declaration criteria, detection abilities, and rate of non-compliance with declaration criteria could bias the estimation of safety variations with age and technology.[30]

3.5 Limitations and Implications

In this section, we discuss the limitations of our analysis and point out why we nevertheless believe that it gives important insights into nuclear safety. In addition, we derive some implications from our results.

The insights of our study are restricted in two ways. First, our results are not produced by using an experiment. Thus, even after including fixed effects (that control for time-fixed endogeneity) and time dummies (that deal with unobserved changes of the setting over time), there could be some remaining idiosyncratic heterogeneity that we fail to detect. On the other hand, however, the robustness checks of our study, and in particular the analyses made in tables B.3 and B.4, challenge this possibility. The consistency of our results throughout a variety of specifications give little sign for worrying. In addition, our results are consistent with those obtained by the standard reliability literature for a wide range of other technical systems, which can be interpreted as an additional evidence for their plausibility. In the absence of more detailed reactor-level data, our results are perhaps the most robust one can derive using

[29] We stress that the evidence of a transparency bias is not equivalent to the evidence of intentional non-compliance from operators. As was mentioned in the exposition and identification sections, transparency issues can occur due to detection failures or subjective divergence in the consideration of an event and of its relation to the reporting criteria.

[30] This result contrasts the results of Rose (1990), who discarded the importance of pilots' subjectivity in airline incidents reports.

a statistical methodology. Finally, building nuclear power plants are long term decisions that reflect a variety of strategic factors, of both economic and political nature. The hope to find a pure natural experiment that has a strong enough impact on this decision is therefore minimal. This can however be seen as an advantage, as it would mean that our fixed effects and time dummies should be sufficient to account for those long-term unobserved factors.

The second limitation is of a conceptual nature: our statistical analysis provides at best a black box picture of the causal effect of age on safety. Our results do not indicate the precise channels for this causal effect. Hence, it has to be emphasized that recommendations on the extensions of the lifetimes of nuclear power plants cannot be made without comprehensive (case-by-case) technical analysis that sheds light on the technical nature of nuclear safety.

There are nevertheless numerous aspects of our results that make them a valuable contribution to the nuclear safety literature. Our study contributes to a scientific and political debate about the role of reactor ageing in nuclear safety.[31] Since the majority of components of a nuclear reactor can be maintained and replaced, there is a widespread view that nuclear reactors do not really age mechanically.[32] The other side in this debate remains more sceptical and points out at the limitations of replacement and surveillance techniques.[33] In particular, not all components of a nuclear plant can be replaced with certainty: reactor pressure vessels and their concrete containment structures have never been replaced so far. Likewise, some life extension techniques, such as annealing[34], could enable the replacement of these parts, but have not yet proven their effectiveness. Finally, the durability of some inaccessible components, such as underground power cables, is hard to assess (Tollefson, 2016). The process of ageing of these components is not fully understood yet, and limited surveillance opportunities adds an additional uncertainty component. The evidence presented in our study, although not specifying a channel, emphasizes the need to increase efforts on the surveillance and understanding of ageing, and of its effect on the mechanical properties of a reactor's various parts, in order to better determine maintenance and replacement efforts required from nuclear operators.

Another implication of our study is that descriptive evidence on the fleet level such as the one presented in graph 3.3 might fail to capture the real effect of age. This is so because a

[31] Two instances of this scientific debate can be read in wide-audience scientific media such as Scientific American or the MIT Technology Review.

[32] Anecdotal evidence for this view is the statement of Michael Shellenberger, president of the Environmental Progress advocacy group in Berkeley, California, that: "*If you maintain them and replace parts, there is no reason why nuclear plants can't run a very long time.*", see Tollefson (2016).

[33] A representative of this opinion is former NRC chair Allison Macfarlane, see Tollefson (2016).

[34] Annealing is a process that could improve the durability of reactor pressure vessels, by *cancelling out*the defects accumulated in the vessels' steel due to the exposure to radiations.

purely descriptive analysis does not disentangle separate components such as age, technology, infrastructure and management. Whereas the total number of PDD safety events is an interplay of those components and appears to fall with calendar time, the statistical inference revealed that, after an initial period of improvement, age has eventually a negative impact on safety. This finding emphasizes the importance of a carefully designed statistical analysis as a complement to technical analyses and descriptive evidence.

Finally, the importance of heterogeneity found in our results suggests that national nuclear energy policies exclusively focused on the age of nuclear reactors might be ill advised. In other words, setting a maximum lifespan irrespective of technological idiosyncrasies could be an inefficient policy. The blunt nature of this policy could entail premature shut-downs of safe reactors or prolonged operation of unsafe ones. Instead, this chapter suggests that delegating lifespan extension decisions (or the definition of the conditions required to grant extended operation licences) to competent safety authorities, capable of focusing on the heterogeneities that characterize nuclear safety across nuclear sites, may be a more efficient policy.

Part II:

Safety Regulation and Disaster Management under Catastrophic Risks

A person is defined to be rational, I believe, if he does the best he can, using reason and all available information, to further his own interests and values. [...] However, rationality is often used loosely in another sense, which is that all behavior is rationalizable as serving to maximize some preference. The two senses of rational are in a way converse. The second says that all behavior is generated by preferences. The first says that when preference exists, behavior serves it.

T. Bewley, in Bewley (2003)

Résumé du chapitre 4

Ce chapitre présente une analyse empirique de l'effet de la surveillance sur le comportement des firmes. Pour ce faire, j'analyse dans ce chapitre les conséquences d'une politique française de surveillance locale et d'information des populations à propos de l'activité menée dans les centrales nucléaires française. Cette analyse combine des données portant sur les déclarations d'événements significatifs de sûreté déclarés dans les centrales françaises entre 2007 et 2017 avec des données économiques locales portant sur l'activité des Commissions Locales d'Informations. Le chapitre propose une méthode instrumentale basée sur des erreurs de prédictions budgétaires, et permettant de mesurer l'effet de l'intensité de l'activité de surveillance et d'information menée par ces commissions locales sur le comportement des opérateurs des centrales nucléaires. En particulier, il apparaît qu'une hausse de l'activité de ces commissions, mesurée par leur budget annuel, mène à une augmentation de la quantité d'événements significatifs de sûreté déclarés par l'opérateur au régulateur. Ces hausses pouvant être dues à des changements dans le niveau de sûreté des réacteurs ou à des changements dans la transparence du comportement déclaratif de l'exploitant, le chapitre propose un modèle théorique dont sont tirés des prédictions empiriques permettant de distinguer ces deux canaux. Il apparaît alors que la hausse des déclarations observée est due à une augmentation de la transparence, plutôt qu'à un effet sur les efforts de sûreté des opérateurs. Si augmenter la quantité d'information à disposition des régulateurs permet ensuite d'augmenter la sûreté nucléaire, cette étude suggère que la surveillance et la publication locale d'informations concernant des activités à risque peut avoir des conséquences positives sur le comportement des firmes impliquées dans ces activités.

CHAPTER 4

Evaluating the effect of local monitoring on nuclear safety: Evidence from France

4.1 Introduction

In this chapter, we empirically assess the effect a nuclear monitoring policy passed in France in 2006 on the reporting behaviours of nuclear plant managers. This policy requires all nuclear power stations to be monitored by local commissions who communicate information regarding the operation of nuclear plants to local populations. Although these commissions have no explicit means to enforce safety within nuclear stations, we hypothesize that through their monitoring and communication activities, they may induce shocks on the costs faced by nuclear plant managers when reporting safety incidents to the nuclear safety authority. Hence, the first aim of this chapter is to assess whether the activities of these commissions, measured by their annual budgets, influence the reporting behaviours of French nuclear plant managers. If so, we also aim to investigate how they influence these behaviours, by disentangling the efforts exerted by managers in order to improve safety, from their decisions to comply with regulatory reporting requirements when they detect significant safety events.

In the context of environmental regulation, informational policies such as public disclosure of information have been shown to increase the compliance of polluting firms with existing environmental regulations, and to decrease their levels of emissions (see e.g. Shimshack (2014) for a review). This chapter investigates whether these results carry forward to the nuclear industry, in which standard-based safety regulations are implemented by centralized safety authorities, and where compliance with these standards is assessed by a voluntary reporting mechanism

[0]We would like to thank the French nuclear safety authority and the personnel of all commissions for local information for providing us with the data that motivated this study and for providing us with multiple insights regarding the nature of their activities. We also thank all participants to the 6th ELAEE conference in Rio de Janeiro, to the 22nd YEEE seminar in Nürnberg Energy Campus, and to the 2017 ZEW Energy Conference in Mannheim for their feedbacks on previous versions of this chapter. All errors are entirely our owns.

combined with audits and penalties for non-compliant facilities. In addition to this command-and-control safety regulation, and since 2006, each French Department[1] hosting a nuclear station has to subsidize a local commission, who monitors the activity of the local nuclear station, and communicates with the local population on the results of this monitoring.[2] We empirically study the effect of this monitoring policy on the decision of local plant managers to exert safety care and to comply with the French voluntary reporting mechanism.

Despite its importance, the question of the effect of monitoring on nuclear safety has remained largely unanswered by the literature. Its empirical evaluation is hampered by three problems. The first one, inherited from the general specifics of nuclear safety, is data scarcity. In particular, severe nuclear accidents are very rare, so that statistical analysis in this context is hardly feasible. The literature has dealt with this issue by either using Bayesian methods in the context of technical probabilistic risk assessment (Rangel and Lévêque, 2014), or by using extended sets of accidents from both nuclear power plants and fuel-cycle facilities (Sovacool, 2008; Hofert and Wüthrich, 2011; Burgherr and Hirshberg, 2014; Wheatley et al., 2017). The second problem stems from the possibility for an operator not to report safety events. These non-compliant behaviours are economically meaningful only if perfect observability of the behaviour of the agent by the principal is not feasible. A first solution to this problem is to study detected non-compliance, as is done in Feinstein (1989). In our case, detected non-compliance is not observable to the econometrician, which implies that observed variations of reporting behaviours can be affected by both safety care and non-compliance. Finally, the intensity of the monitoring of an agent may be partially determined by how safe the agent is perceived by the principal. As a result, safety care, non-compliant behaviours and monitoring intensity are jointly determined, which may induce a simultaneity bias in the estimation.

We contribute to the literature on nuclear safety and non-compliance in several ways, with a particular focus on the three aforementioned problems. First, we use a novel dataset on declared safety incidents in French nuclear power stations. These incidents, although of small magnitude, consist in deviations from the safety standards and operation guidelines set by the French safety authority. They are therefore considered as significant for nuclear safety.[3] The declaration

[1]France's administration is organized in several levels below the national government. The French territory is first divided into thirteen administrative regions. Regions are then divided in a total of a hundred Departments, which are divided in over thirty-six thousands counties.

[2]Most of these Departments had already created such agencies before 2006, on a voluntary basis. The law made the existence of these agencies mandatory in all Departments hosting either nuclear reactors or fuel-cycle facilities.

[3]Through probabilistic risk and reliability analyses - a process of case-by-case scenario analysis performed by the operator and the regulator - an incremental probability of nuclear core meltdown is associated to each of these events.

of these events by plant managers to the safety authority is mandatory. Yet, these events may remain undetected, and managers face countervailing incentives when choosing whether to report detected events. Reporting an event may be costly to the manager, for instance if his salary is based on the safety performance of the plant, or if he has to incur some cost after the declaration in order to redeem a compliant status. Not reporting an event can be costly as well, in case of a public backlash or in case of more stringent regulations. To perform our analysis, we also use data on the monitoring exerted by local commissions between 2008 and 2015. This local monitoring varies in intensity, as Departments provide local commissions with heterogeneous levels of resources. As a result, the monitoring performed by these commissions may range from organizing regular meetings with local managers to hiring independent experts to assess, for instance, the environmental impact of the operation of the power station. We use the annual budget of each commission as a measure of the intensity of their monitoring activities. The heterogeneity in the size of these budgets provides a source of identification.

Second, to disentangle the effects of local monitoring on safety care and non-compliance with declaration guidelines, we adopt a theoretical model for the behaviour of plant managers, adapted from the environmental literature (see e.g. Evans et al. (2009); Gilpatric et al. (2011)). According to this model, the decisions to exert safety care and to comply with declaration guidelines are jointly determined by the perceived sanction incurred by the manager when reporting an event, and the expected penalty faced when incidents are not reported. We derive testable hypotheses from this model and test these by using several observables such as specific or general counts of reported safety events, and nuclear reactor reliability indices. In particular, in a strategy related to Hausman (2014), we look at automatic shut-downs of reactors, a type of events that is perfectly detected by managers and reported to the authority due to the resulting changes in the production of electricity. As non-compliance with declaration guidelines is not possible for this subset of events, the effect of local monitoring on the sanction perceived by managers is identified. In a second step, the estimated effect of monitoring on safety care is measured by observing evolutions of the reliability of nuclear power plants. Finally, conditionally on the results of the former regressions, we can identify the effect of monitoring on compliance by measuring the effect of monitoring intensity on the reports of significant safety events.

Third, we address the endogeneity of the measurement of monitoring intensity by using an instrumental variable (IV). Our instrument is based on the difference between the forecast and the realized annual operating revenues of the French Departments. This forecasting error has several attractive features. First, a forecasting error, once realized, may lead to a reassessment of the forecast for the current or coming fiscal year, and thus induce a change in the budget of the monitoring commissions. Second, such a forecasting error is almost per definition unanticipated, which prevents endogenous forward-looking behaviour of the local authorities. Finally, the source of the error is simply a financial miscalculation due to overall uncertainty or human failure related to tax returns, and thus it can be argued that it is not related to the unobserved factors affecting compliance and safety at the level of nuclear power stations. The second and the third properties of these errors qualify them as a quasi-experiment.[4]

To the best of our knowledge, Davis and Wolfram (2012), Hausman (2014) and Feinstein (1989) are the only papers that analyse the impact of economic incentives on nuclear safety and non-compliance. Davis and Wolfram (2012) and Hausman (2014) identify the effect of market deregulation in the U.S. on the reliability and safety levels of some US nuclear reactors. The proxy used in Hausman (2014) for safety consists of automatic reactor shut-downs and is thus closely related to our identification strategy. Likewise, our use of reliability indices is similar to the empirical assessment performed by Davis and Wolfram (2012). Our main focus, however, is on non-compliance instead of safety and we use the reliability indices and counts of automatic shut-downs only to disentangle the effect of monitoring on safety care from its effect on non-compliant behaviours in a back-door-identification-type strategy. Feinstein (1989) uses data on inspections of US power plants to study the factors of non-compliance and the effect of non-compliance on safety. His identification depends crucially on strong parametric assumptions on the distribution of the unobservables. These assumptions, however, are not guided by economic theory and are in this sense rather arbitrary.

More generally, within the literature dedicated to monitoring on compliance, this chapter is closely related to the work of Duflo et al. (2013) and Telle (2013), who use randomized controlled trials to assess the effectiveness of monitoring programs on self-reporting behaviours. Telle (2013) studies the effect of deterrence on self-reporting, and shows that whereas specific deterrence (e.g. fines) does increase self-reporting, an increased frequency of audit does not. Duflo et al. (2013) show that preventing conflicts of interests in audit mechanisms leads to less non-compliance and more mitigation efforts. In the environmental literature, another related paper is Lin (2013), who assesses the effect of increases in the probability of being monitored on mitigation efforts and truthful self-reporting, using rainfalls as an instrument for the monitoring probability. Compared to these three papers, we investigate a similar question in a different

[4]This instrumental variable is similar in spirit to the natural experiment used in Bressoux et al. (2009), who utilise random administrative mistakes to instrument for the endogenous assignment of teachers to schools in France.

industry, and use an instrumental variable different from the one used in Lin (2013), and instead of a randomized experiment. We also contribute to this literature by introducing possible non-detection of events by the firm and imperfect audit results for the regulator.

Our main results are that nuclear plant managers react to informational incentives provided by this French local monitoring program. These incentives do not induce increases in safety care, but significantly reduce non-compliant behaviours. In particular, a €2.000 increase in the annual budget of a commission leads local managers, in expectation, to increase by 1% their level of compliance with declaration guidelines. While the non-significant impact of monitoring intensity on safety care contradicts the findings of Hausman (2014) and Duflo et al. (2013), the positive effect on compliance is in line with the findings of Feinstein (1989); Telle (2013) and Duflo et al. (2013).

4.2 Institutional setup

4.2.1 Nuclear-power safety in France

The French nuclear fleet is constituted of 58 reactors, located in 19 sites (or plants in the following), owned by a single utility: EDF. Nuclear safety is regulated by the Nuclear Safety Authority (ASN in the following) who sets technical standards regarding the construction, operation and maintenance of all nuclear reactors. In addition, the safety authority establishes reporting criteria which characterize a set of events considered as significant for safety. Upon the detection of any of these events, a nuclear plant manager has to report the event to the safety authority. This self-reporting mechanism aims to foster knowledge spillovers across reactors, and to detect generic design weaknesses or organizational failures, in order to improve nuclear safety.

Non-compliance with this self-reporting mechanism is deterred by the use of periodic and random inspections by ASN inspectors. During these inspections, inspectors access the paperwork describing all the situations detected by plant managers but not considered significant enough for reporting. The firm can be prosecuted for failing to declare significant safety events. In addition, anecdotal evidence suggests that failing to declare safety events can have other costly consequences for plant managers such as production losses due to temporary shut-downs or public backlashes.[5]

Although all French nuclear plants are owned by a single firm, many decisions are delegated to

[5] For instance, the French station Fessenheim was invaded by Greenpeace activists in 2014, after it became public that its managers had understated the magnitude of an incident that happened earlier that year.

the management of each plant. For instance, the reporting of safety events has to be done rapidly after detection, and is thus left to the discretion of power plant managers. Thus, the reported quantity of significant safety events captures both the incentives faced by plant managers when deciding on how much care to dedicate to the limitation of their occurrences, and whether to report observed events to the safety authority. Though, these incentives may be countervailing.

First, if the occurrences of significant safety events lead to extended maintenance periods, or to the shut-down of a reactor, exerting care to limit these occurrences may yield private benefits, for instance through an increased level of reliability of the power stations (e.g. larger production levels). Yet, these private benefits will be offset by the costs of exerting care, such as hiring more staff, fostering safety culture and skills through dedicated trainings, or investing in better equipments.

Second, reporting can be considered costly to plant managers, as they face sanctions when reporting safety events. Regulatory sanctions can consist in necessary investments required by the authority to mitigate the causes of the reported safety events.[6] Internal sanctions could also deter reporting, for instance if managers are incentivized to enhance some plant performance measures based on the counts of these events.[7] On the other hand, failing to report a safety event may also be costly to the manager, due to the possibility of undergoing prosecution if the authority discovers the event, or in case of a public backlash against the power station due to the lack of transparency of its managers.

Finally, it is to be noticed that a significant event can remain unreported for two distinct reasons. First, after observing the event, an operator can deliberately decide not to report it, because the cost associated with the report is larger than the expected cost associated with hiding the event. On the other hand, an event may also not be reported because the manager failed to detect it in the first place. Nevertheless, the consequences of non-reporting seem to be unaffected by the cause of the non-reporting.

In the next paragraphs, we describe the organization of the French local monitoring commissions, and the way they may interplay with the reporting of safety events.

4.2.2 Local monitoring

Since 1981 and the partial meltdown that occurred at the Saint-Laurent-des-Eaux nuclear power station - two years after the Three Mile Island accident - some French departments hosting

[6] In the environmental literature, these sanctions are usually thought of as a tax on reported emissions.

[7] Although the existence of such incentive schemes is not clear, anecdotal evidence on the distribution of event counts suggest it.

nuclear stations have organized a form of local monitoring of nuclear power stations through dedicated commissions whose purpose was to foster transparency regarding nuclear power.[8] In 2006, a law made the existence of these monitoring commissions compulsory in all French Departments hosting nuclear power reactors or fuel cycle facilities.

These monitoring commissions are now composed of four groups of members: locally elected officials (mayors from cities neighbouring the power station or regional counsellors), members of local environmental associations, members of the nuclear plant workers unions, and competent local citizens.[9] These members are not remunerated for their participation, and some restrictions regarding the composition of the commissions are set by law. Elected officials must represent at least 50% of the commission, while each of the other three groups has to constitute at least 10% of the members.

Commissions are funded by the French Departments, as well as by the Nuclear Safety Authority. As the law does not set any rule regarding their budgets, the ASN matches, for each commission, the endowment granted by its Department. Local commissions thus obtain very heterogeneous budgets, which span between 5,000 €/year to more than 190,000 €/year. Due to these variations in endowments, commissions undertake heterogeneous activities, which we now describe in more details.

First, each commission organizes two to three periodic meetings per year, during which plant managers and the safety authority present the main actions undertaken in the nuclear station. Commission members are provided with a set of documents regarding the operation of the nuclear facility to prepare the meeting, and may ask for specific topics to be addressed. In particular, they receive an account of the occurrences of significant safety events within each reactor of their local station.

Based on these meetings, commissions communicate information to the public. To do so, most commissions invite the press to the periodic meetings, and often make public statements regarding the major decisions made by the plant managers or by the safety authority. Depending on their budgets, commissions also publish contents on dedicated websites, distribute journals in city halls, or mail periodic information letters to neighbouring populations. A minority of

[8]In a note circulated in 1981 to local prefects - local state representatives - the French Prime Minister suggested the creation of local commissions dedicated to the monitoring of industrial activities prone to large risks and to the information of the public. At the time, this suggestion aimed to promote a sharing of responsibilities among local collectivities and the State regarding the information of populations about the nuclear risk. The original note authored by Prime Minister Pierre Mauroy (in French) can be downloaded on this website: http://www.cli-gravelines.fr/Services-en-ligne/Espace-documentaire/Documents-a-telecharger/Les-textes-reglementaires/Circulaire-MAUROY-du-15-decembre-1981.

[9]Additional examples can be found on the website of the commission of the Paluel and Penly nuclear power stations.

commissions even organize additional open meetings for interested local inhabitants, and invite local populations from neighbouring countries[10].

Finally, if their budgets allow it, local commissions can hire independent experts in order to carry out assessments of some aspects of the operation of the plant. For instance, past investigations have assessed the environmental impacts of the operation of nuclear stations through radioactivity measurements in local water streams. Results of these investigations can then be discussed during the periodic meetings of the commission with the plant managers and the safety authority.

4.3 A model of monitoring and compliance

In summary, the existence of countervailing incentives faced by plant managers is important to our analysis as the variety of actions undertaken by local monitoring commissions may alter the different costs and benefits incurred by managers when exerting safety care and when deciding whether to report significant safety events. To shed light on the effect of local monitoring on the behaviour of plant managers, we present in the next section a theoretical framework derived from a principal-agent model, which captures the interplay of local monitoring with the incentives for reporting. We derive from this framework the hypotheses that enable the identification of our empirical results. Contrarily to the existing theoretical literature on audit mechanisms[11], we do not model explicitly the optimization problem of the regulator, and only model the best-response of an agent to the exogenous audit mechanism set by the principal. The determination of the optimal audit mechanism is irrelevant in our context, as our empirical estimation will consist in using an instrumental variable method to assess the effect of exogenous changes in monitoring intensity on the behaviour of plant managers.

4.3.1 The model

In the following, we suppose that an agent (the manager) operates a nuclear power reactor subject to a self-reporting mechanism enforced by a principal (the safety authority). This model is adapted from Evans et al. (2009) and Gilpatric et al. (2011) who introduced the possibility of imprecise monitoring technology in environmental emission auditing mechanisms.

[10]Since 2015 and France's new energy transition law, the organization of a third meeting, open to the public, is mandatory for each commission. But as this law has not been implemented yet, the existence of these public meetings is out of the scope of our study.

[11]See for instance e.g. Macho-Stadler and Pérez-Castrillo (2006); Evans et al. (2009); Gilpatric et al. (2011), and Zahran et al. (2014).

Formally, let E_{tot} be a continuous variable capturing the total number of events that occur during a year in a nuclear power reactor. For tractability of the model, we forego the count nature of these events, and assume that this quantity E_{tot} decreases when the agent increases his level of safety care. In other words, we assume that the agent can choose the number of events E_{tot} that occur each year in his nuclear reactor.[12] We further assume that safety care is costly, but provides some private benefits, such as increased reliability of the power station. Hence, we assume there is a function $B(E_{tot})$ concave in E_{tot} with $B'(0) > 0$ and $B'' < 0$, that captures the costs of safety care and its private benefits. In the absence of a principal, the agent would privately choose a level of care associated with a number of events $\bar{E}$, satisfying $B'(\bar{E}) = 0$. We assume that B is a concave function that reaches its maximum, e.g. that $\bar{E}$ is finite.

Second, in order to model the detection abilities of the agent, and his level of compliance with declaration guidelines, we assume that a fraction ρ of events are privately observed by the agent, and that a fraction z of privately observed events are declared to the principal. In other words, the agent observes $E_{obs} = \rho E_{tot}$ and declares to the principal the quantity $zE_{obs} = z\rho E_{tot}$. In the following, we assume that the detection ability of the agent is exogenous. In this case the agent is only left with the discretion of choosing safety care (E_{tot}) and the level of compliance (z).[13] In section 4.6, we let ρ be endogenously chosen by the agent, and comment on how the existence of this third behavioural channel affects our results.

The fact that some events are not observed by the agent may be due to an imperfect knowledge of his equipment, or to a limited time spent trying to detect these events. Yet, as the agent knows that he may fail to detect a certain number of events, ρ is assumed to be known by the agent. In other words, the agent observes E_{obs}, but knows the total number of observable events E_{tot}. Yet, given the nature of the reporting mechanism, which requires to provide numerous details about the causes and consequences of each event, we assume that the agent cannot report an event which he did not really observe.

In addition, as inspectors have limited time to perform their inspections, we will assume that audits do not perfectly reveal all unreported events. More specifically, we assume that audits may reveal any safety event, regardless of whether they were observed by the agent. To do this, we adapt the model developed by Evans et al. (2009) and Gilpatric et al. (2011) to capture the

[12] If the agent can exert a costly level of safety care s, and if E_{tot} is a decreasing function of s, and if s only affects E_{tot}, thn choosing s or choosing E_{tot} is equivalent in terms of the agent's preferences.

[13] It is to be noticed that holding ρ constant does not mean that detection ability is constant, but rather that the rate of privately observed events to the total number of events remains constant. This can be interpreted by saying that increasing safety care reduces the total number of events occurring in a reaction and reduces proportionally the quantity of events which are not privately observed by the plant manager.

imprecision that characterizes CO_2 emissions measurement technologies. Note that a crucial difference from their frameworks is that the imprecision of the audit describes both the fact that the agent may fail to detect safety events, and the fact that the principal has only limited audit resources. The novelty in our case is the fact that the agent may endogenously affect the outcome of the audit.[14]

Hence, let u be a random variable distributed according to a cumulative distribution F and density f over $[0;1]$. The value taken by u represents the fraction of E_{tot} detected by the principal during audits. When u takes values between z and 1, the audit reveals a number of events larger than what the manager publicly reported. When u takes values between ρ and 1, the audit reveals a number of events larger than what the manager privately observed. This captures both the imperfection of inspections performed by the principal, and the possibility of non-detection of safety events by the agent. The expected quantity Q of unreported events revealed by the audit is:

$$Q = E_{tot} \int_{z\rho}^{1} (u - z\rho) f(u) du. \tag{4.1}$$

Upon the observation of an event, the agent can report the event and face a sanction α, which embodies the direct consequences of reporting, such as mandatory investments. Using previous notations, the sanction associated with reporting $z\rho E_{tot}$ events is $\alpha z\rho E_{tot}$. The agent may also decide not to report the detected event to avoid the sanction. To deter this behaviour, the principal audits the agent with a given probability q, and levies a penalty β when unreported events are detected. The probability of inspection can be thought of as the frequency of planned or unplanned inspections. The penalty embodies the consequences of non-declaration, such as legal prosecution, public backlashes or increases in the stringency of the regulatory oversight. The expected penalty faced by the agent is thus $q\beta Q = q\beta E_{tot} \int_{z\rho}^{1} (u - z\rho) f(u) du$.

Under this self-reporting mechanism with imperfect audits and imperfect observation of events, a risk-neutral agent maximizes the following quantity:

$$\max_{E_{tot}, z} B(E_{tot}) - \alpha z\rho E_{tot} - q\beta E_{tot} \int_{z\rho}^{1} (u - z\rho) f(u) du \tag{4.2}$$

In this program, the agent faces a cost associated with the occurrences of each safety event which we note note $\mu(z) = \alpha z\rho + q\beta \int_{z\rho}^{1} (u - z\rho) f(u) du$.

[14]More specifically, when choosing ρ, the agent can affect the extent to which the principal may find more events than the quantity declared by the agent. The distribution of u, however, cannot be affected by the agent.

4.3.2 Comparative statics and testable hypotheses

Suppose that detection ability ρ is exogenous. Let $z^\star$ and $E^\star_{tot}$ be the best response played by the agent given an exogenous audit mechanism characterized by α and $q\beta$. Provided $\alpha < q\beta$, the existence of an interior solution for $z^\star$ is ensured. This condition captures the fact that if the perceived sanction for reporting is higher than the expected penalty for non-reporting, then the agent never reports and an interior $z^\star$ cannot exist. Likewise, provided $\mu(z^\star) < B'(0)$, there exist an interior $E^\star_{tot}$ that maximizes equation (4.2).[15]

We can then derive the following comparative statics describing the effect of a change in the value of parameters α and $q\beta$ on safety care $E^\star_{tot}$, compliance $z^\star$ and the total observed quantity of reports $z^\star \rho E^\star_{tot}$, which we note $z^\star E^\star$ for simplicity.

Proposition 4.3.1. *Comparative statics*

At an interior solution, the following results hold:

- $\frac{\partial E^\star_{tot}}{\partial \alpha} < 0$*: a marginal increase in* α *leads to a decrease in* $E^\star_{tot}$*,*
- $\frac{\partial z^\star}{\partial \alpha} < 0$*: a marginal increase in* α *leads to a decrease in* $z^\star$*,*
- $\frac{\partial E^\star_{tot}}{\partial q\beta} < 0$*: a marginal increase in* $q\beta$ *leads to a decrease in* $E^\star_{tot}$*,*
- $\frac{\partial z^\star}{\partial q\beta} > 0$*: a marginal increase in* $q\beta$ *leads to an increase in* $z^\star$*,*

This result is identical to the result derived by Evans et al. (2009). A proof is proposed in appendix A.2. Two interesting direct corollaries from this proposition are worth being noticed. First, the first two comparative statics in proposition 4.3.1 show that a marginal change in the level of perceived sanctions α has an unambiguous effect on the total quantity of reports $z^\star E^\star$. Second, the last two comparative statics show that a marginal change in $q\beta$ has an ambiguous effect on $z^\star E^\star$. To see this, we can write:

$$\frac{\partial z^\star E^\star}{\partial q\beta} = E^\star \frac{\partial z^\star}{\partial q\beta} + z^\star \frac{\partial E^\star}{\partial q\beta} \tag{4.3}$$

We know that the first term in the right-hand side of (4.3) is positive, while the second term is negative. Therefore, the variation in observed reports induced by a marginal change in $q\beta$ is determined by the relative size of these two terms, and in particular by the relative amplitude

[15] If $\mu(z^\star) > B'(0)$, the agent exerts a level of safety care associated with no occurrences of safety events. In other words, we can interpret this assumption as the fact that reducing the number of safety events to 0 would be infinitely costly, so that there always exist an interior solution for the optimal level of safety care.

of the variations in compliance and in safety care. For instance, if $\frac{\partial E^\star}{\partial q\beta}$ is small enough, then an increase in $q\beta$ should be followed by an increase in the quantity of observed reports.

Proposition 4.3.1 and its two corollaries depict the fundamental problem of identification in our set-up. First, any observed change in the observable outcome $z^\star E^\star$ can be due to changes in either perceived sanctions α or expected penalties $q\beta$. Moreover, conditionally on a constant perceived sanction α, the effect of a marginal change in $q\beta$ on the observable outcome $z^\star E^\star$ consists of two effects: the effect on compliance: $E^\star \frac{\partial z^\star}{\partial q\beta}$, and the effect on safety care: $z^\star \frac{\partial E^\star}{\partial q\beta}$. As these two effects have opposite signs, any change in $z^\star E^\star$ can be explained by either an increase or a decrease in $q\beta$.

Thus, the fundamental problem of identification consists in assessing whether the activities undertaken by local monitoring commissions induce changes in perceived sanctions or changes in expected penalties, and then to disentangle the different channels which determine the compound effect of monitoring on the observable outcome $z^\star E^\star$. We refer to this identification issue as the *channel identification problem*. It adds additional complexity to the problem of potentially endogenous level of monitoring and motivates our identification strategy, which we outline in section 4.5.

4.4 Data and descriptive statistics

The data we use to conduct this study emanates from three sources: the French Nuclear Safety Authority, the French utility EDF, and fourteen local monitoring commissions. Our unit of observation is set at the level of the reactor.year. In other words, each observation in our dataset consist of a pair (reactor, year). Our dataset consists in an unbalanced panel of 234 observations of reactor-years, spread across 50 different nuclear reactors observed between 2008 to 2015. As the French fleet contains 58 nuclear reactors, the largest possible dataset that we could have gathered over the same time period would have contained 464 observations. The 8 missing reactors are located in 4 sites, whose commissions could not provide us with any reliable data regarding their activities. The rest of the missing data is due to the fact that many commissions could not provide us with financial data prior to 2010. The following paragraphs describe our different variables as well as their sources.

Table 4.1: Descriptive statistics: treatment variables.

Variable		Mean	Std. Dev.	Min.	Max.
Commission controls	*budget*	52.415	48.146	4	198
	meet	2.271	0.446	2	3
	multiple	0.169	0.376	0	1
	saintlaurent	0.051	0.22	0	1

236 observations in 50 reactors from 2007 to 2015 (522 possible)

Note: This table describes the activities of local monitoring commissions, measured by several observables characterizing the frequency of their meetings and their communication strategies.

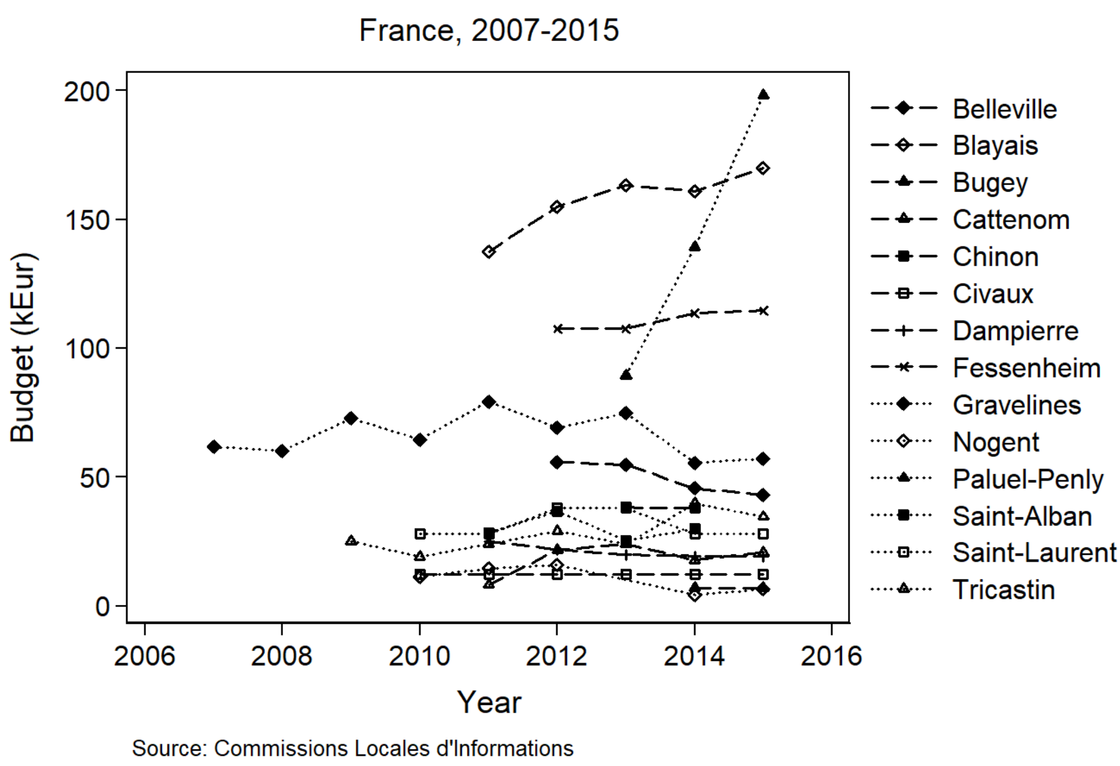
4.4.1 Treatment variables

We first gathered annual activity reports describing the activities of local monitoring commissions. This dataset contains information regarding their annual budgets, i.e. the endowment received from the Department councils and the subsidies granted by the ASN. We also retrieved information regarding the administrative statuses and composition of the commissions. We finally gathered data regarding the frequency of their meetings, and whether these meetings are open to the press or the public. We also retrieved information regarding the number of independent studies commissioned by these agencies. Finally, we identified the commissions which have multiple facilities to monitor, as some nuclear sites in France host more than one nuclear facility. Descriptive statistics regarding these commissions are gathered in table 4.1. The budgets of the commissions varies in the data from 4 000 €/year to 198 000 €/year. Notice also that most commissions only organize two meetings a year, and that only one of them has multiple sites to monitor.

We use the annual budgets of local commissions as a proxy for the intensity of their monitoring activities. We deem this variable to be a good proxy for the intensity of the commissions as these budgets are used to finance environmental impact assessments by independent experts, in order to train commission members, or to pay for the diffusion of the information gathered by these commissions. Commissions endowed with larger budgets are thus likely to be able to induce greater shocks on the perceived sanctions and expected penalties faced by plant managers when deciding how much safety care to exert and whether to report observed events. The variations of these budgets over time and across nuclear stations are presented on figure 4.1.

On the other hand, these budgets are an endogenous measure of the intensity of the monitoring performed by the commissions. Commissions budgets and the behaviour of plant managers may be simultaneously determined, as budgets can increase because commission members expect the safety of their local plant to be diminished, while managers may increase safety care

Figure 4.1: Budget variations across local monitoring agencies



Note: Total annual budget received by local commissions over the period 2008-2015. The Paluel and Penly nuclear stations are located in the same Department and are monitored by the same commission. The graph illustrate both the heterogeneity of the budgets and the unbalanced nature of our panel.

and compliance when they expect local commissions to exert more monitoring pressure on them thanks to larger budgets. This endogeneity issue is addressed in section 4.5.

The other treatment variables described above are also prone to endogeneity. For instance, the number of annual meetings can also be affected by the reverse causality described in the previous paragraphs. Yet, it can be noticed that only the budget variable is varying over time and across nuclear sites, whereas the other variables, such as the number of meetings or the variables describing the communication of each commission, are time invariant. This means that if these variables are endogenous, they have to be related to the expectations held by commission members when their values were set. In this sense, the endogeneity of these variables is less troublesome than the one characterizing the commissions' budgets.

4.4.2 Reporting and reliability

As a proxy for nuclear safety, we use a dataset obtained from the French Nuclear Safety Authority which contains the significant safety events reported by plant managers. Although these events only have minor consequences, their number is substantially larger than the number of nuclear accidents. This dataset contains over 19.000 safety events, declared between 1972 and 2015 in currently operated nuclear power stations. We restrict our analysis to the events reported between 2008 and 2015, to match our data regarding the local commissions' activities.[16]

Within this dataset, we focus on counts of events annually reported in the French reactors. In order to implement the identification strategy described in the following section, several counts of events will be considered: the count of all events declared during a reactor-year (ALL), and specific counts of events declared during a reactor-year, such as automatic shut-downs (ASD) or unplanned uses of safeguard systems (SFG). These two types of events were identified jointly with the safety authority as being subject to perfect detection and declaration, a property which we will use to disentangle the effect of monitoring intensity on safety care and compliance with declaration guidelines. Automatic shut-downs have an impact on the electrical output of the power station, and are thus impossible to hide. Events requiring the use of safeguard mechanisms are deemed particularly severe and easy to detect by the authority. These two categories are jointly referred to as perfectly detected and declared events, and measured by the variable PDD. It is to be noticed that we have $PDD = ASD + SFG$.

In order to control for the various differences across reactors that may also explain the occurrences of safety events, we rely on two datasets obtained from the Nuclear Safety Authority and the French utility EDF. These datasets contain detailed information regarding the annual production levels, as well as information regarding the reliability of nuclear reactors. In particular, we use data on the annual length of maintenance activities conducted in each reactor, and on the share of electricity lost due to unplanned maintenance extensions (K_m) or due to fortuitous stops (K_f).

In addition, we construct several variables that account for the history and technological design of the reactors. We first construct an age variable that describes the age of a reactor during the calendar year of observation. Age is defined here as the time elapsed between the

[16] Within the events reported, we dropped the so-called 'generic' events, which are specific to the whole fleet or to a large group of reactors. When computing counts of events for any particular reactor over a given period, the events that affected systems common to several reactors within a site were accounted for in the counts associated with each affected reactor.

Table 4.2: Descriptive statistics: reactor-level data.

Variable		Mean	Std. Dev.	Min.	Max.
	ALL	12.856	4.778	2	27
Event counts	PDD	1.017	1.13	0	5
	ASD	0.809	0.955	0	5
	SFG	0.208	0.492	0	3
Reactor reliability	K_m	0.048	0.063	0	0.583
	K_f	0.033	0.051	0	0.427
Reactor controls	age	28.169	5.659	8	37
	$size$	3.966	1.38	2	6
	$FOAS$	0.559	0.498	0	1
	$FOAK$	0.008	0.092	0	1
	$production$	6.866	1.747	2.165	11.622
	$maintenance$	67.568	49.839	0	279

236 observations in 50 reactors from 2007 to 2015 (522 possible)

Note: the first four variable describe counts of safety events reported each year by each reactors. The second two variables capture the reliability of each nuclear reactor, measured by the share of electricity lost to unplanned maintenance works and to fortuitous stops. Finally, the last variables are reactor controls that describe their yearly production levels and some of their technical features.

period of observation and the first divergence of the core of the reactor.[17] We also construct three design fixed effects dummy variables that match the three power plant designs that coexist in the French fleet. In order to capture possible learning-by-doing effects, we finally construct dummy variables which identify the first reactors built within each nuclear site and the first reactors built within the groups of reactors sharing a common plant design. These variables are described further in table 4.2. The age of the reactors considered ranges from 8 to 37. Sites include from 2 to 6 reactors, each of which produced an average of 7 TWh per year over the elapsed period of time, and underwent an average 68 days of annual maintenance.

4.4.3 Attrition bias, local monitoring and reporting behaviours

As our study relies on an unbalanced panel of pairs of reactor-year, table 4.3 proposes comparisons of means for several observables between the sample studied and the sample of excluded observations, for which we could not obtain data regarding the budget of the local commission. Table 4.3 shows that some variables take significantly different values in each sample. The difference in mean age is intuitive, as most missing data would characterize the period 2008-2010. The difference in reports (ALL) suggests that there may be some degree of attrition bias, but the difference is not very significant.

[17]Other possible definitions of the age of a reactor is the time since the beginning of its construction, its connection to the network, or the start of its commercial operation.

Table 4.3: Descriptive statistics: attrition bias.

Variable	**Sample mean**	**Out-of-sample mean**	**t-statistic**	**p-value**
ALL	12.85	13.67	1.81	0.07
SDD	1.09	1.19	0.89	0.38
ASD	0.87	0.89	0.23	0.82
SFG	0.23	0.30	1.49	0.14
K_m	0.04	0.05	1.25	0.21
K_f	0.03	0.04	2.24	0.03
Age	29	24	9.49	0.00
Production	6.86	7.28	2.50	0.01
Maintenance	66.98	68.59	0.34	0.73

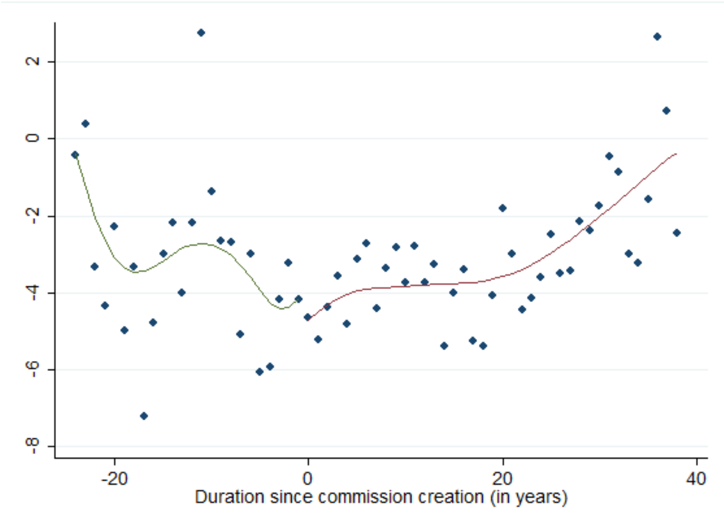
Note: In this table, we compare the mean value of some observables for the pairs of reactor-year within our panel (sample means), with the pairs of reactor-year for which we did not obtain data regarding the budgets of local commissions (out-of-sample mean). For most variables, the difference is not significant. The difference in mean age is normal, as most missing data would characterize the period 2008-2010. The difference in reports (ALL) suggests that there may be attrition bias, but the difference is not highly significant.

In addition, and similarly to the descriptive lowess regression provided by Hausman (2014), we conducted a simple analysis of the quantity of events reported per reactor and per year before and after the creation of their local monitoring commission. This test aims to provide some tentative evidence of the effect of local commissions on the reporting behaviour of nuclear plant managers.

To do so, we first regress the yearly reports of each reactor on yearly dummies, to obtain a time-corrected estimation of the quantity of events reported each year in each reactor. This time correction controls for changing declaration criteria, which may evolve over time. We then plot this quantity against a time indicator, calculated as the difference between the year of declaration and the year of creation of the local monitoring commission.[18] Figure 4.2 shows this plot. Two lowess fits are calculated on the two sub-samples defined by the date of creation of local commissions. These lowess fits suggest that the introduction of local monitoring commission was correlated with an increase in the number of events reported annually by nuclear plant managers. Yet, this statistic accounts for neither the intensity of the local monitoring, nor for the endogeneity that characterizes its measurement. We tackle this issue in the following two sections.

[18] As was mentionned in the institutionnal setup, some local commissions were created as early as 1981, although there existence only became compulsory in 2006.

Figure 4.2: Descriptive statistics: local monitoring and reporting behaviours.



Note: Two-part lowess of the time-corrected annual quantity of events reported by French nuclear reactors. The x-axis shows years since creation of the local monitoring commission. Simple lowess fits suggest an increasing trend in declarations after the creation of local commissions.

4.5 Empirical strategy

4.5.1 Endogeneity of monitoring intensity

Another identification problem that formally and logically precedes the Channel problem is the identification of the total effect (that is, regardless through which mediation channel) of the activity of the commissions on $z^{\star}E^{\star}$. This problem stems from two potential endogeneity sources. First, the budget of these commissions, which we use as a proxy for the intensity of their monitoring activity, is potentially endogenous due to reverse causality: a local commission may have a high level of activity because the Department council - which allocates their budget - is aware that the power station has an abnormal level of declaration of events. Second, the incentives provided by the regulator to the plant managers might be related to the intensity of the local monitoring. In a period of intensified political or public debate regarding nuclear power, for instance, both the threat of new regulation and the expenditures for local monitoring might be increased. Since we do not measure the intensity of the regulatory oversight, the intensity of the commissions' activity could be endogenous due to an omitted variable bias (OVB in the following).

In order to solve the two endogeneity issues mentionned above, we use an instrumental variable method. Our instrumental variable, $INSTR$, is based on a quasi-experiment triggered by forecasting mistakes. More precisely, we use *ex ante* budget data published every year by

Department councils, in which a forecast of the balance of their revenues and expenditures for the upcoming year is provided. We also use *ex post* data on the Departments realized financial revenues in order to compute the forecasting errors made by the Departments. These errors are attractive in many respects. They are first, by nature, unanticipated, which precludes endogenous forward looking behaviours of local authorities. Second, as these errors are the results of a failure from the local officials to predict accurately the revenues levied by local taxes, it seems fair to assume that this error will be independent from the unobserved factors influencing the commissions' budgets. Finally, failing to predict accurately their revenues may lead local officials to reassess the funding provided to local monitoring commissions.

The instrument $INSTR$ is defined using public financial data from French Departments[19]. More precisely, we used the detailed reports describing both the anticipated and realized budgets for each Department hosting a nuclear power station. These budgets include two main sections: revenues and expenditures, whose total are equal. Within each section of these budgets, one can find two main categories: investment revenues/expenditures, and operating revenues/expenditures. Both categories are separated: investment revenues can only finance investment expenditures, while operating revenues are used to finance operating expenditures. Operating budgets account for approximately 85% of the budget of the French Departments. Yet, it appeared from a careful analysis of these datasets that the total forecast error is mostly driven by the investment revenue forecast error. Though, subsidies granted to the monitoring commissions are part of the operating budgets of the Departments. Therefore, the $INSTR$ variable was defined for each reactor and every year between 2008 and 2015 as the two-year lagged value of the forecast error on Department-level operating revenues. The two-year lag is introduced because real budgets are usually published with a one-year delay. This suggests that forecasting errors made in year t are only known at the end of year $t+1$, and may affect the decision to subsidize local monitoring commissions at year $t+2$.

4.5.2 The channel identification problem

Observable channels

Using the instrumental variable described above, we can now estimate the effect of an exogenous change in monitoring intensity on the observed number of events reported annually by each nuclear reactor. Yet, given the model developed in section 4.3, we cannot at this stage disentangle

[19]French local territories publish their budget forecasts and realized budgets every year on this governmental website.

the effect of a change in monitoring on safety care (e.g. changes in E_{tot}) from its effect on compliance (e.g. changes in z). This channel identification problem was presented in section 4.3 in equation (4.3). To solve this issue, we first need to identify whether an exogenous change in monitoring intensity affects the sanctions perceived by plant managers, or the expected penalty faced when choosing not to report safety events. We then need to disentangle the effect of monitoring intensity on safety care and compliance.

To do this, we first identify within the reports of significant safety events a subset of events which are perfectly detected by managers and necessarily declared to the authority. These events first contain automatic shut-downs of reactors (ASD), which are perfectly detected and declared due to the fact that they instantaneously stop the electrical production of a reactor. They also contain events labelled "abnormal uses of safeguard systems" (SFG). These rare events are considered particularly severe by the safety authority, and are particularly simple to detect during audits.[20] We refer to these events as PDD events (for Perfectly Detected and Declared), and note their total yearly number per reactor E_{PDD}. For this subset of events, we have $z = 1$ and $\rho = 1$. Hence, assuming that perceived sanctions and expected penalties are constant across all types of events, the first order condition defining the optimal effort exerted by the manager to prevent these events is the following:

$$B'(E_{PDD}) = \alpha \tag{4.4}$$

Therefore, if variations in the intensity of the activity of local commissions lead to a change in the occurrences of these perfectly detected and declared events, then we can conclude that the activity of these commissions do induce a change in the sanction perceived by the plant managers, which led to a change in the level of safety care exerted by the agent. Conversely, the absence of an effect of the monitoring performed by local commissions on the occurrences of PDD events, combined with a significant effect of local monitoring on the unrestricted quantity of events reported to the authority will be consistent with the hypothesis that local monitoring affects the expected penalty faced by plant managers when failing to report an event.

In addition, we can also assess the effect of monitoring on safety care by estimating the effect of the monitoring performed by local commissions on the reliability of reactors, which is a proxy for the level of safety care exerted by plant managers. We expect reliability to be correlated with safety care since reducing the frequency of these events would limit the likelihood of fortuitous

[20] These two categories of events were selected jointly with the nuclear safety authority during informal discussions.

stops of the reactors or would limit the need to extend maintenance periods for safety reasons. Both of these effects would in turn increase reliability. Finally, as reliability is directly related to the profits made by the power station, we argue that a manager provided with incentives to increase safety care will do so, when possible, in a way that increases his profits.

Reliability is here measured by two variables: the annual share of electricity lost due to unplanned maintenance extensions, and the annual share of electricity lost due to fortuitous stops. These proxies capture the quantity of electricity that was not produced due to unplanned maintenance works or fortuitous stops.

Identification

Our identification strategy is based on the combination of the results of the estimation of the effect of local monitoring on three observed variables: the quantity of events reported, the quantity of PDD events reported, and the reliability of power stations. The first estimation, using counts of reported events, provides the general direction of the effect of monitoring on compliance and safety care. If monitoring affects z or E_{tot}, this first estimation provides us with the sign and amplitude of the variations of $z\rho E_{tot}$. The second estimation, using counts of perfectly detected and declared events, indicates whether local monitoring affects perceived sanctions α. The third estimation, using reliability indicators, provides us with the sign of the effect of monitoring on safety care. We now discuss under what circumstances these estimations allow us to identify the various channels described previously.

First, if monitoring affects the total counts of events $z\rho E_{tot}$ but does not significantly affect sanctions, or if the effect on reports and the effect on sanctions are not coherent with the first corollary of proposition 4.3.1, then we can unambiguously claim that monitoring affects expected penalties. In all other cases, these two estimations are insufficient to identify whether local monitoring affects expected penalties, and only the joint effect of monitoring on compliance and safety care can be identified.

If the effect of monitoring on sanctions and safety care have the same sign, then the third estimation can only confirm (or reject) the predictions of our model, but brings no new information. On the other hand, if the sanction channel is closed by the first two estimations, or if the effects of monitoring on sanctions and on safety care have different signs, then it follows that monitoring has to have an effect on expected penalties. The sign of this effect can be determined by the sign of the effect of monitoring on safety care, using proposition 4.3.1.

Quantitatively, if the third estimation fails to show that monitoring has a significant effect on safety care, then we can neglect the second term of the right-hand side of equation (4.3). It follows that we can interpret the results of the first estimation as the effect of monitoring on compliance z. To see this, note i the intensity of local monitoring, and β_{budget} the coefficient of the *budget* variable used as a proxy for monitoring intensity in the different estimations. When considering the estimation of the effect of local monitoring on observed reports of events, we have $\beta_{budget} = \frac{\partial zE}{\partial i} = z\frac{\partial E}{\partial i} + E\frac{\partial z}{\partial i}$. Then, if $\frac{\partial E}{\partial i} = 0$, we have that $\beta_{budget} = E\frac{\partial z}{\partial i}$. Unless the sanction channel is convincingly closed by the results of the second estimation, the effect of monitoring intensity on compliance can in general be caused by both a change in perceived sanctions and in expected penalties.

When local monitoring intensity significantly increases expected penalties, such that the resulting effects on safety care and compliance have different signs[21], then the results of the first estimation can still be interpreted quantitatively. Indeed, as $\beta_{budget} = \frac{\partial zE}{\partial i} = z\frac{\partial E}{\partial i} + E\frac{\partial z}{\partial i}$, then we have $\beta_{budget} > z\frac{\partial E}{\partial i}$ and $\beta_{budget} < E\frac{\partial z}{\partial i}$. Hence, observing a positive β_{budget} constitutes a lower bound for the (positive) effect of monitoring intensity on compliance. Likewise, observing a negative β_{budget} constitutes an upper bound for the (negative) effect of monitoring intensity on safety care.[22]

Econometric framework

A linear specification allowing to carry out the estimations presented above is:

$$Y_{it} = \beta \cdot X_{it} + \beta_{budget} \cdot budget_{it} + \eta_i + \delta_t + \epsilon_{it} \tag{4.5}$$

where indices i and t respectively refer to the reactor and the year of observation. As was described above, the dependant variable Y_{it} will be defined in turn as the total number of events reported per reactor and per year (ALL), as the number of reports of perfectly detected and declared events (ASD and SFG), and finally as the performance indicators describing the reliability of nuclear stations (K_m and K_f). δ_t represents year fixed effects, and controls for potentially varying declaration guidelines, or particular time-varying factors, such as generic efforts exerted by EDF at a national scale. η_i represents reactor fixed effects, which capture potentially varying local factors influencing the safety of nuclear reactors. In all regressions,

[21] This can arise in multiple situations, for instance if observed reports increase but reliability decreases, or if sanctions are constant but observed reports vary.

[22] As safety care is measured here as a number of events, it is to be noticed that a negative effect of monitoring intensity on safety care means that a more intense monitoring leads to a decrease in the total number of occurrences of events.

control variables X include reactor age, electrical production[23] and the overall number of days of maintenance during the year.

Given the endogeneity of the *budget* variable, we estimate equation 4.5 using our instrument $INSTR$ and a two-stage-least square estimator with robust standard-errors.[24] Robustness checks regarding the linear specification of the model, the nature of the estimator, and the definition of the treatment variables are carried out in Appendix C.1.

Estimation results are presented in the following way. First, table 4.4 contains the results of five regressions. In all regressions, the dependant variable Y_{it} is defined as the number of reported significant safety events. The first regression is an OLS regression including reactor specific fixed effects and our time-varying controls. The next four regressions present the result of a 2SLS estimator, in which each regression differs in the definition of the fixed-effect included. Four levels of fixed effects are included: no fixed effects, technological fixed effects controlling for the main three designs of reactors that coexist in the French fleet, Site fixed effects and reactor fixed effects. The additional control variables described above are included wherever it makes sense. These regressions measure the effect of monitoring intensity on the general reporting behaviour of local managers.

Second, table 4.6 and 4.7 present the results of four regressions in which the dependant variable Y_{it} is defined as, respectively, the annual number of reports (per reactor) of automatic shut-downs ASD, the annual number of reports of unplanned uses of safeguard systems SFG, the rate of lost production due to unplanned prolonged maintenance works K_m and the rate of lost production due to fortuitous stops K_f. Regressions presented in table 4.6 test whether increased monitoring intensity induce a change in the sanctions perceived by plant managers. The last two regressions presented in table 4.7 test whether increases in monitoring intensity lead to changes in safety care.

The results of the first-stage regressions are reported in table 4.5. These first-stage regressions support our instrument as the coefficient of the $INSTR$ variable is positive and highly significant. In addition, the test statistics reported in tables 4.5 and 4.4 support our instrumental variable.

[23] Production can be seen as a form of exposure, as all power stations do not produce the same amount of energy each year.

[24] As we only have one endogenous regressor and one instrument, the GMM-IV, two-stage least-square and limited information maximum likelihood estimators are equivalent, see for instance p.189 in Wooldridge (2002b), or chapter 8.6 in Hayashi (2000). In particular, the GMM-IV estimator has the appealing property to be approximatively unbiased and to achieve close to perfect nominal coverage in the just-identified case (see e.g. Angrist and Pischke (2009b,a))

Table 4.4: Monitoring intensity and reports of significant safety events

	OLS		2SLS		
VARIABLES	ALL	ALL	ALL	ALL	ALL
budget	-0.0222	0.0651**	0.0571**	0.137*	0.132*
	(0.0272)	(0.0281)	(0.0244)	(0.0747)	(0.0736)
age	-0.183	-0.626***	0.178	0.332	-0.653
	(0.183)	(0.225)	(0.206)	(0.241)	(0.803)
production	-1.642***	-0.0253	-1.137*	-0.758	-0.942
	(0.586)	(0.390)	(0.603)	(0.616)	(0.731)
maintenance	0.000883	0.0314***	0.0124	0.0183	0.0135
	(0.0109)	(0.00938)	(0.0130)	(0.0121)	(0.0140)
status		8.547***	9.357***		
		(2.536)	(2.487)		
multiple		-2.147*	-2.355**		
		(1.261)	(1.199)		
meet		8.206***	7.513**		
		(3.071)	(2.941)		
saintlaurent		-2.156	-1.958		
		(3.628)	(3.675)		
size		0.987	1.554**		
		(0.647)	(0.695)		
FOAS		0.525	-0.481	-0.214	
		(0.781)	(0.678)	(0.807)	
FOAK		0.703	-0.558	0.961	
		(1.788)	(1.502)	(2.336)	
1300 MW			9.191***		
			(2.546)		
1450 MW			20.20***		
			(4.504)		
Constant	30.56***	1.446	-12.06	13.12	32.85
	(6.146)	(8.998)	(9.829)	(8.665)	(24.43)
Observations	234	234	234	234	234
Year FE	Y	Y	Y	Y	Y
Indiv. FE	Reactor	No	Capacity	Site	Reactor
R-squared	0.388	-0.033	0.170	0.307	0.384
KP rk Wald		26.59	26.83	11.82	10.03
Wu-Hausman		15.42	13.46	6.441	6.320

*** $p<0.01$, ** $p<0.05$, * $p<0.1$. Robust standard-errors

The table present the results of five regressions. The first regression is an OLS regression for the analysis of the endogeneity bias. The next four 2SLS regressions differ in the definition of the individual fixed effects. All four regressions show a significant positive effect of monitoring intensity on the quantity of events reported. The first-stage statistic provides support for the instrument.

4.6 Empirical results

4.6.1 Monitoring and reporting behaviours

The regressions presented in table 4.4 show that increased intensity of local monitoring leads nuclear managers to increase significantly the number of safety events reported. At this stage, this increase may be due to either an increase in the total number of occurrences of events, or to an increase in the level of compliance of plant managers with declaration guidelines.

The coefficient associated with the budget of local commissions reported in table 4.4 is positive and highly significant for all definitions of the fixed effects in the 2SLS estimator. In appendix C.1, we show that our results are also robust to count specifications such as the GMM-IV Poisson proposed by Cameron and Trivedi (2013) or the Poisson Control-Function approach proposed by Wooldridge (2002b, 2015).

Table 4.4 also contains the result of a fixed-effect OLS regression in which the endogeneity of the budget variable is unaccounted for. In this specification, the coefficient associated with the budget variable is negative, although not significant. This downward bias of the OLS regression is consistent with the reverse causality that may exist between the measurement of local monitoring intensity and the behaviour of local plant managers. Indeed, if local commissions lobby for higher budgets when they expect their local managers to hide information, then we would expect to see high budgets where declarations are rather low, which could be consistent with the observation of a negative coefficient associated with the treatment variable.

The other results provided in table 4.4 are intuitive. Production is negatively correlated with reporting, which is consistent with the fact that most events occur during maintenance works. Age and declarations are negatively correlated when no fixed effects are included.[25] An analysis of the technological fixed effects shows that reactors from the group of 1450 MW and 1300 MW reactors report more events than those in the 900 MW group, which is consistent with the increase in technical complexity of these reactors.

Finally, first-stage statistics are reported in table 4.5, and provide support for our instrument, whose coefficient in both first-stage regressions (with and without reactor fixed effects) is positive and highly significant. The positive sign is coherent with the definition of the instrument: when department fail to forecast accurately their tax revenues, they adapt their spendings accordingly. In other words, when departments collect larger operating budgets than expected, they also allocate larger budgets to their local monitoring commissions. In addition, the Kleibergen-Paap F-statistics reported in table 4.4 are all larger than 10 and thus provide additional support for our instrument. It is to be noticed that in fixed effects regressions, the value of the Kleibergen-Paap statistics is lower than the Stock-Yogo-15% statistic.

[25]Though, this coefficient cannot be interpreted as the causal effect of ageing on the reporting of safety events, as this regression fails to include cohort-specific variables. Hence, the age coefficient captures both the effect of ageing and cohort-specific effects.

Table 4.5: First-stage results of 2SLS regressions

VARIABLES	budget	budget
INSTR	0.298***	0.772***
age	3.332***	5.230***
production	-3.798*	-2.566
maintenance	-0.0647	-0.0659
Status		-13.79
multiple		43.59***
meet		-110.3***
saintlaurent		-137.2***
size		-27.41***
FOAS		-7.981
FOAK		45.36***
1300.Power_Group		-41.85***
1450.Power_Group		-55.97*
Constant		275.9***
Observations	234	234
Fixed effects	R-Y	Y

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors.

Table 4.6: Monitoring intensity and perceived sanctions

VARIABLES	ASD	SFG
budget	0.00464	-0.0277**
age	-0.0859*	0.0420*
production	-0.285**	-0.202**
maintenance	-0.00252	-0.00148
Observations	234	234
R-squared	0.076	-0.314
KP rk Wald	10.03	10.03
Wu-Hausman	0.549	8.108

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors.

4.6.2 Monitoring, perceived sanctions and safety care

The two regressions presented in table 4.6 use counts of perfectly detected and declared events as the dependent variable. The first regression shows that the intensity of local commissions' monitoring activities has no significant impact on the quantity of automatic shut-downs reported by plant managers. On the contrary, the second regression shows a small but statistically significant effect of local monitoring intensity on the occurrences of unplanned uses of safeguard systems.

Conditionally on the assumptions made in the identification section, the results obtained when using reports of ASD as a dependant variable suggest that local monitoring has no impact on perceived sanctions. This result is contrasted by the results obtained when using reports of

Table 4.7: Monitoring intensity and reliability

VARIABLES	K_m	K_f
budget	-0.000106	-0.000322
age	-0.00134	0.00122
maintenance	0.000289***	-0.000282**
production	-0.0104*	-0.0158**
Observations	234	234
Fixed effects	R-Y	R-Y
R-squared	0.460	0.091
KP rk Wald	10.03	10.03
Wu-Hausman	0.0499	0.0271

*** p<0.01, ** p<0.05, * p<0.1. Robust standard errors

unplanned safeguard events as a dependant variable, where the results indicate a small but significant increase in the sanctions perceived by local managers.

In the two regressions of table 4.7, we find no significant effect of the commissions' monitoring activity on the on reliability of nuclear reactors. As reliability is assumed to be correlated with safety efforts, these two regressions fail to reject the hypothesis that the level of safety care exerted by plant managers remains constant when the intensity of the local monitoring increases. Overall, we conclude from the results gathered in tables 4.6 and 4.7 that the intensity of the monitoring performed by local commissions has a small effect on perceived sanctions, but no significant effect on safety care.

The two estimations presented in table 4.6 show mixed results regarding the age of nuclear reactors, which is positively correlated with the declarations of SFG events, but negatively correlated with automatic shut downs. Conversely, table 4.7 shows that age has no significant effect on reliability. In all four regressions, these results are small when compared to the results on production which significantly decreases the quantity of reports and the reliability of nuclear station. The results on reliability are coherent, given the fact that production and reliability ratios are closely related. The relation between production and declarations of PDD events is also coherent with the fact that significant safety events seldom occur when the power station is producing electricity at its nominal capacity.

4.6.3 Monitoring and compliance

Given the results presented in tables 4.4 and 4.6, observing an increase in the number of events reported can only be consistent with an increase in the expected penalty faced by nuclear plant managers when deciding not to report a safety event. Indeed, under constant expected penalties, the observed increase in perceived sanctions can only be consistent with a decrease in reports. Moreover, under the previous observation that monitoring intensity does not significantly affect safety care[26], the only way to observe an increase in reports is to have an increase in expected penalties, leading to an increase in compliance.[27]

Given the results presented in table 4.7, safety care is constant when monitoring intensity increases. As a consequence, the coefficient β_{budget} obtained in the second table can be interpreted as the total effect of monitoring intensity on compliance. From the last column of table 4.4, we conclude that a budget raise of 1.000€ leads to an average increase of the annual number of reports of 0.13. Compared to the reports made by the reactors included in our sample, this amounts to a 1% increase in compliance with declaration guidelines. As noted in the previous section, if $\frac{\partial E}{\partial i} \neq 0$, then the quantitative interpretation of β_{budget} is still a lower bound on the effect of monitoring intensity on compliance.

4.6.4 Endogenous detection abilities

These results rely on several assumptions. First, we neglect the fact that all French plant managers work for a single firm, which could provide them with collective incentives that are not captured in our formal model, or accounted for in our empirical estimation framework. Second, we assume that safety care cannot be specific to a particular type of event, and that the ratio of the number of events undetected by the operator to the total number of events occurring in a power station is independent from the level of care exerted by the operator[28].

Assume now that the agent may also choose (at no cost) any detection ability ρ between 0 and 1. Then, all the results derived in section 3 can be applied to a new variable $Z' = z\rho$. Z' captures the *transparency* of nuclear plant managers, defined as the combination of their compliance rate and of their detection abilities.

The identification strategy described above is then valid, provided one replaces all mentions of z by the new variable Z', and substitutes transparency to compliance. Likewise, our empirical results can be reinterpreted by replacing the level of compliance z by the level of transparency Z':

[26]We stress the fact that the absence of a significant effect of monitoring intensity on safety care combined with significant effects on perceived sanctions and expected penalties are not incompatible with proposition 4.3.1. Indeed, if an increase in safety care has a very small effect on the observable number of events, then it is possible that our proxy for safety care is just to coarse to enable us to really measure the effect of increases in perceived sanctions and expected penalties on safety care.

[27]Indeed, a decrease in expected penalties would lead to a decrease in compliance, which would not be consistent with the observed increase in reports combined with the observed constant level of safety care.

[28]This is different from saying that safety care as no impact on the detection of safety events. In our model, exerting safety care diminishes both the total number of events and the number of events undetected. Yet, we assume that the ratio of these two quantities remains constant.

plant managers react to the incentives provided by local monitoring, and if they don't seem to increase their level of safety care, they show an increase in transparency, which can be explained by either an increase in detection abilities or by an increase in compliance. The unobserved nature of these two choice parameters makes us unable to distinguish them empirically.

4.7 Conclusion and policy implications

This chapter empirically studies the effect of a French informational policy that organizes the monitoring of nuclear power stations by dedicated local commissions. These commissions can provide no monetary incentives to local plant managers, but can communicate the results of their monitoring activities to local populations and to the safety authority. This chapter investigates whether this policy induces changes in the behaviour of nuclear plant managers.

To do so, we design an original empirical strategy to identify the causal impact of this policy on the choices of local managers regarding both safety care and compliance with self-reporting guidelines. We first clarify the incentives faced by managers when exerting safety care and reporting significant safety events. We consider two main drivers of these decisions: the perceived sanctions incurred for reporting a safety event, and the expected penalty faced when an unreported event is discovered by the safety authority. Using this formal model, we estimate the effect of increased monitoring on safety care and compliance using an instrumental variable method. Our instrument, defined as budget forecasting errors at the Department level, aims to correct the possible biases induced when considering annual budgets of local commissions as a proxy for the intensity of their monitoring activities. This instrument exhibits several interesting features that qualify it as a quasi-experiment.

Empirically, we study the effect of monitoring intensity on three observables: the annual level of reliability of nuclear reactors, the quantity of events declared per reactor and per year, and the quantity of events declared that belong to a certain category of events, characterized by perfect detection and declaration. Using this latter observable, we identify the effect of monitoring intensity on perceived sanctions. Results suggest that the sanctions perceived by plant managers only marginally increase under increased monitoring intensity. Second, we find that increased monitoring intensity leads to an increase in the total number of declarations of safety events, and to no significant changes in the reliability of nuclear reactors. We conclude from these observations that local monitoring significantly increases the expected penalty faced by managers who decide not to comply with reporting schemes, which in turn induces an increase in their compliance with self-reporting criteria. Quantitatively, a budget raise of 2.000 € leads to a one percent increase in the number of events reported.

These results rely on several strong assumptions. First, we neglect the fact that all French plant managers work for a single firm, which could provide them with collective incentives that are not captured in our formal model, or accounted for in our empirical estimation framework. Second, we assume that safety care cannot be specific to a particular type of event, and that the ratio of the number of events undetected by the operator to the total number of events occurring in a power station is independent from the level of care exerted by the operator[29]. When relaxing this assumption, our empirical evidence can still be interpreted as a positive and significant impact on the overall *transparency* of nuclear managers, where transparency is defined as the combination of a manager's ability to detect safety events and his propensity to declare them.

Quantitatively, we estimate in appendix C.2 the effects of a policy that would consist in equating the budget of local commissions under a fixed total public expenditure. In other words, we estimate the number of reports that our fitted model would predict if the budgets of all local commissions were equal to the actual average budgets. We perform these predictions based on the data of year 2014, and following the results of a reactor fixed effects Poisson control-function estimator, presented in appendix C.1.[30] These out-of-sample predictions suggest that this policy would lead to an increase in the average number of events reported in the fleet. Indeed, in 2014, the 48 reactors considered in this study reported 11.9 events on average, whereas our results would predict an average 13.6 reports per reactor for the year 2014. Over these 48 reactors, this amounts to an additional 82 events reported in a single year.

Regarding policy implications, we believe these results call for optimism, as they suggest potentially cost-effective ways of improving the institutional design of nuclear safety regulation. At a national level, there has been a debate in France regarding the subsidies given to local monitoring commissions. Indeed, the French law passed in 2006 mentioned that Departments ought to provide their local monitoring commissions with a fixed share of a special tax imposed on nuclear installations. As of this writing, this article of the law is yet to be implemented. Our findings suggest that implementing this measure could foster significant improvements in the

[29]This is different from saying that safety care as no impact on the detection of safety events. In our model, exerting safety care diminishes both the total number of events and the number of events undetected. Yet, we assume that the ratio of these two quantities remains constant.

[30]Table C.4 presents the results of several reactor fixed effects specifications. We chose to use the Poisson control function estimator as it is a count model, which prevents negative predictions. In addition, the coefficient of the treatment variable in this case is positive and highly significant. It is also smaller than the coefficient of the GMM IV-Poisson estimator presented in the same table.

level of compliance of nuclear operators. More generally, our results suggest that informational policies such as this French local monitoring policy could be used as efficient complements to traditional command-and-control nuclear safety regulation mechanisms, in France or abroad.

Résumé du chapitre 5

Ce chapitre propose une analyse par la théorie des jeux des stratégies de communication adoptées par les gouvernements ou des régulateurs afin de répondre à des crises. En particulier, ce chapitre porte sur les situations dans lesquelles un gouvernement, informé de l'intensité d'un événement potentiellement catastrophique, doit informer sa population de la conduite à tenir afin de faire face à cet événement, mais redoute certains effets externes de la réaction de la population (paniques, épidémies...). Le modèle proposé étend à une audience hétérogène le jeu de cheap-talk présenté par Crawford et Sobel (1982), mais en conserve les propriétés classiques. En particulier, en raison de l'effet externe des réponses individuelles à la catastrophe, les échelles d'intensités apparaissent comme des stratégies crédibles afin d'avertir les populations. Les résultats de l'analyse sont conformes à ceux obtenus par la littérature concernant la communication de crises. Par ailleurs, le modèle se prête également à l'analyse de la coordination de la communication de crises avec de nombreuses mesures de prévention et de réponses aux désastres, comme par exemple l'utilisation de fonds publics d'indemnisation des victimes.

CHAPTER 5

Communicating Disasters

5.1 Introduction

In this chapter, we investigate the communication strategies adopted by governments, public officials or safety regulators in the wake of major disasters. More precisely, we consider harmful events during which the actions that individuals would like to take in order to protect themselves differ from those a planner would like them to take in order to minimize the total consequences of the event. Nuclear accidents, tsunamis, earthquakes, pandemics, wars or financial crises fit this description. Preventing people from trampling over each other when evacuating from fires, requiring farmers to kill livestocks to prevent the diffusion of diseases, limiting access to retail bank desks after a financial crisis to avoid bank runs, or recommending people to stay home to avoid exposure to a nuclear fallout are cases in which the government tries to convince individuals not to take an action they may have taken otherwise.

In these situations, a salient issue is the credibility of the communication strategy chosen by governments. Indeed, when major disasters occur, governments may have incentives to manipulate the information they disclose to populations in order to avoid socially costly outcomes such as panic, or to foster pro-social behaviours. Yet, if the population anticipates this manipulation, it is unclear whether or how the government will manage to persuade individuals to take any particular action. In other words, for a government informed of an incoming catastrophe, distorting information may be done at the expense of the trust of the public towards the messages and recommendations made by its governing bodies.

This trade-off has been discussed in the crisis communication literature, a review of which can be found in a recent report for the European Commission on the management of rare disasters (Pljansek et al., 2017). Based on case studies and experiments,[1] this crisis communication

[1] For instance, Haynes et al. (2008) studies trust in volcanic risk communication during the 2008 Montserrat eruption. Maidl and Buchecker (2014) analyse the flood preparedness policies implemented in Zurich, Switzerland.

literature claims that governments should foster long-term trust in the transparency of their crisis communication rather than profit from short-term benefits obtained by downplaying risks. It also claims that last-mile communication - i.e. differentiated communication strategies with each affected community - should be used to account for the heterogeneity of local sub-cultures towards catastrophic events. Finally, this literature claims that two-way communication between affected communities and the government is an effective way to improve the response of populations to disastrous events.

In this chapter, we aim to shed a new light on crisis communication using a game-theoretic approach, which allows us to test the robustness of the conclusions drawn by the literature stated above to a stylized analysis of the strategies of both the government and the population. To do so, we develop a cheap-talk model with multiple audiences. A sender - the government - is informed of the extent of a catastrophe, and communicates publicly with a continuum of receivers - the population - equally exposed to harm, but characterized by heterogeneous costs of avoiding harm. When the receivers' actions are characterized by positive or negative externalities, the equilibria of this game are structurally identical to the equilibria presented by Crawford and Sobel (1982) (CS in the following). We use these equilibria to study several government interventions aiming to prevent, prepare for or respond to disastrous events. We do so by allowing in turn the sender to use private communication channels, to invest in the mitigation of the consequences of the disaster, or to use monetary transfers to shift the incentives of the receivers.

Our analysis confirms the conclusions of the crisis communication literature. The structure of the communication equilibria confirms that considering the response of populations is paramount in the design of disaster communication strategies. Moreover, it appears that transparency matters, as under both positive and negative externalities, the most informative equilibria always lead to the largest expected welfare. We also show that private communication (i.e. differentiated modes of communication between the sender and each receiver) always extends the scope of communication between the government and the population. In other words, when public communication exists, some level of private communication is possible, and increases welfare. In addition, there exist cases in which public communication is impossible, but private communication can be sustained and improves expected welfare. This result differs from the classical multiple audience literature in which private communication is sufficient to ensure public com-

---

Boer et al. (2014) present the result of a risk communication experiment near Rotterdam in the Netherlands. Engel et al. (2014) describes the variety of local sub-cultures regarding flood risks and flood risks communication in the Netherlands.

munication (see e.g. Farrell and Gibbons (1989)). Yet, it confirms that fostering private modes of communication between authorities and affected communities is a sound aim in the wake of a major disasters.

Our analysis also yields some new insights into the coordination of crisis communication strategies with disaster management actions. First, it appears that preparedness does not always improve communication credibility. When trying to avoid panic, reducing the extent of the negative externalities associated with individual actions (for instance by building larger tsunami escape roads) aligns the incentives of the sender with those of the population and increases welfare. On the other hand, fostering pro-social behaviours also increases welfare, but shifts the incentives of the government away from those of the receivers, reducing its credibility. When preparedness is costly to the government, multiple optimal investments may be faced before engaging in cheap-talk. Selecting among these solutions may require the government to trade-off ex-post welfare with communication transparency.

Finally, we show that monetary transfers can interact with the incentives of the cheap-talk game. In other words, if a government can commit funds to the curtailment of the consequences of a disaster, we question how these potential transfers can change the equilibria of the communication game. When the government can commit ex ante to perform monetary transfers based on individual actions, and when individual actions entail positive externalities, it is optimal to subsidize these actions. Conversely, when externalities are negative, it is optimal for the government to provide partial public insurance to those who cannot protect themselves. This result contributes to the literature on the Samaritan dilemma, which argues that public insurance may foster careless behaviours from agents exposed to the risks of rare disasters (such as building houses in flood-prone areas, see e.g. Coate (1995); Kunreuther (1996); Epstein (1996); Kunreuther (2006); Kunreuther et al. (2013) or Teh (2017)). Our result shows that partial public insurance can be beneficial to the government during the communication phase, as it allows more transparent communication strategies and better individual responses.

Although the economics literature on disaster communication is scarce, this chapter participates to a large body of applications of cheap-talk models to economic situations in which informed parties can transfer unverifiable information to uninformed agents whose actions matter for the informed player. The classical CS model with one sender and one receiver has been used for instance to describe the communication strategies used by central banks to advertise their objectives and future policies (Stein, 1989; Moscarini, 2007). Likewise, Kawamura (2011) analyses public consultation procedures by designing a cheap-talk model in which multiple interested senders try to weigh in the decision of a single public receiver. Allon and Bassamboo (2011) use cheap talk to study the provision of information by a retailer to a set of homogeneous customers in a dynamic setting.[2] The paper closest to ours is Allon et al. (2011), who analyse the provision of information by firms to queueing customers, where queuing determines the costs faced by the firms and the utility derived by customers. The models used in this paper and ours are similar: they both entail binary actions characterized by external effects.[3] Our chapter differs as we explicitly deal with the heterogeneity of receivers, and as the actions taken by receivers do not affect the state of nature in our setting. In addition, the interplay between committed transfers and the equilibrium of the cheap-talk game is a feature that is absent from the aforementioned references.[4]

This chapter is organized as follows. Section 5.2 presents the game and characterizes its equilibria. Section 5.3 studies the effect of ex ante investment in preparedness on the equilibria of the cheap-talk game. Section 5.4 compares private and public communication. Section 5.5 explores the possibility for the sender to engage in various types of monetary transfers, and characterizes their effect on communication equilibria. Section 5.6 concludes.

5.2 A cheap-talk model with multiple audiences

5.2.1 The model

The game involves one sender (S or she in the following), and a continuum of receivers of type θ uniformly distributed over $\Theta = [0;1]$. The sender is uninformed about the type of each receiver, but she is privately informed of the value of a random variable r that characterizes the extent of a catastrophe. This random variable can equivalently be thought of as the type of the sender. The common prior shared by the sender and the receivers is that r is distributed uniformly over $[0;1]$. Each receiver has to choose an action a in $\mathcal{A} = \{0,1\}$.

The timing of the game is as follows. First, nature draws the state of the world $r \in [0;1]$. The sender observes r at no cost. Then, the sender sends an unverifiable public message perfectly

[2] Additional uses of cheap-talk models have been made outside of economics. Political scientists such as Trager (2010) and Ramsay (2011) have used cheap-talk models to study diplomacy and international crises. They show that diplomacy can be used to deter wars. Cheap talk has also been used in ecological economics to analyse animal forms of multi-modal communication (Wilson et al., 2013).

[3] In Allon et al. (2011), customers choose whether to queue, thus affecting the expected waiting time for others. Receivers in our model can choose whether to seek protection from a catastrophe, but seeking protection entails a social cost.

[4] A similar feature can be found in an unpublished working paper by Antic and Persico (2016), who study the competition for information provision among shareholders. In their paper, investors interact on the market for shares and then communicate on a risky venture. In this game, the action of an informed investor on the market determines his credibility during the communication phase.

observed by all receivers. The sender cannot commit ex ante on this communication strategy. Upon the reception of the message, receivers update their prior over the state space, and choose their actions. Finally, pay-off is realized.

The preferences of a receiver of type $\theta \in \Theta$ are described by the following utility function:

$$u_R(a, \theta, r) = \begin{cases} -r, & \text{if } a = 0 \\ -\theta, & \text{if } a = 1 \end{cases} \tag{5.1}$$

The state of nature thus symbolizes the damage incurred by individuals, while θ embodies the private cost of avoiding harm. This specification of the preferences of the receivers implicitly assumes that all receivers are equally exposed to harm, but have heterogeneous costs of avoiding this harm.

In addition, we assume that the actions taken by the receivers have a social cost, that we note $\gamma \in \mathbb{R}$. More precisely, we assume that the preferences of the sender regarding the choice of receiver θ can be represented by:

$$u_S(a, \theta, r) = \begin{cases} -r, & \text{if } a = 0 \\ -\theta - \gamma, & \text{if } a = 1 \end{cases} \tag{5.2}$$

The parameter γ can be both positive or negative, as the response of an individual to a disaster can have both positive and negative external effects on other individuals. For instance, people rushing to their cars to flee from a nuclear accident will likely create traffic jams, reducing the overall speed of evacuation from danger. On the other hand, destroying irradiated food or killing contaminated livestocks in the aftermath of a nuclear accident lead to a possible reduction of the overall exposition of the population to nuclear radiations.

Under complete information regarding the value of r, the receiver of type θ chooses $a = 1$ if and only if $r > \theta$, while the sender would rather have him make this choice if and only if $r > \theta + \gamma$.

When only the sender is informed about the value of r, she can design a communication strategy, e.g. a mapping from the set of states to a set of public signals that we note $\mathcal{M}$. We assume that communication is non-committed and payoff-irrelevant. We note $m(r)$ the message sent by the sender in equilibrium after observing r. Conditionally on the information revealed by the sender, any receiver takes action $a = 1$ if and only if the private cost of avoiding harm is lower than the expected damage incurred, obtained by revising the prior according to Bayes rule :

$$\mathbb{E}(r|m(r)) > \theta. \tag{5.3}$$

As communication is public, and as receivers share a common prior regarding the value of the random variable r, condition (5.3) implies that any equilibrium outcome of the public communication game can be described by a function $\theta^\star(m(r))$ such that action $a = 1$ is chosen by all receivers whose type θ is lower than $\theta^\star(m(r))$, while all receivers whose type is larger than $\theta^\star(m(r))$ choose action $a = 0$.

In this framework, the set of message is irrelevant, and only the mapping between states of the world and the actions chosen by the receivers is necessary to characterize an equilibrium. In other words, the message function chosen by the sender is equivalent, in equilibrium, to providing a recommendation regarding the equilibrium outcome, such as: *"given the information we have, we recommend to all individuals below type $\theta^\star$ to seek shelter"*.[5] Hence, in the following, we note $\theta^\star(r)$ the equilibrium outcome obtained when the sender sends the message $m(r)$.

Thus, in the following of this chapter, the welfare associated with the state of nature r and a communication strategy leading to the equilibrium outcomes $\theta^\star(\cdot)$ is:

$$W(r, \theta^\star(r)) = -\int_0^{\theta^\star(r)} (\theta + \gamma)d\theta - \int_{\theta^\star(r)}^1 r d\theta \tag{5.4}$$

and the expected welfare for the sender is finally defined as $\mathbb{E}W(\gamma) = \int_0^1 W(r, \theta^\star(r))dr$.

This environment is a classical cheap-talk model with constant bias. The equilibria of this game thus share many similarities with the properties enunciated in Crawford and Sobel (1982). We now review the properties of these equilibria.

5.2.2 Communication equilibria

Properties of communication equilibria

In the following, we focus on pure communication strategies and the solution concept used to solve the game is Bayes-Nash equilibrium. For the sender, the game consists in choosing a communication strategy that is credible. In other words, given a message function $m : \Omega \rightarrow \mathcal{M}$

[5]This is consistent with observed disaster management practices recently witnessed in Florida in preparation for Hurricane Irma. Although over 5 million people were ordered to evacuate, some people such as hospital patients with limited mobility were not evacuated. In addition, civil servants such as policemen or fire-fighters were not allowed to evacuate. This piece of anecdotal evidence is documented in two press releases in the New York Times and Slate.

chosen in equilibrium, m has to satisfy the following incentive compatibility condition:

$$\forall r \in [0;1],\ m(r) \in \underset{m(r')}{\operatorname{argmax}}\, W(r, \theta^\star(m(r'))) \tag{5.5}$$

In other words, equation (5.5) states that in equilibrium, the sender must have no incentives to deviate from the communication strategy chosen $m(r)$.

Conditionally on a credible message $m(r)$ sent by the sender, each receiver chooses an action that maximizes his expected pay-off in equilibrium:

$$a(r, \theta) = \underset{a}{\operatorname{argmax}}\, \mathbb{E}\left[u_R(a, r, \theta) | m(r)\right]. \tag{5.6}$$

Lemma 5.2.1 states the main properties of communication equilibria, which are identical to the ones derived by CS. The proof follows the same steps.

Lemma 5.2.1. *Communication equilibria are necessarily monotonic and only involve a finite number of messages.*

Proof. In equilibrium, $\theta^\star$ is necessarily increasing in r. To see this, suppose there exist $\theta_1^\star$ and $\theta_2^\star$ induced in equilibrium such that $\theta_1^\star < \theta_2^\star$. Then, there is r such that S is indifferent between $\theta_1^\star$ and $\theta_2^\star$, and the function $W(r, \theta_1^\star) - W(r, \theta_2^\star)$ is decreasing in r. Therefore, for any r_1 and r_2 that respectively induce $\theta_1^\star$ and $\theta_2^\star$, we have $r_1 < r_2$.

Next, notice that as $\theta^\star$ is increasing, it has to be continuous and differentiable almost everywhere on $[0;1]$. Then, suppose there exists a point r_0 at which $\theta^\star$ is differentiable and its derivative is not null. Then, the belief held by all receivers upon reception of the message associated with this point is the singleton $\{r_0\}$, which implies that $\theta^\star(r_0) = r_0$. Then, there exist $\epsilon > 0$ such that $\theta^\star(r_0 + \epsilon) \geq r_0$ and $r_0 + \epsilon - \gamma < r_0$. When $r = r_0 + \epsilon$, the sender has an incentive to deviate from $m(r_0 + \epsilon)$, as she is strictly better off when sending $m(r_0)$.

Hence, $\theta^\star$ is constant wherever it is continuous, which implies that equilibria involve at most a countable number of messages. Finally, all equilibrium outcomes induced in equilibrium have to be separated by at least $|\gamma|$, or they would otherwise violate the receivers' Bayesian rationality constraint. This yields that all equilibria involve finitely many actions. □

Hence, for any $n \in \mathbb{N}$, possible communication equilibria are constituted of a finite and increasing set of equilibrium outcomes $(\theta_i^\star)_{1 \leq i \leq n}$, respectively induced over intervals of the state space noted $[r_{i-1}; r_i]$, for all $i \leq n$, $i \neq 0$, and where $r_0 = 0$ and $r_n = 1$.

Babbling equilibria

As in other cheap-talk frameworks, our game always has a babbling equilibrium, e.g. an equilibrium characterized in the previous description by $n = 1$. In this equilibrium, the sender sends a single message, irrespective of the state of the world. This equilibrium is uninformative, as each receiver can do no better than reacting according to the common prior. All receivers of type $\theta < \frac{1}{2}$ play action $a = 1$ whereas receivers of type $\theta > \frac{1}{2}$ play action $a = 0$. It is clear that the sender can never profitably deviate from the communication strategy chosen, and that the response of the population is optimal. The expected welfare associated with the babbling equilibria is $\mathbb{E}W_{bab}(\gamma) = -\frac{\gamma}{2} - \frac{3}{8}$.

The babbling equilibrium of this game can be thought of as a world in which the government has completely lost credibility, and people no longer listen to the government's recommendations. Another interpretation of this hypothesis is a world in which the incoming catastrophe wiped out all telecommunication lines. Then, people have no outside information regarding the extent of a catastrophe, and can only base their decision on their prior and their own private cost of taking protective actions.

After the Fukushima-Daiichi nuclear accident, the absence of communication from the government during the days that followed the catastrophe led people not to trust the announcements made after a few days, and to take maladjusted actions which led to excessive exposures to irradiations or psychologically-driven migrations (see e.g. Figueroa (2013) or Zhang et al. (2014)). Perko (2011) provides additional pieces of evidence of the adverse effects of the distrust of populations in their governments based on a case study of the Three-Mile Island accident, which was followed by a massive panic reactions from the population living near the power station. Therefore, a salient question which we address in the following paragraphs is whether an informed government can, in general, do better than this uninformative equilibrium.

Information transmission

To address this question, we now look for equilibria in which the sender transmits information to the receivers (e.g. for which $n > 1$).

Following the characterization stated above, an increasing sequence of equilibrium outcomes $(\theta_i^\star)_{1\leq i\leq n}$ and a set of associated intervals defined by a sequence of thresholds $(r_i)_{0\leq i\leq n}$ constitute an equilibrium if and only if they satisfy the sender's incentive compatibility constraint, and the receivers' Bayesian rationality constraint.

Regarding the receivers Bayesian rationality constraint, and conditionally on the reception of a credible message over any interval $[r_{i-1}; r_i]$, the public belief obtained by Bayes' rule is uniform over this interval. Then, the equilibrium outcome has to satisfy the following necessary condition:

$$\forall r \in [r_{i-1}; r_i], \theta^\star(r) = \theta_i^\star = \frac{r_{i-1} + r_i}{2}. \tag{5.7}$$

In addition, thresholds $(r_i)_{0<i<n}$ and equilibrium outcomes $(\theta_i^\star)_{1\leq i\leq n}$ have to satisfy the sender's incentive compatibility constraint:

$$\forall i < n,\ W(r_i, \theta_i^\star) = W(r_i, \theta_{i+1}^\star) \tag{5.8}$$

Developing equation (5.8) using equations (5.7) and (5.4) yields the following condition:

$$\forall i < n,\ r_{i+1} - r_i = r_i - r_{i-1} - 4\gamma \tag{5.9}$$

Condition (5.9) is identical to the findings of CS[6]. From this condition, lemma 5.2.2 characterizes all the equilibria of the game.

Lemma 5.2.2. *For all values of* γ*, and all* $n \in \mathbb{N}$*, there exists a unique equilibrium characterized by* n *equilibrium outcomes* $(\theta_i^\star)_{1\leq i\leq n}$ *and* $n+1$ *thresholds* $(r_i)_{0\leq i\leq n}$ *if and only if* $|\gamma| < \frac{1}{2n(n-1)}$*. In addition, we have that* $\forall i \leq n$*,* $r_i = \frac{i}{n} + 2\gamma i(n-i)$*, and* $\theta_i^\star$ *is induced over* $[r_{i-1}; r_i]$*.*

Proof. If an equilibrium characterized by n equilibrium outcomes exists, then $r_i = \frac{i}{n} + 2\gamma i(n-i)$ is a necessary condition implied by equation (5.9). This condition ensures uniqueness of the equilibrium. In addition, the equilibrium exists if the sequence of thresholds defined above is increasing. When $\gamma > 0$ and $n \geq 2$, the sequence of thresholds is increasing if and only if $r_{n-1} < r_n = 1$. Likewise, for $\gamma < 0$ and $n \geq 2$, the n-message equilibrium is increasing if and only if $0 = r_0 < r_1$. These two conditions yield the result stated in the lemma. □

Hence, for any value of γ, there is a finite number of credible communication strategy, involving from 1 to $N(\gamma)$ equilibrium outcomes. In the following, and for all values of γ, we refer to the equilibrium characterized by the largest number of equilibrium messages as the *most informative equilbrium.* The uniqueness of the equilibrium associated to a any given pair

[6] Our result matches those from CS up to the sign of γ, which is negative in CS, and can be both positive or negative here. It can also be noted that if the communication strategies that are credible in equilibrium are the same in CS and in our game, the welfare levels derived by the sender are different, ex ante and ex post.

$(\gamma;n)$, where $n < N(\gamma)$ makes it interesting to derive the welfare expected from the game characterized by γ and the unique communication strategy leading to n equilibrium outcomes. Expected welfare is defined here as the expected utility of the sender from its interaction with all receivers:

$$\mathbb{E}W(\gamma,n) = \sum_{i=1}^{n} \int_{r_{i-1}}^{r_i} \left(- \int_0^{\theta_i^\star} (\theta+\gamma) d\theta - \int_{\theta_i^\star}^{1} r d\theta \right) dr \tag{5.10}$$

The expected utility of the receivers, $\mathbb{E}U_R(\gamma,n)$ is defined by replacing in equation (5.10) the social cost of action $\theta+\gamma$ by its private counterpart θ. Lemma 5.2.3 presents the explicit form of the two quantities $\mathbb{E}W$ and $\mathbb{E}U_R$. The calculations leading from equation (5.10) to these quantities are omitted.

Lemma 5.2.3. *For any γ, the expected welfare and expected receivers' utility associated with an existing n-message equilibrium are equal to:*

$$\mathbb{E}W(\gamma,n) = -\frac{\gamma^2(n^2-1)+3\gamma+2}{6} - \frac{1}{24n^2} \tag{5.11}$$

$$\mathbb{E}U_R(\gamma,n) = -\frac{\gamma^2(n^2-1)+2}{6} - \frac{1}{24n^2} \tag{5.12}$$

Equations (5.11) and (5.12) lead to the following proposition.

Proposition 5.2.1. *For any γ, the highest level of welfare is obtained under the most informative equilibrium. Likewise, the expected utility of the receivers is also increasing in the number of messages used in equilibrium.*

Proof. To see the first part of the result, notice that $\forall i \in \mathbb{N}$, $W(i, \frac{1}{2i(i-1)}) = W(i-1, \frac{1}{2i(i-1)})$, that $W(i, \frac{-1}{2i(i-1)}) = W(i-1, \frac{-1}{2i(i-1)})$, and that $W(i,\gamma) - W(i-1,\gamma)$ is a quadratic polynomial in γ, with negative first coefficient. Hence, when the equilibrium with i messages exists, it dominates the equilibrium with $i-1$ messages. The same reasoning holds for the second part of the result. □

The first part of this proposition is usual. The more informative the communication strategy chosen in equilibrium, the larger the expected utility derived by the sender from the game. Conversely, although the second part of proposition 5.2.1 is conform to CS, its interpretation differs because of our multiple audience setting. Here, there are receivers that are strictly better off when less information is transmitted by the sender. For instance, for any γ, consider an equilibrium characterized by $i < N(\gamma)$ equilibrium outcomes $(\theta_j^\star)_{1\leq j\leq i}$. Then, the receivers characterized by the types corresponding to these equilibrium outcomes cannot profit from the disclosure of more information, as they are already provided with exactly the information they need to take their preferred action.

This is so because the expected utility $\mathbb{E}U_R$ captures the expected utility averaged over all receivers. We can interpret this quantity as the expected utility considered by any receiver before learning his own type. For instance, if the sender and the receivers agree ex ante on which equilibrium will be played after the receivers learn their types θ and the sender learns r, then the fact that both $\mathbb{E}U_R(\gamma, n)$ and $\mathbb{E}W(\gamma, n)$ are increasing in n provide strong support for choosing an equilibrium selection criterion which always selects the most informative equilibrium.[7] This interpretation is consistent with our catastrophe setting, as disaster communication plans and mitigation strategies are usually determined ex ante, when people do not know when a catastrophe will strike, nor how able they will be to avoid it when it occurs.[8]

Hence, in the following of this chapter, we consider that conditionally on the value of γ, the most informative equilibrium is always selected.

Application to severity scales

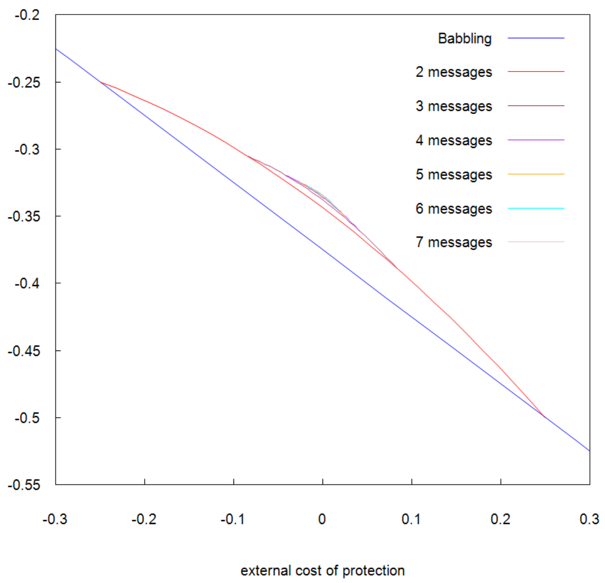
The expected welfare associated with the equilibria characterized above are plotted on figure 5.1. Under both positive and negative externalities, informative public communication is impossible when $|\gamma| > \frac{1}{4}$, and only the babbling equilibria is credible for the government. This means that when preferences are sufficiently biased, the government cannot do better than to remain silent in public. Conversely, when credible and informative communication is possible, all equilibria consists in a partition of the state space. In other words, the sender defines a finite number of messages, each of which correspond to a range of possible catastrophes of similar severities. Depending on the value of the external effects associated with individual actions, the number of messages sent by the sender changes. In particular, the lower the external effects of actions (i.e. $|\gamma|$), the more informative can be the sender's communication strategy.

This casts a new light on the incentives faced by a government engaging in crisis communication. Contrarily to the usual trade-off presented in the crisis communication literature, in which

[7]An equilibrium selection criterion that allows to select the most informative equilibrium for each γ is, for instance, the "no-incentive-to-separate" condition developed by Chen et al. (2008).

[8]$\mathbb{E}U_R$ can alternatively be interpreted as the welfare measure that would be used by a populist regulator, following the terminology defined in the *Happyville* literature (see e.g. Portney (1992); Salanié and Treich (2009)). Within this terminology, we could interpret θ and $\theta + \gamma$ as the subjective perceptions respectively held by the receivers and the sender regarding the ability of a receiver to avoid a given catastrophe. This would be consistent with the analysis of Hasegawa (2013), who stresses the role of public perceptions in the consequences of the Fukushima-Daiichi accident. Within this framework, we adopt a standpoint in which the sender is acting in a paternalistic way, i.e. she acts according to her own perception of the cost of avoiding the catastrophe but accounts for the strategic reaction of the biased receivers.

Figure 5.1: Expected welfare for communication equilibria for $n \leq 7$ and $\gamma \in [-0.3; 0.3]$.



governments must choose between downplaying risks and fostering trust, it appears here that in all equilibria, governments have no incentives to downplay risks systematically. Doing so would result in a proportional adaptation of the response of the population. On the other hand, the structure of the equilibria of the communication game suggest that governments should rather engage in *vague* communication strategies, pooling catastrophes of different severities to foster socially optimal behaviours while remaining credible.

There are multiple examples of vague modes of communication characterized by constant signals over large ranges of possible harm. In the nuclear industry, the International Atomic Energy Agency defines an International Nuclear Event Scale (INES in the following) that ranks nuclear events in 8 categories, separated by a tenfold increase in material consequences.[9] Likewise, the U.S. Nuclear Regulatory Commission (NRC in the following) defines four emergency messages to warn populations after abnormal events in nuclear power station. The most benign events are communicated in *notifications of unusual event*, while larger incidents are communicated through an *alert*. The most severe events are communicated as *site area emergencies* and *general emergencies*. Similar communication strategies are used to signal other types of rare disasters, such as hurricanes intensity which are usually measured on the Saphir-Simpson scale which contains five categories defined by wind speeds, or avalanche risks which are usually advertised using coloured flags.

[9] See e.g. D'Haeseleer (2013) or the International Atomic Energy Agency's website dedicated to the INES scale.

Another interesting takeaway from our model is that the structure of these message scales changes with the sign of the externalities. Under positive externalities, the range of the intervals increases when the catastrophe worsens. Under negative externalities, the larger is the catastrophe, the narrower the intervals get.

This feature is directly related to the credibility of the communication strategy. Under positive externalities, receivers expect the sender to over-report the risks associated with a given event, leaving her no other choice than to pool more high states together than low states to restore credibility. Indeed, if two consecutive messages were sent on two ranges of harm of equal lengths, receivers would expect the sender to deviate from this communication strategy when the states of the world are just below the threshold separating both messages. Likewise, under negative externalities, receivers expect the sender to under-report the risks. Hence, following the same reasoning, two consecutive messages sent in equilibrium have to be sent on intervals of decreasing width to restore credibility.

When the severity of a catastrophe can be measured according to some dimension related to its consequences for the population, this characteristic of credible communication strategies can be used to analyse real disaster communication strategies. In the case of nuclear power, the range of potential harm associated with each level of the INES scale is increasing as two consecutive levels of the INES scale are supposed to capture a tenfold increase in the consequences of nuclear events. According to our model, this scale is thus a good candidate for a communication strategy when individual actions are characterized by positive externalities. On the other hand, when trying to avoid panic, and according to our model, communicating with the INES scale is not credible. A similar critique of the INES scale is made by Wheatley et al. (2017), who challenge its transparency and adequacy to characterize the severity of nuclear accidents. A more credible communication strategy aiming at reducing the risks of panic should actually pool together minor nuclear incidents (such as events rated lower than level 4 on the INES scale), and should refine further the distinctions among larger events.

The structure of the cheap talk equilibria studied above sheds light on the incentives faced by a government during the communication phase that occurs in the wake of a major disaster. Yet, another salient aspect of disaster management is the allocation of public resources to the mitigation of catastrophes. For instance, governments can take actions to prevent, prepare for and respond to disasters. In the following sections, we question how these actions taken by governments can interact with disaster communication strategies, and how these actions can be combined to foster more effective crisis communication strategies.

5.3 Reducing externalities: the case of preparedness

The aim of the following paragraphs is to study the interplay between the actions a government may take before a catastrophe and the communication stage. Preparedness is usually defined as the action of getting a population prepared to the possibility of catastrophic events by fostering the adoption of good courses of actions. Preparedness can be increased by performing drills, engaging citizens in adapted trainings, or through the communication of good practices.

For instance, Pljansek et al. (2017) describe a case of preparedness measures taken in the island-city of Dordrecht in the Netherlands. During severe floods, only 10 to 20% of the population can be safely evacuated from the city before the water levies fail. Hence, to reduce potential casualties associated with the failures of water levies, recent preparedness programs aimed to convince people to respond to flood alerts by taking shelter in their homes or in elevated public shelters.

These programs limit the negative external effect of the actions taken by individuals in response to an incoming disaster, or increase their pro-social consequences. To study these actions and their effects on disaster communication strategies, we first model them as costless modifications of the value of the externalities induced by individual actions. We then study the case of costly modifications of γ.

5.3.1 Costless selection of externalities

Consider a more general game, in which the sender can choose ex ante (i.e. before observing r) and at no cost the extent of the externalities γ. We analyse a first case in which the sender may choose γ in the positive unit interval, and a symmetric case in which she chooses γ in the negative unit interval. This simple refinement of the game leads to the following two straightforward but interesting findings.

Corollary 1. *When the sender is constrained to choose γ in $[0;1]$, the highest level of expected-welfare is obtained when choosing $\gamma = 0$. When γ has to be chosen in $[-1;0]$, the highest level of expected-welfare is obtained when choosing $\gamma = -1$.*

Proof. For any value of γ, the sender-preferred equilibrium[10] of the cheap-talk is considered. Using proposition 5.2.1, the expected welfare[11] associated with the cheap-talk game under ex-

[10] As discussed earlier, this makes sense as the equilibrium preferred by the sender for any γ is both the most informative equilibrium and the equilibrium preferred by all receivers before learning their types.

[11] One could equivalently define $W_{\text{ex ante}}(\gamma)$ as the closure of the union of the convex hull of the graphs of $\mathbb{E}W(\gamma, n)$, which can be written $W_{\text{ex ante}}(\gamma) = \sup\{z | z \in \cup_{i\in\mathbb{N}} co(\mathbb{E}W(\gamma, i))\}$.

ternalities γ is:

$$W_{\text{ex ante}}(\gamma) = \sup_{n \in \mathbb{N}} \{\mathbb{E}W(\gamma, n)\} \tag{5.13}$$

Given the fact that $\frac{\partial \mathbb{E}W}{\partial \gamma}(\frac{-1}{2n(n-1)}, n) < 0$ and $\frac{\partial \mathbb{E}W}{\partial \gamma}(\frac{1}{2n(n-1)}, n) < 0$ for all n, we have that $W_{\text{ex ante}}(\gamma)$ is strictly decreasing (and quasi-concave) in γ. □

Under negative externalities, corollary 1 shows that the outcome of the game is full communication as the preferences of individuals and the government are perfectly aligned. In this case, there is no trade-off between transparency and social welfare. This is the case described in the flood example presented above. There, having too many people evacuating by land would create traffic jams and slow down the evacuation. Convincing people to take shelter on their roofs reduces this negative externalities, and reduces the incentives for the government to manipulate the information disclosed in equilibrium.

On the other hand, when the sender is constrained to choose the external effect of the receivers' actions within $[-1; 0]$, she will choose $\gamma = -1$. In this case, the equilibrium played is the babbling equilibrium. Here, allowing individuals to have strong pro-social behaviours completely offsets the cost of the catastrophe. In particular, society strictly benefits from any receiver taking the costly action. An example of this result is the management of contaminated food stocks. In this case, it is unlikely to see the government trade-off public health with the ability to communicate, and one could expect the government to require the destruction of the whole food stock.

As a remark, not allowing the sender to choose γ over $[-1; 1]$ implicitly assumes that the sender cannot change the nature of an externality, but only influence its size through preparedness. This assumption is required to illustrate the different incentives at play under positive and negative externalities, which respectively lead the sender to choose the maximum bias and minimum bias. This assumption is relaxed in the following paragraphs.

5.3.2 Costly externality shifts

We now present a more general case, in which shifting the level of externalities is costly to the sender. We here assume that γ can be shifted over the whole $[-1; 1]$ interval. Separating the interval into two positive and negative unit intervals would not change the nature of the following result.

Let the external effect of individual actions be $\gamma_0 \in [-1; 1]$. The sender knows γ_0, and may invest, before learning r, in order to shift γ_0. The resulting level of externalities γ is then observable by the receivers. Once the investment is realized, the cheap-talk game is played as before. We assume that investing in shifting γ_0 is costly, and that this cost is increasing and convex in the shift. The cost function $\psi(|\gamma - \gamma_0|)$ is assumed to be increasing and convex in its argument.

As the first step of the game carries no information regarding the state of the world, the cheap-talk game is played as before, under externalities γ resulting from the investment of the sender. This new game can thus be solved by backward induction. Let $\mathbb{E}W_{\text{prevention}}(\gamma)$ denote the expected welfare of this new game. Then, we have:

Corollary 2. $\mathbb{E}W_{prevention}(\gamma)$ *can admit multiple global optima, which we note*

$$\Gamma = \underset{\gamma}{\operatorname{argmax}}\{\mathbb{E}W_{prevention}(\gamma)\}.$$

Proof. The expected welfare derived from the cheap-talk game is $W_{\text{ex ante}}(\gamma)$. Hence, the sender chooses γ so as to satisfy:

$$\gamma \in \underset{\gamma' \in [-1;1]}{\operatorname{argmax}} \mathbb{E}W_{\text{prevention}}(\gamma') = \mathbb{E}W_{\text{ex ante}}(\gamma') - \psi(|\gamma' - \gamma_0|) \tag{5.14}$$

Using corollary 1, the objective function defined by equation (5.14) is not necessarily quasi-concave, as the quasi-concavity of $\mathbb{E}W_{\text{ex ante}}(\cdot)$ is not preserved by the addition with $-\psi(\cdot)$. □

Hence, either $\mathbb{E}W_{\text{prevention}}(\gamma)$ has a single global maximizer (for instance, this is the case if $\mathbb{E}W_{\text{prevention}}(\gamma)$ is quasi-concave), either it has several of them.

In the first case, we note the global maximum γ_{max}. As ψ is convex and positive, γ_{max} is necessarily weakly inferior to γ_0. Intuitively, the sender always has an incentive to reduce the extent of the external effects of the receivers' actions.

In the second case, Γ contains at least two elements. Although these elements cannot be distinguished using this ex ante welfare criterion, these values of γ entail different outcomes, that trade-off the expected welfare resulting from the cheap-talk game with the cost of preparedness actions, and differ in the amount of information disclosed by the sender in equilibrium.

Choosing among these optimal solutions requires to select among their different properties. One of the properties which does not carry any value in our framework is transparency. If there are multiple solutions, some entail more messages in equilibrium than others. In particular, the element of Γ characterized by the lowest absolute value (i.e. $\operatorname{argmin}_{\gamma\in\Gamma}\{|\gamma|\}$) is the solution that leads the government to communicate in the most truthful way in equilibrium.

Two other elements of Γ have particular properties. The maximum element of Γ is the one that requires the smallest amount of investment ex ante. A budget-constrained sender may opt for this solution. On the contrary, the minimum element of Γ is the one that requires the largest amount of investment ex ante, but is also the one that yields the largest expected welfare from the cheap-talk game. This global optimum may be the one preferred by a deep-pocket sender.

An interesting takeaway from this second case is the fact that preparedness programs aiming to reduce the external effects associated with protective actions interact with the communication game. Reducing externalities does not necessarily increase the transparency of communication strategies, and may lead to several optimal solutions for a government. These optima differ in the way they trade-off the distribution of outcomes of the cheap-talk game with the investment required ex-ante.

An illustrative example of multiple preparedness measures is the case of fire evacuation procedures. Displaying maps within buildings indicating multiple escape routes is a relatively cheap preparedness measure that aims to use optimally all the exits of a building without creating slow crowds at some exits. Another usual preparedness measure that requires more efforts is the designation of specific agents in charge of checking that all people within a floor do comply with the fire alarms and effectively follow the emergency evacuation procedures. In this case, potential negative externalities associated with responses to fire alarms (i.e. slow crowds trying to escape a building) are turned into positive ones (i.e. people caring for the fact that everyone leaves the building through an assigned evacuation route).

5.4 Observable types and private communication

5.4.1 A private communication benchmark

In the previous sections of this chapter, we assumed that the sender could only engage in public communication. In the case of major disasters, this assumption may be warranted as communication networks allowing a government to communicate specifically with affected communities may have been destroyed. Another possible explanation of this assumption is that the sender cannot observe the type of each receiver, but only knows their distribution. For instance, in the case of nuclear accidents, a government may not be fully aware of how meteorological conditions will disseminate the radiological fallout.

In this section, we relax these assumptions and tackle the issue of last-mile communication. In other words, we ask whether it makes sense for a government to engage in private communication with the different populations instead of communicating publicly with all the parties affected by a disaster. To do so, we assume that the type of each receiver is common knowledge, and that the sender disposes of private communication channels with each receiver. In practice, early warning systems exists in areas prone to disasters such as floods or hurricanes. Text messages or sirens can be triggered locally when a disaster occurs. Local news channels can be used to provide different pieces of information in different areas.

Define the scope of communication as the range of values of the parameter γ over which informative communication can be sustained. Following lemma 5.2.2, the scope of public communication is $[-\frac{1}{4}; \frac{1}{4}]$, as informative equilibria could only be sustained if $|\gamma| < \frac{1}{4}$. Under this definition, the following proposition holds.

Proposition 5.4.1. *Private communication strictly extends the scope of communication between the sender and the receivers*

Proof. Under private communication, the communication of the sender with any receiver characterized by type θ can result in two equilibria: a babbling equilibrium in which no information is disclosed, or an equilibrium in which the sender discloses whether the state of nature is above or below $\theta + \gamma$. Any other communication strategy involving two distinct messages on two different intervals would not be credible to the receiver.

When $\gamma > 0$, a necessary condition for an informative private communication equilibrium to be sustained is $\theta + \gamma < 1$, otherwise only one message is sent in equilibrium. With these receivers, the communication of the sender induces two different actions when:

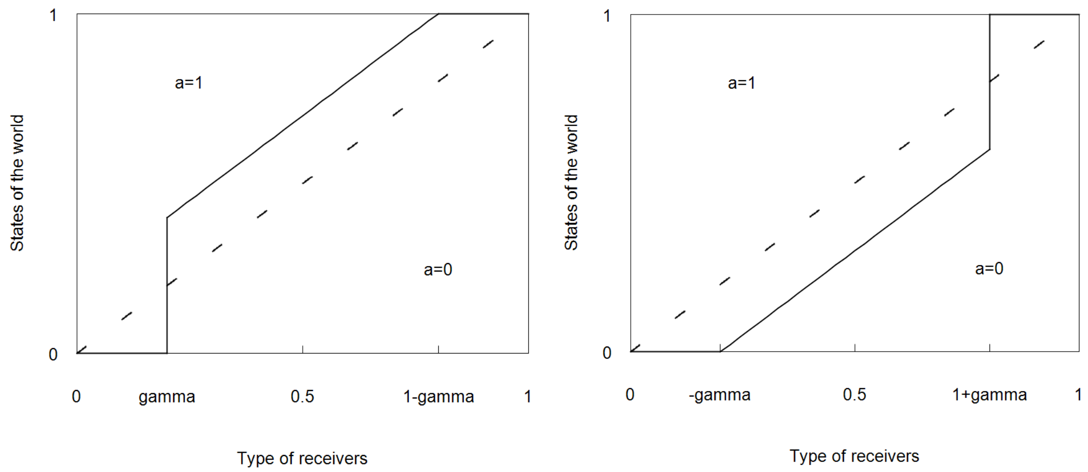
$$\begin{cases} \mathbb{E}(r|r > \theta + \gamma) & > \theta \\ \mathbb{E}(r|r < \theta + \gamma) & < \theta \end{cases} \tag{5.15}$$

Because $\gamma > 0$, the first condition of equation (5.15) is always satisfied, and the second one is equivalent to $\theta > \gamma$. Communication is possible with all $\theta \in [\gamma; 1 - \gamma]$, which is non-empty if and only if $\theta < \frac{1}{2}$.

Likewise, when $\gamma < 0$, the second condition of equation (5.15) is always satisfied, and the first equation is equivalent to $\theta < 1 + \gamma$. For more than one message to be sent in equilibrium, we need $\theta + \gamma > 0$, which is satisfied when $\theta > -\gamma$. Hence, communication is possible with all $\theta \in [-\gamma; 1 + \gamma]$, which is non-empty when $\gamma > -\frac{1}{2}$. Hence, the scope of private communication is $[-\frac{1}{2}; \frac{1}{2}]$, which strictly contains the scope of public communication. □

Under private communication, and under mild negative or positive externalities, the sender

Figure 5.2: Distribution of actions in a private communication setting ($\mid \gamma \mid < \frac{1}{2}$)



uses communication to convince all receivers located in a centred interval. This interval structure is represented on figure 5.2. The figure on the left-hand side represents the actions taken by each receiver in each state of the world when γ is negative. The right-hand side represents the case of positive externalities.

Under negative externalities, the receivers located in the extremities of the unit interval are not convinced through communication, as they either have a very small cost of protection, or a very high one. Receivers characterized by a low θ are particularly hard to convince due to their high ability to avoid harm. In equilibrium, they all take the protective action. On the other hand, no credible communication need to be sustained with receivers characterized by high θ, who would anticipate that the preferred action of the receiver is outside the set of possible catastrophes. In equilibrium, they receive no information and choose not to avoid harm.

Under positive externalities, this interval structure also exists, but the interpretation of the motivations of the receivers left outside of this interval are interchanged. Receivers with a low cost of avoiding harm anticipate that the preferred action of the sender is for them to always take the pro-social action. No credible communication strategies need to exist between them, as they take this preferred action when left without information. Conversely, receivers characterized by a high cost of protection are too hard to convince. This case is plotted on the right-hand side of figure 5.2.

The next proposition shows that in addition to extending the scope of communication, private communication also improves welfare.

Proposition 5.4.2. *For all values of γ and all public communication equilibria, the expected* *welfare obtained under private communication is always weakly greater than the expected welfare obtained under public communication.*

Proof. Assume $\gamma \in [0; \frac{1}{2}]$. Note $\mathbb{E}w(\theta,\gamma)$ the expected utility derived by the sender from its interaction with receiver θ, this quantity takes the following form:

$$\mathbb{E}w(\theta,\gamma) = \begin{cases} \int_0^{\theta+\gamma} -rdr - \int_{\theta+\gamma}^1 (\theta+\gamma)dr = \frac{(\theta+\gamma)^2}{2} - (\theta+\gamma) & \text{if } \theta \in [\gamma; 1-\gamma] \\ \int_0^1 -rdr = -\frac{1}{2} & \text{if } \theta > 1-\gamma \\ \int_0^1 -(\theta+\gamma)dr = -\theta-\gamma & \text{if } \theta < \gamma \end{cases} \tag{5.16}$$

Hence, the expected welfare for the sender $\mathbb{E}W_{private}(\gamma)$, derived from its private interaction with all receivers can be written as:

$$\mathbb{E}W_{private}(\gamma) = \int_0^{\gamma} \mathbb{E}w(\theta,\gamma)d\theta + \int_{\gamma}^{1-\gamma} \mathbb{E}w(\theta,\gamma)d\theta + \int_{1-\gamma}^{1} \mathbb{E}w(\theta,\gamma)d\theta \tag{5.17}$$

Straightforward integration yields the following result:

$$\mathbb{E}W_{private}(\gamma) = \begin{cases} -\frac{4\gamma^3}{3} + \frac{\gamma^2}{2} - \frac{\gamma}{2} - \frac{1}{3} & \text{if } \gamma \in [0; \frac{1}{2}] \\ -\frac{\gamma}{2} - \frac{3}{8} & \text{if } \gamma \in [\frac{1}{2}; 1] \end{cases} \tag{5.18}$$

Similarly, when $\gamma < 0$, the following result holds:

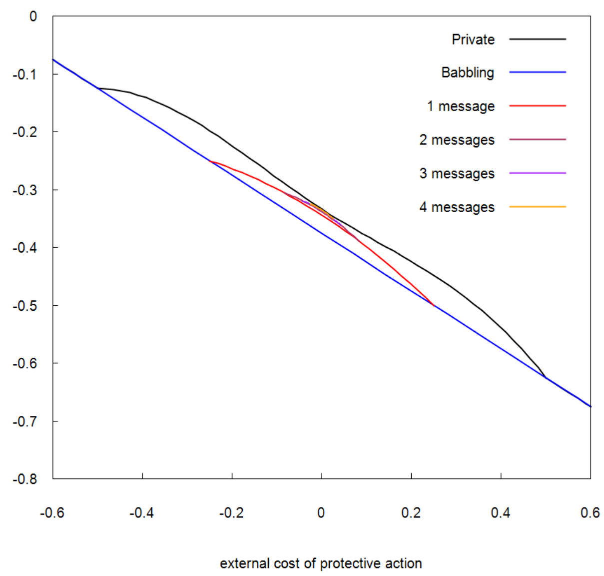
$$\mathbb{E}W_{private}(\gamma) = \begin{cases} \frac{(1+2\gamma)^3}{6} - \frac{(1+2\gamma)^2}{2} + \frac{\gamma^2+\gamma}{2} & \text{if } \gamma \in [-\frac{1}{2}; 0] \\ -\frac{\gamma}{2} - \frac{3}{8} & \text{if } \gamma \in [-1; -\frac{1}{2}] \end{cases} \tag{5.19}$$

As illustrated in figure 5.3, the expected welfare under private communication is equal to the expected welfare from public communication when $|\gamma| > \frac{1}{2}$, and always weakly larger than the welfare obtained under public communication equilibria. □

5.5 Disaster response: communication and transfers

In this section we suppose that the sender can perform positive monetary transfers indexed either on the actions taken in equilibrium or on the types of the receivers, and subject to a budget constraint. We assume that the government can spend up to B to compensate people. This constraint embodies the existence of public relief funds which are funded before a catastrophe

Figure 5.3: Expected welfare associated with private communication



strikes, and depleted afterwards. We also assume that $B < \gamma$ so that solutions are necessarily second-best.[12]

5.5.1 Partial public insurance and subsidies

In a first simple benchmark case, we assume that the sender commits to transferring t_0 and t_1 to all receivers engaged respectively in action $a = 0$ and $a = 1$, regardless of the state r. We note $\Delta t = t_1 - t_0$, and assume that the sender's budget is constrained by a positive quantity B such that:

$$\forall r, \int_0^{\theta^\star(r)} t_1 d\theta + \int_{\theta^\star(r)}^1 t_0 d\theta \leq B < \gamma \tag{5.20}$$

We assume that the utility of the sender is unaffected by these transfers, and that the utility of the receivers is now described by:

$$u_R(r, \theta, a) = \begin{cases} -r + t_0 & \text{if } a = 0 \\ -\theta + t_1 & \text{if } a = 1 \end{cases} \tag{5.21}$$

A receiver now chooses action $a = 0$ if and only if $\theta > \mathbb{E}(r) + \Delta t$. This changes the Bayesian rationality constraint that defines the equilibria of the cheap talk game. Indeed, conditionally on a credible message sent in equilibrium on any interval $[r_a; r_b]$, the equilibrium outcome function

[12] For instance, in the case of negative externalities, the sender could transfer γ to all receivers choosing $a = 0$, which would perfectly align the incentives of the sender with those of the receivers.

$\theta^\star(r)$ now has to satisfy:

$$\theta^\star = \frac{r_a + r_b}{2} + \Delta t \tag{5.22}$$

The incentive compatibility constraint of the sender is also affected, and equation (5.5) now yields:

$$r_{i+1} - r_i = r_i - r_{i-1} - 4(\gamma + \Delta t) \tag{5.23}$$

Equation (5.23) shows that after choosing a transfer program $(t_0; t_1)$, the subsequent cheap-talk game is characterized by the same communication equilibria as the original cheap-talk game under external effects $\gamma' = \gamma + \Delta t$. In other words, transfers allow the sender to access different communication equilibria. Intuitively, when Δt and γ have opposite signs, transfers offset the bias of the sender, who will access more informative credible communication strategies. When Δt and γ have the same signs, transfers increase the existing bias, and credible communication strategies become less informative. The following proposition characterizes optimal transfers.

Proposition 5.5.1. *Under negative externalities, optimal transfers entail a partial public insurance for harmed receivers, i.e.* $t_0 > 0$ *and* $t_1 = 0$*. Under positive externalities, optimal transfers entail a subsidy for the costly protective action, i.e.* $t_0 = 0$ *and* $t_1 > 0$*.*

Proof. We first determine the expected welfare of the new cheap talk game, in which the externalities associated with the receivers' actions are measured by γ, while the communication strategies available to the sender correspond to those that would be played under externalities $\gamma' = \gamma + \Delta t$. Hence, noting $\theta_i'^\star$ the equilibrium outcomes associated with the new communication strategies available to the sender, and r_i' their associated thresholds, the expression of the equilibrium expected welfare is:

$$\mathbb{E}W(\gamma, \Delta t, n) = \sum_{i=1}^{n} \int_{r'_{i-1}}^{r'_i} \left(- \int_0^{\theta_i'^\star} (\theta + \gamma) d\theta - \int_{\theta_i'^\star}^1 r d\theta \right) dr \tag{5.24}$$

which leads to an expression similar to equation (5.11):

$$\mathbb{E}W(\gamma, \Delta t, n) = -\frac{(n^2-1)((\Delta t)^2 + 2\gamma \Delta t + \gamma^2) + 3\gamma + 2}{6} - \frac{1}{24n^2} \tag{5.25}$$

Optimal transfers are still considered with respect to the sender-preferred equilibria. Hence the sender chooses Δt so as to maximize $\sup_n\{\mathbb{E}W(\gamma, \Delta t, n)\}$. For all values of n, this is equivalent to finding the minimum value of the polynomial $(\Delta t + \gamma)^2$. Intuitively, this polynomial is minimum for $\Delta t = -\gamma$. Yet, choosing $\Delta t = -\gamma$ would violate the budget constraint of the sender. Indeed, when $\gamma' = 0$, the sender-preferred equilibrium relies on a truthful communication strategy, and all receivers would ask for t_0 in state $r = 0$ or for t_1 in state $r = 1$, depending on the sign of γ. In both cases, the total transfer realized would be higher than B.

Hence, the best the sender can do is to maximize t_0 (keeping $t_1 = 0$) when γ is positive, and maximize t_1 (keeping $t_0 = 0$) when γ is negative. In each case, the budget constraint has to be binding in the states of nature in which the largest number of receivers receive the transfers. When $\gamma > 0$, the constraint is binding when $r \in [0; r'_1]$, as $1 - \theta_1'^{\star}$ receivers have to be paid. When $\gamma < 0$ the constraint is binding when $r \in [r'_{n-1}; 1]$, as $\theta_n'^{\star}$ receivers have to be paid. This yields:

$$\Delta t = \begin{cases} -t_0 = -\frac{B}{1-\theta_1'^{\star}} & \text{if } \gamma > 0 \\ t_1 = \frac{B}{\theta_n'^{\star}} & \text{if } \gamma < 0 \end{cases} \tag{5.26}$$

Finally, using the definition of $\theta_1'^{\star}$ and $\theta_n'^{\star}$ in equation (5.26) leads to the following expressions of t_0 and t_1 as a function of γ, n and B.

$$t_1 = \frac{1}{4n^2} - \frac{\gamma}{2} + \frac{\gamma}{2n} + \frac{1}{2n}\sqrt{\left(\gamma(n-1) - \frac{1}{2n}\right)^2 + 4nB} \tag{5.27}$$

$$t_0 = \frac{1}{4n^2} + \frac{\gamma}{2} - \frac{1+\gamma}{2n} + \frac{1}{2n}\sqrt{\left(\frac{2n-1}{2n} - \gamma(n-1)\right)^2 + 4nB} \tag{5.28}$$

□

Proposition 5.5.1 has two direct policy implications. First, the sender should provide a subsidy for actions characterized by positive externalities. Second, it shows that the sender should provide partial public insurance to all receivers who cannot protect against disasters when actions are characterized by negative externalities. These transfers offset the credibility issue faced by the sender, allowing him to achieve larger levels of welfare while increasing the level of transparency of his equilibrium communication strategy.

The second implication contributes to the literature on the Samaritan dilemma, which argues that public insurance creates an ex ante moral hazard problem by providing incentives for individuals to lower their level of protective efforts before major disasters (see e.g. Epstein (1996) or Kunreuther (1996)). When the actions taken by the population entail negative externalities, we show that a certain level of public insurance may serve the purpose of the government during the communication phase that immediately follows the catastrophe, by allowing the government to communicate more transparently while curtailing the negative external effects of the population's response to the disaster.

Equations (5.27) and (5.28) show that optimal transfers are non-decreasing in B, but less than proportional to it. This is so because increasing transfers also increases the quantity of receivers demanding these transfers in equilibrium. In the case of positive externalities, an increase in the budget leads the subsidy t_1 to be provided to more numerous receivers when r is large. In the case of negative externalities, an increase in the budget would lead the partial public insurance t_0 to be provided to more numerous receivers when r is low.

5.5.2 On the use of disaster relief funds

This benchmark suffers from an important shortcoming. Indeed, it is possible to derive from propositon 5.5.1 a couter-intuitive result regarding the timing of the depletion of the government's budget. As we have assumed that transfers are committed, the mechanics of the negative and positive external effects cases are again different. In both cases, it is necessary that total expenditure does not exceed total budget when the largest amount of receivers claim the transfer. Hence, under negative externalities, the budget is completely depleted only if r is below r_1. Under positive externalities, the budget is depleted when r is superior to r_{n-1}. In other words, under positive externalities, the relief fund B is depleted when the most severe catastrophe strikes. On the other hand, under negative externalities, the fund is depleted when the least severe catastrophes strike, and the budget is not fully depleted under more severe events. This second part of the result may seem counter-intuitive.

This result is due to the fact that a single transfer program $(t_0; t_1)$ had to satisfy the sender's budget constraint in all states of nature. Yet, a policy-maker may want to withhold the use of a relief fund in order to save it for very severe events. For instance, the use of relief funds are often characterized by a specific threshold which is set to decide when to start depleting the fund. In addition, one could argue that if the size of the relief fund B has to be determined ex ante, the transfers t_0 and t_1 are usually determined ex post, based on the needs of victims or on the basis of the damage undergone.

Hence, it is interesting to study the case in which the pair $(t_0; t_1)$ may depend on the state of nature, thus relaxing the sender's budget constraint. In this case, it is no longer possible to derive closed form solutions for the equilibria of the new cheap-talk game, as the Bayesian rationality constraint of the receivers takes a more complex form:

$$a = 0 \Leftrightarrow \theta > \mathbb{E}(r + \Delta t(r) | m(r)) \tag{5.29}$$

Then, noting $\mathbb{E}_i(\Delta t) = \mathbb{E}(\Delta t(r)|r \in [r_{i-1}; r_i])$, equation (5.29) yields the following induction condition on the structure of equilibria:

$$\forall i < n,\ r_{i+1} - r_i = r_i - r_{i-1} - 4\gamma - 2\left[\mathbb{E}_i(\Delta t) + \mathbb{E}_{i+1}(\Delta t)\right]. \tag{5.30}$$

Characterizing the equilibria satisfying equation (5.30) is outside of the scope of this chapter. To overcome this tractability issue, we propose to study a simple policy instrument that illustrates the *state of natural catastrophe*, which is used in most countries to determine when public funds are used to provide compensations to victims of a disaster. To do so, instead of conditioning transfers on states of the world, we restrict the analysis to the case of constant transfers that can only be made if a catastrophe exceeds a certain threshold.

Suppose that a government has access to a relief fund B, and can determine a threshold r_0 above which this budget will be used in order to compensate victims. When $r < r_0$ the utility of a receiver is given by equation (5.1). When $r > r_0$, their utility is described by equation (5.21). The threshold r_0 serves both as a trigger for the relief fund, and as a communication device to warn the population. In equilibrium, the definition of this threshold needs to be credible for the population. Hence, we necessarily have:

$$W(r_0, \theta_1^\star) = W(r_0, \theta_2^\star) \tag{5.31}$$

where $\theta_1^\star$ and $\theta_2^\star$ describe the actions played in equilibrium when r is respectively below and above r_0. The Bayesian rationality constraints imply that $\theta_1^\star = \frac{r_0}{2}$ and that $\theta_2^\star = \frac{1+r_0}{2} + \Delta t$. Finally, the budget constraint of the sender reads:

$$\int_0^{\theta_2^\star} t_1 d\theta + \int_{\theta_2^\star}^1 t_0 \leq B \tag{5.32}$$

This can easily be shown to be equivalent to $(\Delta t)^2 + \frac{1+r_0}{2}\Delta t + t_0 \leq B$.

Corollary 3. *Conditionally on transfers t_0 and t_1 chosen in equilibrium, there is a unique credible equilibrium characterized by a single threshold r_0:*

$$r_0 = \Delta t + 2\gamma + \frac{1}{2} \tag{5.33}$$

In addition, when $\gamma > 0$, equilibrium transfers consist in partial public insurance granted when $r > r_0$ to receivers who choose action $a = 0$.

Proof. Equation (5.33) is directly obtained by combining equation (5.31) and the Bayesian rationality constraints. This implies that choosing a credible threshold ex ante fully determines the amount of the transfers made ex post (or at least their difference Δt). Thus, we solve the optimization problem of the sender by considering the transfers as her sole decision variable. The optimization problem is to choose the amount of the transfers in order maximize expected welfare, while satisfying the budget constraint. Expected welfare is here defined as:

$$\mathbb{E}W(\Delta t, r_0, \gamma) = \int_0^{r_0} W(r, \theta_1^\star)dr + \int_{r_0}^1 W(r, \theta_2^\star)dr \tag{5.34}$$

To solve this problem, we focus on the case of negative externalities. We can restrict to the cases in which $t_1 = 0$ and $t_0 > 0$. If the solution to our problem entails both types of transfers, then the sender can reduce both types of transfers by the same amount without modifying expected welfare, which only depends on Δt. As we assume that $\gamma > 0$, we necessarily have $t_0 > t_1$, so that we can decrease t_1 until it is null and the sender only subsidizes people who cannot take the protective action. □

We next turn to the determination of the transfers made in equilibrium. Using the definition of the expected welfare and the budget constraint, the optimization problem can be shown to boil down to:

$$\max_{t_0} -\frac{t_0^3}{2} + (2\gamma - \frac{3}{8})t_0^2 - \gamma(2\gamma - 1)t_0 \tag{5.35}$$

$$\text{s.t. } \frac{3}{2}t_0^2 + (\frac{1}{4} - \gamma)t_0 - B \leq 0 \tag{5.36}$$

Then, the following result holds:

Corollary 4. *If the budget of the sender is large enough, it is not necessarily depleted in equilibrium.*

Proof. As equation (5.36) is a quadratic constraint with positive coefficient, this problem consist in maximizing the expression shown in equation (5.35) over a closed interval that we note $[0; t_+]$.[13] This problem always has a strictly positive solution, which can be either the upper corner solution t_+ or the unconstrained maximizer of equation (5.35), noted $t_0^\star$. These two

[13]The interval is closed because the constraint can be binding. t_+ is the unique positive root of the polynomial defined by equation (5.36), which explains why the other end of the constrained interval is 0.

quantities can be explicitly calculated:

$$t_0^\star = \frac{4\gamma}{3} - \frac{1}{4} + \frac{1}{3}\sqrt{4\gamma^2 + \frac{16}{9}} \tag{5.37}$$

$$t_+ = \frac{\gamma}{3} - \frac{1}{12} + \frac{1}{3}\sqrt{(\frac{1}{4} - \gamma)^2 + 6B} \tag{5.38}$$

Simple manipulations show that $t_+ > t_0^\star$ if B is superior to a minimum budget B_{min}. □

This is intuitive, if the budget is too small, the sender increases the size of transfers in order to improve welfare until the budget constraint is binding. If the budget B is large, the sender can maximize expected welfare without depleting the whole fund.

The study of this simple policy instrument shows that it is possible to combine communication with the use of relief funds in order to convey credible information to the population, and to foster socially optimal responses to catastrophic events. In addition, the use of thresholds to trigger the use of relief funds can be a useful instrument to avoid squandering resources during minor events. In this sense, it appears that using simple instruments such as a pre-determined *state of natural catastrophe* is a sensible policy.

This result still holds if we assume that the relief fund has to be depleted, i.e. if equation (5.36) has to be binding. This assumption is consistent, for instance, with the case of a populist sender searching to buy votes by distributing the totality of the relief fund in the wake of a major disaster.

Under this hypothesis, the solution provided above is still optimal up to a symmetric increase of both types of transfers. Indeed, the solution above can first be characterized by a binding budget constraint, in which case this is an optimal solution transfer for the politician. In the other case, the politician can shift the optimal t_0 found above as well as the transfer t_1 made to receivers engaging in the protective action. Because expected welfare only depends on the difference between the two transfers, a parallel shift of t_0 and t_1 allows the politician to deplete the whole fund, without hindering the effect of these transfers on the incentives of the population to take protective actions.

5.6 Conclusion

In this chapter, we present a model of strategic communication between a sender informed of an incoming catastrophe and a continuum of receivers characterized by heterogeneous abilities to avoid harm. In addition, each receiver can take a binary protective action, which entails a private cost and an external effect on society. Hence, the sender tries to convince receivers to take her preferred actions.

We use this model to shed light on disaster communication strategies used by governments and safety regulators in the wake of major catastrophes. Similarly with the disaster communication literature, we find that considering the response of populations to the information disclosed is paramount in the design of credible communication strategies. We also find that transparency increases welfare, and that private communication always improves the outcome of the communication game.

In addition, we study how the policies implemented by a government in order to prevent, prepare for or respond to major disasters interact with the incentives for the government to disclose information in the wake of the catastrophe. When trying to avoid panic, preparedness always increases welfare and fosters transparent communication. On the contrary, under positive externalities, fostering pro-social behaviours leads to reduced levels of transparency, as it increases the conflict of interest of the sender.

When the government can use transfers to offset the bias of the population, we show that an optimal policy is to subsidize costly actions when these actions entail positive externalities, and to provide partial public insurance when protective actions entail negative externalities. This finding provides a rationale for the use of partial public insurance in the case of rare disasters, in contrast with the usual conclusions from the Samaritan dilemma literature. When transfers are non-committed, we describe a simple policy instrument that matches the usual definition of the state of natural catastrophe, which is often used by states to use public funds for the curtailment of the consequences of natural disasters.

Résumé du chapitre 6

Ce chapitre de conclusion reprend les contributions des différents chapitres de la thèse. La première partie de la thèse, composée des chapitres 2 et 3, montre que malgré les incertitudes qui caractérisent le risque nucléaire, sa réglementation peut s'appuyer sur d'autres sources d'informations. D'abord, il est possible de décrire et d'évaluer quantitativement les préférences individuelles face au risque et à l'incertitude. Ensuite, des sources d'informations permettent de caractériser la sûreté nucléaire malgré la rareté des accidents. Dans la seconde partie de la thèse, la mise en œuvre des réglementations de la sûreté et des politiques de gestion des désastres sont analysées sous l'angle de la communication et de la transmission d'information. Le chapitre 4 montre que la publication locale d'information peut influer sur la transparence du comportement déclaratif des exploitants et augmenter la quantité d'information à disposition du régulateur. Le chapitre 5 montre que la communication de crise est essentielle à la mitigation des conséquences des événements désastreux, et décrit les problèmes qui peuvent apparaître dans la coordination de la communication avec les actions de gestion post-accidentelle.

CHAPTER 6

Conclusion

In the introduction of this thesis, I claimed that the characteristics of nuclear accidents do not correspond to the classical framework developed in the traditional economics literature dedicated to safety regulation. In particular, I argued that their rare and catastrophic nature severely hampered the measurement of the effects of safety care on safety levels, and challenged the normative results of the economics of safety regulation.

I believe this observation could have led this thesis in two different directions. A first direction could have consisted in the development of a new conceptual framework to study safety regulation when risks are rare and catastrophic. This direction would have prevailed if the conflict between uncertainties and the classical public economics of risks had proven to be insoluble. On the other hand, a second research direction consists in finding ways to reconcile these characteristics with the usual conclusions from the economics literature.

This thesis resolutely embraced the latter direction. This thesis shows that in some circumstances, uncertainties can be coped with by using theoretical or empirical methods taken from the economic toolbox. In particular, decision and game-theoretic frameworks can be developed to account for uncertainties in the making of public policies characterized by rare and catastrophic risks. Likewise, the empirical analysis of minor incidents can provide new and informative insights regarding the evolution of safety or the effectiveness of safety-related policies.

From a methodological point of view, this thesis developed positive tools to solve the measurement issue raised by rare and catastrophic disasters. Applying econometrics methods to well identified sets of safety incidents can be a solution to measure the causal effects of safety regulations on safety care and safety levels. Likewise, conceptual frameworks can be developed to better describe the uncertainties that characterize the low probabilities associated with major disasters, to assess the effect these uncertainties have on individual behaviours and disaster mitigation policies, and to provide public decision-makers with tools that allow the comparison of alternatives characterized by uncertainties with more certain ones.

These methods led to several normative conclusions regarding nuclear safety and its regulation. First, I proposed a way to adapt cost-benefit analysis to the case of rare and catastrophic risks, in order to provide decision-makers with assessments of the social costs of nuclear accidents that account for the multiplicity of models describing their probabilities of occurrence. This thesis also showed that local monitoring policies were an interesting complement to classical command-and-control safety regulation, as these policies provide positive incentives for plant managers to comply with safety regulation. Finally, I showed that providing public insurance to the victims of a major nuclear disaster does not only have a negative moral hazard effect on safety investments, as it also allows more transparent and more effective disaster communication strategies.

Overall, this thesis stresses the need to acknowledge and address the limits of classical command-and-control safety regulations in industries characterized by rare and catastrophic disasters. In particular, chapter 2 proposes a method to account for individual attitudes towards uncertainty when assessing the costs associated with various policy alternatives that may involve nuclear power. By doing so, I account for the fact that the individual willingness to avoid nuclear accidents may exceed the expected sum of their physical consequences. Therefore, motivating energy policies by relying solely on the monetary values of the consequences of these accidents and on the technical assessments of their likelihood of occurrence would fail to acknowledge that individuals are not indifferent to the uncertainties that characterize these events.

Chapter 3 suggests that the information derived from minor safety incidents can be used as a proxy to characterize the variations of nuclear safety. This proxy conveys information regarding safety, which can be valuable for decision-makers. For instance, using this safety metric, it appears that safety is negatively correlated with the age of reactors, but positively correlated with time. In a context in which multiple reactors are going to seek life extensions beyond their fortieth anniversary in the near future, this information should feed the debate regarding the early closure of potentially safe nuclear stations.

Chapter 4 shows that even without explicit sanctions, local monitoring can exert some form of pressure on plant managers, which improves compliance. In France, fostering more intense monitoring activities from local monitoring agencies could be used to marginally improve compliance with reporting guidelines, which would in turn increase the quantity of information obtained by the regulator regarding the safety of the fleet. In other countries, implementing local monitoring policies could be a cost-effective improvement of nuclear safety regulation, given the large cost associated with technical safety improvements, and the relatively low budgets required by the French local monitoring comities.

Chapter 5 shows that preparedness actions, partial public insurance for the victims of accidents, public subsidies for pro-social behaviours, and disaster relief funds are effective instruments to improve the quality of the response of a population to catastrophic disasters such as nuclear power accidents. Not only do these instruments convey direct incentives to the population to take better actions, they also reduce the conflict of interest of the informed parties once a disaster occurs. By doing so, they allow for more transparent and credible communication strategies, which in turn improve the outcome of the response of the population.

This thesis leaves many questions unanswered and opens some perspectives for future research. From an empirical perspective, the analysis of significant safety events could be pursued in the veins of the reliability literature, by analysing the various hazard rates associated with different types of events and their evolution over time. Questions related to monitoring intensity and its effect on compliance and safety care could be investigated further by analysing data on the inspections performed by the nuclear safety authority, and by measuring its effect on the behaviour of plant managers in the period of time that follow or precede inspections. Comparing data from the planned and random audits performed by the safety authority could also provide meaningful results regarding the deterrence effect of monitoring.

From a theoretical perspective, the mechanisms a regulator should implement in order to foster optimal levels of safety care when safety is only ambiguously measurable is still an open question. Likewise, the combination of the moral hazard effect created by disaster relief funds with its positive effect on communication credibility could be investigated further. Another open question pertains to social choice theory, as it remains unclear whether social choice functions should exhibit catastrophe-aversion or ambiguity aversion, as rare disasters can be characterized by both ambiguous probabilities and catastrophic outcomes.

Part III:

Appendices

If people do not believe that mathematics is simple, it is only because they do not realize how complicated life is.

J. Von Neumann in Alt (1972)

APPENDIX **A**

Proofs

A.1 Proof of lemma 3.3.1

Proof of lemma 3.3.1. Note that

$$\mathbb{E}\{Y|W=w\} = \mathbb{E}\{\mathbb{E}\{Y|W=w,D,O\}\} = \mathbb{E}\{f(D,O)\}, \tag{A.1}$$

with $f(D,O) = \mathbb{E}\{Y|W=w,D,O\}$. In the following, we suppress the dependency on W for the sake of representational simplicity. It holds

$$\mathbb{E}\{f(D,O)\} = \sum_{d,o} f(d,o)\mathbb{P}\{D=d,O=o\}. \tag{A.2}$$

We observe that due to assumption A2, $\mathbb{E}\{Y|O=0\} = \mathbb{P}\{Y=1|O=0\} = 0$, and thus, under A1, A2 and A3, equation (A.2) gives

$$\mathbb{E}\{Y|W=w\} = f(1,1)\mathbb{P}\{D=1,O=1\} = \mathbb{E}\{Y|W=w,D=1,O=1\}. \tag{A.3}$$

□

A.2 Proof of proposition 4.3.1

Let $\mu(z) = \alpha\rho z + q\beta\int_{z\rho}^{1}(u - z\rho)f(u)du$. From equation (4.2), one can derive the following first-order condition characterizing the existence of an interior solution for the agent's choice:

$$
\begin{aligned}
z^\star &= \frac{1}{\rho}F^{-1}(1-\frac{\alpha}{q\beta}) \quad (A.4) \\
B'(E^\star_{tot}) &= \mu(z^\star) \quad (A.5)
\end{aligned}
$$

The effect of a change in α or $q\beta$ on $z^\star$ derives directly from equation (A.4). Then, using the envelope theorem in equation (A.5), we can differentiate $\mu(z^\star)$ with respect to either α or $q\beta$, which yields the second part of the result. This result is identical to the results provided by Evans et al. (2009) and Gilpatric et al. (2011).

APPENDIX B

Model selection and bias in the relation between age and safety

B.1 Significant safety event data

Not all reported significant safety events were used for our analysis. First, generic events characterizing the whole fleet or particular cohorts of reactors are excluded. These generic events consist in detections of conception failures which are specific to either a group of reactors, or to the whole fleet. According to the regulator, these events capture specific efforts made by EDF to increase its knowledge of the conception of his reactors, as well as their reliability. Second, we discarded the events declared during construction, as the safety authority considers the reporting criteria to be tailored for the operation and maintenance of reactors rather than for their construction periods. Third, although our dataset initially included events reported between 1973 and 1996, we excluded this period of observation due to incompleteness of the description of the events. In 1996, a reform of the reporting criteria led to a more stringent and complete reporting (and numerical storage) process. Conversely, when aggregating safety events into counts of reports per reactor and per year, the events that affected some systems common to multiple reactors within a given power station were counted in each affected reactor.

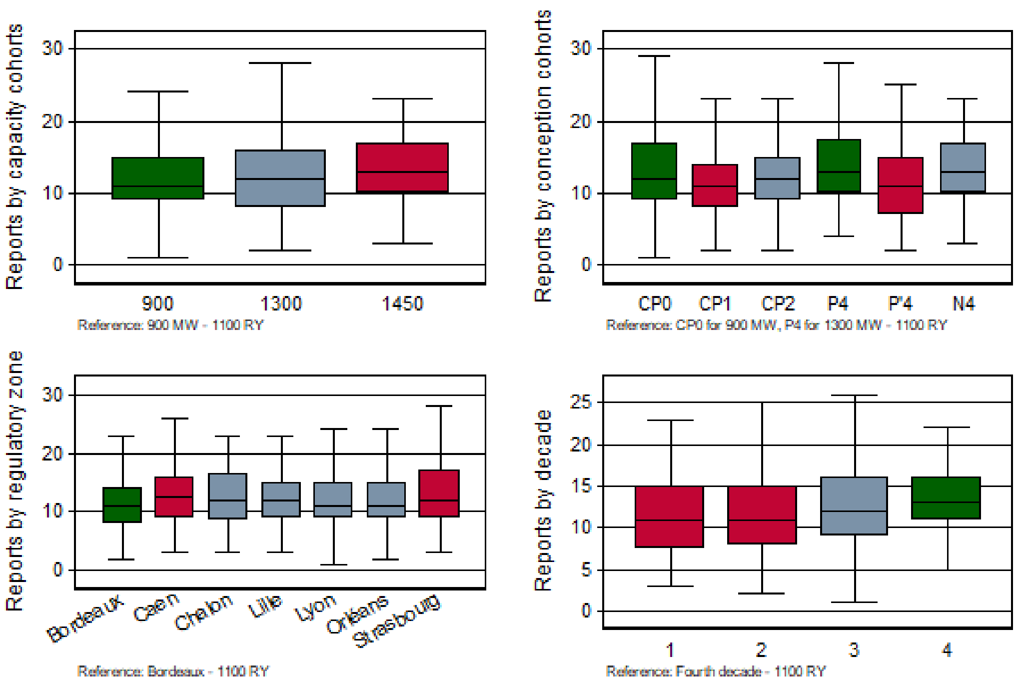
B.2 Additional descriptive evidence

In this appendix, we compare the frequencies of reports of safety significant events (without any restriction to the PDD subsample, and using all available data regarding events reported between 1973 and 1997) across groups of reactors. [1] Figure B.1 presents four box plots showing how the

[1] We refer to statistics describing counts of events occurring in a given reactor during a given year as reactor.year statistics. For instance, the Gravelines-5-2004 reactor.year is constituted of all the events declared in the fifth reactor of the Gravelines power station in 2004.

annual declarations of events are distributed across four definitions of groups of reactors. In each box plot, multiple comparisons of group means are performed. We present the results of these comparisons by indicating a reference group in green, and the groups that differ significantly (at the 5%-level of significance) from this reference in red. The statistics on which these comparisons are based account for multiple pairwise comparisons[2].

Figure B.1: Annual declarations by reactor for different groups of reactors



For the upper two sub-figures in figure B.1, the x-axes represent the capacity and conception cohorts, ordered by date of construction. It appears that the reactors that belong to the most recent capacity cohort (1450 MW) declare a significantly larger average number of events when compared to the 900 MW cohort.[3] Some heterogeneity appears within capacity cohorts when split into the six underlying conception cohorts. Within the 900 MW group, reactors of the oldest design ($CP0$) declare a significantly higher number of events than the reactors of the other two designs. The same conclusion can be drawn for the reactors of the $P4$ conception cohort with respect to their younger siblings of the $P'4$ conception cohort. Note that in this second box plot, only the results of pairwise comparisons within each capacity cohorts are represented.

The bottom-left graph on figure B.1 is dedicated to France's regional regulatory subdivi-

[2]The significance of the difference in means among the pairs identified in figure B.1 is robust to several methods of adjustment for multiple comparisons (i.e. Tukey, Sidak and Bonferonni's methods).

[3]According to the safety authority, this is be due to the comparatively more complex design of these recent plants, which led to large quantities of events in their early years of operation.

sions.[4] Regional distributions of reports of significant safety events are rather equally distributed, as only two pairwise comparisons of the mean annual number of reports are significant. It appears that the reactors overseen by the Bordeaux division (e.g. the 8 reactors located in the Blayais, Golfech and Civaux power stations) declare significantly lower numbers of events than the reactors overseen by the Strasbourg division and the Caen division.

On the bottom-right graph, four age groups are defined by the decade of operation of a reactor at the time of observation.[5] Reactors in their third and fourth decade of operation declare significantly larger numbers of events than reactors in their first decade of operation. It is interesting to notice the relative under-dispersion of the fourth-decade group with respect to the three younger groups. This could be due to a relative lack of observation of reactors in their fourth decade. Indeed, as of 2014, only 45 reactor.years in the fourth decade have been observed, whereas respectively 743, 563 and 414 reactor.years have been observed in the first, second and third decade of operation.

B.3 Model selection

Our model selection is based on standard arguments, well established in the empirical and theoretical literature. First, the count specification presented in equations (3.4) and (3.5) has several advantages over the standard linear model, see e.g. Wooldridge (2002a) or Cameron and Trivedi (2013). Most importantly, the linear model might produce negative predictions for feasible values of the observed covariates, something we would like to preclude.

Second, concerning the estimation procedure, we estimate this model using a Negative Binomial estimator with quadratic over-dispersion and site-clustered standard errors. We perform model selection using the Akaike and Bayesian information criterion, whose values are reported in table B.1. Calculations of the Pearson statistics for the Poisson regressions (on the ASD and SFG subsamples) allow to strongly reject the Poisson distribution for both automatic shutdowns and safeguard events. Moreover, the standard Akaike and Bayesian information criteria (AIC and BIC), computed using specification (1) under linear and quadratic over-dispersion support the use of the negative binomial models with site-clustered standard errors and quadratic over-dispersion specification.

As a result, the Negative Binomial estimator with quadratic over-dispersion and site-clustered

[4] The ASN delegates the duty of inspection to its 7 territorial subdivisions, who have some level of discretion in their interaction with the plant managers.

[5] For instance, an event observed in a 26 year-old reactor is associated with the third decade of operation of this reactor.

Table B.1: Test-statistics for regression model selection

Regression model	Over-dispersion	S-E clusters	Ln L	Pearson	AIC	BIC
Poisson	-	Reactor	-1473	1198	2956	2981
Poisson	-	Site	-1473	1198	2982	3072
Neg. Bin.	NB1	Reactor	-1471	-	2952.2	2977.2
Neg. Bin.	NB1	Site	-1471	-	2978.1	3068.1
Neg. Bin.	NB2	Reactor	-1471	-	2952.1	2977.1
Neg. Bin.	NB2	Site	-1471	-	2978.2	3068.3

ASN data. This table presents test statistics associated with different count estimators using the model specification presented in equation (3.5) and automatic shut-downs as a dependent variable (i.e. regression (1) in table 3.4). Two regression models, e.g. the Poisson and Negative Binomial model are compared using the Pearson over-dispersion test statistics. For the Negative Binomial model, we also compare linear and quadratic over-dispersion specifications. For all models, we compare standard-error clustering at the reactor or site level. The log-likelihood, Akaike and Bayesian Information Criteria are computed for each regression models.

standard errors dominates the Poisson-QMLE estimator and the Negative-Binomial estimator with linear over-dispersion. One advantage of the Negative Binomial estimator is that it better fits over-dispersed data, consistently with the descriptive evidence provided in table 3.2 on page 77. A second advantage of the Negative Binomial estimator is that it is more efficient than the Poisson-QMLE estimator in some cases, see Cameron and Trivedi (2013) for a discussion.[6]

The interpretation of the coefficients obtained through a Negative-Binomial regression model can be done using incidence rate ratios: given any explanatory variable X and its coefficient β_X, e^{β_X} represents the ratio of the expected counts of events obtained after and before a unit increase of X. When β_i is small for all i, β represents the vector of semi-elasticities of the dependent variable Y in the explanatory variables X. In addition, when coefficients are small and explanatory variables are included in logarithmic form, then β can be interpreted as a regular elasticity. As an example, using the notations defined in equation (3.5), $e^{\beta_{Age}}$ represents the (multiplicative) average effect of a unit increase in the age of a reactor on the expected number of occurrences of events.

B.4 Is unobserved heterogeneity site-specific and time-constant? Some robustness checks

Assumption A1 consists of two components: (a) that the unobserved heterogeneity is at the site level, and (b) that this unobserved heterogeneity is time constant. To fix ideas, suppose that the

[6] An advantage of the Poisson-QMLE estimator is that it is robust to functional form misspecification. Therefore, we check the robustness of our results using the Poisson estimator. Results are in line with the main results and available upon request.

unobserved factors r_{it} can be decomposed as $r_{it} = q_{it} + \varepsilon_{it}$, where q_{it} is potentially correlated with age and ε_{it} is the independent component of the error term. Assumption A1 postulates that $q_{it} = q_j$, where j indexes the different sites. Part (a) of the assumption requires that for any pair of reactors (i, i') in site j, it holds $q_{it} = q_{i't} = q_{jt}$. Part (b) of the assumption amounts to omitting the dependence of q_{it} on t.

In the absence of a quasi-experimental setting, we first provide indirect evidence for part (a). In principle, allowing for reactor-specific fixed effects would be desirable since there could exist some unobserved heterogeneity at the reactor level. However, allowing for reactor-specific FE is not possible due to perfect multicollinearity of age, time dummies, and reactor FE. We tackle this problem in an alternative way: instead of using more coarse definition of cohorts (as in the main results), we now aggregate on the time scale. In particular, instead of adding time dummies for each year to the regression, we add three period dummies which respectively take the value of 1 when the year of observation belongs to the periods 1997-2003, 2004-2009, 2010-2015. This strategy allows us to include reactor-specific FE. By doing so, and if there exist unobserved heterogeneity correlated with age at the reactor level, then the results of our main results should be biased and thus differ from those obtained in this aggregated time-span specification. Nevertheless, the results are shown in table B.3 in appendix F, and are in line with our main results.

We now target part (b). While this assumption is not testable, indirect evidence can be provided. To do so, we use a novel approach that compares the estimates produced with a standard linear within and first difference (FD) approaches. Under assumption A1, the within and FD estimators are asymptotically equivalent. Therefore, although in finite samples the estimates produced with these two approaches under A1 are likely to differ[7], one should not expect substantial qualitative differences in the implications. On the other hand, if A1 is violated, then the two estimators might produce very different results. In particular, omitting a time varying first differenced variable q_{it} could lead to a different bias then when the omitted variable is an averaged q_{it} as in the within approach. A very large difference in the conclusions produced with these two approaches would therefore point at violation of A1.

We therefore reestimate the main linear specification (whose results are presented in table 3.4 on page 81) with an FD estimator. The estimates are shown in table B.4 on page 173. The results in 14 out of 19 nuclear plants have the same sign as in the main regression, and in 12 out

[7]This is because the FD uses less observations: the first period of observation cannot be lagged and is therefore dropped; in addition, there is a pure mechanical difference in how the values are calculated, which leads to deviations in small samples and vanishes asymptotically).

of those 14 they have also they same statistical significance indication. The bathtub is found consistently with the main results. In addition, the magnitudes of the coefficients remain very close to the initial ones. Thus, both qualitative and quantitative implications derived on basis of the two methods are comparable. As a representative example, we again look at the Chooz site. The estimated age effect for this reactor exhibits a bathtub pattern as in the initial result, with a maximum safety achieved after approximately 12 years (compared to the 7 years with the within estimation).

Remark. Note that this exercise cannot be taken as a formal test for time varying unobserved heterogeneity. In particular, there could be cases of time patterns of q_{it} for which within and FD estimators remain very similar. Put differently, it is not clear against which alternatives there will be statistical power. Nevertheless, the results presented above add additional evidence for the plausibility of assumption A1.

B.5 Additional tables

Table B.2: Additional regresssion results: analysis of the ACP bias

VARIABLES	(1) SFG	(2) SFG	(3) SFG	(4) SFG
900×AGE	-0.041	0.12***	-0.17	-0.066
1300×AGE	-0.022	0.12***	-0.068	0.049
1450×AGE	-0.045	0.11*	-0.024	0.11
900×AGE^2			0.0029	0.0040
1300×AGE^2			0.0017	0.0023
1450×AGE^2			-0.00036	0.00043
FOAS	-0.0026	-0.23**	-0.0023	-0.22**
FOAK	-0.074	0.18	-0.089	0.13
Capacity = 1300MW		1.25**		-0.22
Capacity = 1450MW		2.51***		0.46
Blayais	-0.40		0.68	
Bugey	0.54		1.59	
Cattenom	-0.69***		-0.69***	
Chinon B	-0.48		0.55	
Chooz B	-0.043		-0.44	
Civaux	-0.60		-0.99	
Cruas	-1.00		0.059	
Dampierre	-0.33		0.74	
Fessenheim	0.86		1.86	
Flamanville	0.18		0.15	
Golfech	-0.98***		-0.98***	
Gravelines	-1.16		-0.098	
Nogent	-0.045		-0.044	
Paluel	0.16		0.12	
Penly	-0.39***		-0.37***	
Saint-Alban	-0.15		-0.18	
Saint-Laurent	-0.025		1.05	
Tricastin	-0.42		0.65	
1998	-0.25	-0.42	-0.23	-0.39
1999	-0.20	-0.51	-0.17	-0.44
2000	-0.21	-0.68**	-0.16	-0.59**
2001	-0.11	-0.72**	-0.056	-0.62*
2002	0.077	-0.68**	0.13	-0.56
2003	-0.31	-1.23***	-0.26	-1.11**
2004	-0.62	-1.69***	-0.57	-1.57***
2005	0.014	-1.22***	0.044	-1.11**
2006	-1.19**	-2.57***	-1.17**	-2.47***
2007	-0.71	-2.24***	-0.72	-2.15***
2008	-1.53**	-3.21***	-1.56***	-3.16***
2009	-0.20	-2.03***	-0.26	-2.01***
2010	-0.69	-2.67***	-0.78	-2.69***
2011	-0.40	-2.53***	-0.52	-2.59***
2012	-0.0071	-2.30***	-0.17	-2.40***
2013	-0.60	-3.04***	-0.81	-3.20***
2014	0.0050	-2.59***	-0.25	-2.81***
2015	-1.03	-3.76***	-1.34	-4.06***

ASN data. This table presents four estimations, using a panel of 1100 observed reactor.years across 58 reactors and 19 years (1997-2015). In specifications (1) and (2), AGE is interacted with capacity group dummies. Specification (1) includes site fixed effects and year dummies. Specification (2) includes capacity group and time dummies, to measure the biases associated with the omission of cohort controls. Specifications (3) and (4) are respectively identical to (1) and (2), but allow for non-linear treatment effects. Standard errors are clustered at the site level. Intercepts are omitted. Significance: ***1%; **5%; *10%.

Table B.3: Robustness checks: aggregated time periods and reactor fixed-effects

VARIABLES	(1) ASD	(2) SFG
Belleville×AGE	-0.38***	0.0017
Blayais×AGE	0.18**	-0.52***
Bugey×AGE	0.024	-0.095
Cattenom×AGE	-0.00027	0.23***
Chinon×AGE	-0.18***	0.069
Chooz×AGE	-0.25***	-0.21***
Civaux×AGE	-0.14***	0.26***
Cruas×AGE	-0.065	-0.19**
Dampierre×AGE	-0.44***	0.25**
Fessenheim×AGE	-0.63***	0.28*
Flamanville×AGE	0.095	-0.16*
Golfech×AGE	-0.14***	-0.54***
Gravelines×AGE	-0.18***	0.50***
Nogent×AGE	-0.14**	0.045
Paluel×AGE	0.30***	-0.45***
Penly×AGE	0.042	0.62***
St-Alban×AGE	0.13*	-0.55***
St-Laurent×AGE	0.050	-1.01***
Tricastin×AGE	-0.18**	-0.20*
Belleville×AGE^2	0.0081***	0.00040
Blayais×AGE^2	-0.0053***	0.011***
Bugey×AGE^2	-0.0018	-0.00067
Cattenom×AGE^2	-0.00054	-0.0087***
Chinon×AGE^2	0.0033***	-0.0020
Chooz×AGE^2	0.0088***	0.0068***
Civaux×AGE^2	0.00068	-0.014***
Cruas×AGE^2	0.0014	0.0026
Dampierre×AGE^2	0.0077***	-0.0068***
Fessenheim×AGE^2	0.011***	-0.0031
Flamanville×AGE^2	-0.0044***	0.0033
Golfech×AGE^2	0.0043***	0.017***
Gravelines×AGE^2	0.0031***	-0.014***
Nogent×AGE^2	0.0015	-0.000097
Paluel×AGE^2	-0.0091***	0.0095***
Penly×AGE^2	-0.0013	-0.025***
St-Alban×AGE^2	-0.0040**	0.013***
St-Laurent×AGE^2	-0.0018	0.018***
Tricastin×AGE^2	0.0030*	0.0039*
$\mathbb{1}_{1997-2003}$	-0.19	1.00
$\mathbb{1}_{2004-2009}$	-0.019	0.69
$\mathbb{1}_{2010-2015}$	-0.021	0.94*
Reactor FE	Y	Y

ASN data. This table presents the result of two estimations of the effect of age on safety, using a panel of 1100 observed reactor.years across 58 reactors and 19 years (1997-2015). In both specifications, non-linear and heterogeneous treatment effects are allowed for. The first specification uses automatic shut-downs (ASDs) as the dependent variable. The second specification uses safeguard events as the dependent variable to test whether results obtained on ASDs can be generalized to other PDD events. Both specifications include aggregated time dummies and reactor fixed effect, to test the possibility of reactor-level unobserved heterogeneity. Standard errors are clustered at the site level. Intercepts are omitted. Significance: ***1%; **5%; *10%.

Table B.4: Robustness checks: first-difference estimation

VARIABLES	(1) ASD	(2) SFG	(3) ALL
Belleville*AGE	-0.53***	-0.24***	-4.16***
Blayais*AGE	0.79***	-0.080	1.19*
Bugey*AGE	-0.24	-0.65***	-3.43***
Cattenom*AGE	-0.16*	0.084*	-0.10
Chinon B*AGE	-0.16	-0.074	-0.41
Chooz B*AGE	-0.88***	-0.20***	-1.27***
Civaux*AGE	-0.17***	0.065*	1.34***
Cruas*AGE	-0.71***	-0.16**	-5.07***
Dampierre*AGE	-0.67***	0.062	-2.15***
Fessenheim*AGE	-0.66***	-0.23**	3.78***
Flamanville*AGE	0.32**	-0.23***	0.60
Golfech*AGE	-0.26***	-0.46***	-1.55***
Gravelines*AGE	-0.53***	-0.061	-3.24***
Nogent*AGE	-0.68***	-0.20***	-2.59***
Paluel*AGE	0.31**	-0.31***	-1.07*
Penly*AGE	-0.053	0.0077	-1.63***
Saint-Alban*AGE	0.10	-0.82***	-2.19***
Saint-Laurent B*AGE	-0.62***	-1.53***	-4.13***
Tricastin*AGE	-0.46***	-0.065	-2.50***
Belleville*AGE^2	0.012***	0.0057***	0.13***
Blayais*AGE^2	-0.017***	0.0014	-0.020
Bugey*AGE^2	0.0028	0.010***	0.068***
Cattenom*AGE^2	0.0042	-0.0024*	0.010
Chinon B*AGE^2	0.0033	0.0014	0.015
Chooz B*AGE^2	0.031***	0.0062***	0.038**
Civaux*AGE^2	0.0058*	-0.0037**	-0.053***
Cruas*AGE^2	0.016***	0.0033**	0.12***
Dampierre*AGE^2	0.011***	-0.0012	0.043***
Fessenheim*AGE^2	0.013***	0.0055***	-0.059***
Flamanville*AGE^2	-0.0084**	0.0036**	-0.0072
Golfech*AGE^2	0.0069**	0.013***	0.053***
Gravelines*AGE^2	0.0096***	0.0012	0.064***
Nogent*AGE^2	0.016***	0.0062***	0.075***
Paluel*AGE^2	-0.0082**	0.0052***	0.026*
Penly*AGE^2	-0.0010	-0.0012	0.048***
Saint-Alban*AGE^2	0.00025	0.018***	0.063***
Saint-Laurent B*AGE^2	0.0100***	0.027***	0.078***
Tricastin*AGE^2	0.0081**	0.0015	0.055***
Year FE	Y	Y	Y

ASN data. This table presents the result of three first-difference estimations of the effect of age on safety, using a panel of 1042 observed reactor.years (the FD operation requires to drop one observation per reactor) across 58 reactors and 19 years (1997-2015). In all specifications, non-linear and heterogeneous treatment effects are allowed for. The first specification uses automatic shut-downs (ASDs) as the dependent variable. The second specification uses safeguard events as the dependent variable to test whether results obtained on ASDs can be generalized to other PDD events. The third specification uses counts of safety significant events to check whether the results obtained for counts of PDD events can be generalized to non PDD events. All specifications include aggregated time dummies. Standard-errors are clustered at the site level. Intercepts are omitted. Significance: ***1%; **5%; *10%.

Table B.5: Bathtub effect from the OLS fixed-effect estimation on the ASD subsample

Site	Bathtub	Design	Conception	Size	Subdivision
Bugey	0	900	CP0	4	Lyon
Fessenheim	1	900	CP0	2	Strasbourg
Blayais	0	900	CP1	4	Bordeaux
Gravelines	1	900	CP1	6	Lille
Tricastin	1	900	CP1	4	Lyon
Dampierre	1	900	CP1	4	Orléans
Cruas	0	900	CP2	4	Lyon
Saint-Laurent B	0	900	CP2	2	Orléans
Chinon B	1	900	CP2	4	Orléans
Flamanville	0	1300	P4	2	Caen
Paluel	0	1300	P4	4	Caen
Saint-Alban	0	1300	P4	2	Lyon
Golfech	1	1300	P'4	2	Bordeaux
Penly	0	1300	P'4	2	Caen
Nogent	0	1300	P'4	2	Chalon
Belleville	1	1300	P'4	2	Orléans
Cattenom	0	1300	P'4	4	Strasbourg
Civaux	1	1450	N4	2	Bordeaux
Chooz B	1	1450	N4	2	Chalon

ASN Data. This table summarizes the estimation of the bathtub effect performed in the OLS fixed effect specification using the ASD subsample and presented in table 3.4. In the Bathtub column, a site is associated with the value 1 if the estimated coefficient of the $AGE \times Site$ variable is negative and significant and if the coefficient of the $AGE^2 \times Site$ variable is positive and significant. Otherwise, the site is associated with the value 0. Information regarding reactor design, conception, size and local regulatory subdivision are added to the table to search for potential patterns explaining the bathtub effect.

APPENDIX C

Robustness and consequences of the effect of monitoring on compliance

C.1 Robustness checks for chapter 4

In this appendix, we present robustness checks of our main results (e.g. the regressions based on the unrestricted counts of events) to several definitions of our treatment variable and of our estimators. In particular, we propose additional results obtained under the five following estimators: the OLS estimator, the Poisson estimator, the 2SLS estimator, the GMM-IV Poisson estimator suggested by Cameron and Trivedi (2013) and the Poisson Control Function approach suggested by Wooldridge (2002b, 2015). The final two estimators are included in order to account for the count nature of the dependent variable. We run these estimators four times, where each iteration of the estimator is characterized by different levels of fixed-effects. In table C.1, only year fixed-effects are included. In table C.2, technological design fixed-effects are included.[1] In table C.3, site fixed-effects are included. In table C.4, reactor fixed-effects are included.

In these tables, we propose alternative definitions of the treatment variable. Specifically, we test the robustness of our results to changes in the definition of the *budget* variable. To do so, we compare the results obtained when the treatment is defined as the annual budget of the local agencies, as the log of these budgets, and as the log of the ratio of these budget on the average budget calculated on the whole sample.

We also add to table C.1 and C.2 four alternative measures of the activity of a local agency. Due to a lack of data, these alternative treatments are time-invariant. We first propose to measure activity by a dummy variable $Meeting$ capturing whether an agency organizes meetings

[1] More specifically, these technological design fixed-effects capture the three major reactor designs used in the French fleet, which mainly differ by their nominal capacity.

that the public can freely attend. We also propose to measure activity by a variable *Publishing* that characterizes the frequency with which an agency publishes public communication materials. A third alternative measure of activity is captured by a treatment variable $Members$, defined as the number of members of an agency.[2] A final alternative measure of a commission's activity is captured by variable *Expertise*, defined as the number of independent expertise commissioned by a local agency since 2000.[3]

As all agencies communicate with the public and interact with plant managers, it is not clear whether the effects measured in chapter 4 are caused by the communication of the agencies or by their monitoring activities. In other words, measuring a significant positive causal effect of the agencies' activities on reporting behaviours could be explained in two different ways. First, managers could anticipate that the communication performed by the agency may lead to costly public backlashes (e.g. an increase in expected penalties). Second, the monitoring performed by the agency could directly affect the penalties perceived by plant managers. The four alternative measures proposed in this appendix aim to provide tentative evidence regarding these different channels. However, their time-invariance prevent us from considering these results as robust.

From the observation of these four tables, it appears that the coefficients associated with the *budget* variable are positive and significant across all specifications that account for endogeneity. The coefficients associated with the $\log\left(\frac{\text{budget}}{\mathbb{E}(\text{budget})}\right)$ treatment are also positive and significant under no fixed effects and under technological fixed-effects. However, our results are not robust to the definition of the treatment as log(budget). This is probably due to the fact that taking the log of the budget variable reduces its already small variations in a way that precludes the identification of any significant trends.

Another takeaway from these table is that accounting for the endogeneity of the treatment variable changes the sign of the treatment coefficient obtained under classical OLS or Poisson specifications. In these uncorrected estimations, and for the three different definitions of the budget treatment, the coefficients associated with the treatment variable are almost systematically negative and significant (regardless of the definition of the fixed-effects included). As discussed previously in chapter 4, the sign of this bias is intuitive, yet the high level of significance of these

[2]The members of a local agency are the people designated to represent the local population during the annual meetings with plant managers, it does not account for possible administrative personnel. This number varies from 40 to 71 in the sample. This variation is possible as the law only sets standards regarding the proportion of members that have to be elected, or that have to be part of EDF's workers' unions, part of an environmental agency, or part of the local population.

[3]This measure is particularly prone to measurement biases, as these missions were retrieved through e-mails and phone calls with the agencies' administrative personnel as well as through the exploration of the agencies' websites.

Table C.1: Robustness checks: year fixed-effects only

	Year fixed-effects only				
Treatment	OLS	Poisson	2SLS	IV-Poisson	CF-Poisson
budget	-0.0212***	-0.00168***	0.0651**	0.00512**	0.00509***
	(0.00755)	(0.000600)	(0.0281)	(0.00211)	(0.00170)
log(budget)	-1.176***	-0.0928***	2.402*	0.173**	0.187**
	(0.386)	(0.0290)	(1.310)	(0.0883)	(0.0787)
$\log\left(\frac{\text{budget}}{\mathbb{E}(\text{budget})}\right)$	-1.037***	-0.0811***	4.303**	0.328**	0.340**
	(0.377)	(0.0288)	(1.873)	(0.132)	(0.136)
Public Meeting	-2.553***	-0.189***	14.28	0.907	1.002
	(0.733)	(0.0570)	(13.59)	(0.688)	(0.728)
Communication	0.463	0.0301	2.388	0.175	0.173*
	(0.316)	(0.0231)	(1.581)	(0.126)	(0.1000)
Members	-0.0291	-0.00196	0.290	0.0238	0.0204
	(0.0243)	(0.00185)	(0.289)	(0.0244)	(0.0213)
Expertise	-0.997***	-0.0782***	1.623*	0.102*	0.113**
	(0.224)	(0.0182)	(0.853)	(0.0547)	(0.0552)

The table presents the value and statistical significance of $\beta_{treatment}$ for different definitions of the treatment variable and for several estimators. All significance levels are based on robust standard errors, except for the control function estimator, which was estimated using bootstrapped standard errors.

uncorrected coefficients may cast some doubt on the strength of our instrument.

When including reactor fixed-effects, the results associated with the control-function Poisson estimator under the logged definitions of the budget treatment become negative and highly significant. These results are due to the fact that the first stage of this estimator, in which the treatment variable is regressed on the instrument, is not significant when these fixed effects are included.

Finally, the alternative measures of an agency's activities associated with communication (e.g. Meeting and Publishing) are all negative in tables C.1 and C.2. On the contrary, the coefficients associated with the Expertise treatment are positive and significant in all specifications accounting for endogeneity and including no fixed effects. Although the robustness of the results obtained under these alternative measures is weak, they suggest that the effect of the agencies' activities may be due to their monitoring activities rather than to their communication activities.

C.2 Out-of-sample predictions for chapter 4

In order to derive a quantitative analysis of the possible policy recommendations associated with our results, we propose an analysis of the predicted values of one of our fitted models

Table C.2: Robustness checks: design fixed-effects

	Design fixed-effects				
Treatment	OLS	Poisson	2SLS	IV-Poisson	CF-Poisson
budget	-0.0161**	-0.00138**	0.0571**	0.00469**	0.00435***
	(0.00808)	(0.000641)	(0.0244)	(0.00193)	(0.00168)
log(budget)	-1.031***	-0.0858***	1.110	0.0902	0.0898
	(0.385)	(0.0286)	(1.058)	(0.0804)	(0.0748)
$\log\left(\frac{\text{budget}}{\mathbb{E}(\text{budget})}\right)$	-0.589	-0.0513*	3.717**	0.287**	0.290**
	(0.406)	(0.0305)	(1.595)	(0.116)	(0.118)
Public Meeting	-2.331***	-0.173***	6.694	0.525	0.441
	(0.694)	(0.0545)	(9.544)	(0.658)	(0.616)
Communication	-0.359	-0.0296	1.252	0.0898	0.0948
	(0.370)	(0.0280)	(1.577)	(0.117)	(0.0972)
Members	-0.0343	-0.00241	0.0878	0.00607	0.00455
	(0.0271)	(0.00211)	(0.235)	(0.0216)	(0.0185)
Expertise	-1.122***	-0.0907***	0.856	0.0555	0.0551
	(0.229)	(0.0191)	(0.771)	(0.0537)	(0.0538)

The table presents the value and statistical significance of $\beta_{treatment}$ for different definitions of the treatment variable and for several estimators. All significance levels are based on robust standard errors, except for the control function estimator, which was estimated using bootstrapped standard errors.

Table C.3: Robustness checks: site fixed-effects

	Site fixed-effects				
Treatment	OLS	Poisson	2SLS	IV-Poisson	CF-Poisson
budget	-0.0192	-0.00164	0.137*	0.0131*	0.0109**
	(0.0177)	(0.00174)	(0.0747)	(0.00683)	(0.00540)
log(budget)	-2.658***	-0.207***	-68.32	-2.198	-5.817***
	(0.839)	(0.0785)	(137.4)	(1.381)	(2.158)
$\log\left(\frac{\text{budget}}{\mathbb{E}(\text{budget})}\right)$	-2.196***	-0.177**	-55.99	-1.859	-4.551**
	(0.716)	(0.0797)	(112.5)	(1.270)	(2.060)

The table presents the value and statistical significance of $\beta_{treatment}$ for different definitions of the treatment variable and for several estimators. All significance levels are based on robust standard errors, except for the control function estimator, which was estimated using bootstrapped standard errors.

in an hypothetical world. In this hypothetical world, we assume that budgets are equally distributed among commissions under a constant public expenditure constraint. In other words, we distribute the sum of the budgets granted in reality equally among all commissions, and estimate the counts of reports predicted by one of our regression models.

To perform these out of sample predictions, we use the Control-Function Poisson estimator presented in table C.4, in which reactor fixed effects are included, and the treatment variable is the budget of the reactors. In this specification, the coefficient associated with the treatment variable is positive and very significant, and the count nature of the estimator prevents negative

Table C.4: Robustness checks: reactor fixed-effects

Treatment	Reactor fixed-effects OLS	Poisson	2SLS	IV-Poisson	CF-Poisson
budget	-0.0222	-0.00197	0.132*	0.0127*	0.0105*
	(0.0228)	(0.00172)	(0.0736)	(0.00683)	(0.00622)
log(budget)	-2.691**	-0.215***	-63.66	-2.096	-5.421*
	(1.077)	(0.0751)	(124.7)	(1.280)	(3.053)
$\log\left(\frac{\text{budget}}{\mathbb{E}(\text{budget})}\right)$	-2.226**	-0.183**	-52.21	-1.777	-4.253***
	(1.077)	(0.0752)	(102.7)	(1.194)	(1.290)

The table presents the value and statistical significance of $\beta_{treatment}$ for different definitions of the treatment variable and for several estimators. All significance levels are based on robust standard errors, except for the control function estimator, which was estimated using bootstrapped standard errors.

predictions.

This estimator is computed following Wooldridge (2002b) and Wooldridge (2015), by performing a fixed-effect linear regression of the treatment variable *budget* on the instrument:

$$budget_{it} = \beta_{INSTR} INSTR_{it} + \beta_A \cdot Age_{it} + \beta_P \cdot Prod_{it} + \beta_M \cdot Maint_{it} + \delta_t + \gamma_i + \epsilon_{it} \quad \text{(C.1)}$$

where δ_t and γ_i respectively capture year and reactor fixed-effects. Age, $Prod$ and $Maint$ respectively refer to the age of reactors, their annual production levels and the quantity of maintenance undergone annually. We then predict the residuals of this first stage, $\hat{e}_{it}$, and include them in a Poisson regression of our dependent variable:

$$\mathbb{E}(ALL|X) = \exp\left(\beta_{budget} budget_{it} + \beta_{\hat{e}} \hat{e}_{it} + \beta_A \cdot Age_{it} + \beta_P \cdot Prod_{it} + \beta_M Maint_{it} + \delta_t + \gamma_i + \epsilon'_{it}\right) \quad \text{(C.2)}$$

in which exp denotes the exponential function, X denotes the regressors present on the right hand side of equation (C.2), and the variance of ALL is supposed to be equal to its mean, as is usual in Poisson regressions.

We fit this model on our initial dataset, and obtain the results from which the upper cell of the final column of figure C.4 is taken.[4] Then, we look at the counts of reports predicted by this model, using the year 2014 as a reference. 2014 is chosen as a reference as this is the year in our sample that has the largest number of observations. For the observations made in 2014 only, we replace the value of the treatment (i.e. the budget) by its average value within this sub-sample. We then estimate the counts of reports predicted by the fitted model within these

[4]This estimator is computed using bootstrapped standard errors.

48 reactors using the new equally distributed budget as the value of the treatment.

The results of this prediction show that in the original 2014 sub-sample, the average number of declarations per reactor was 11.9. This number could have been as large as 13.6 according to the predictions of our model and provided the 2014 budgets had been distributed equally among commissions. This quantitative analysis shows that fine-tuning the way budgets are allocated across these commissions could lead to significant increases in the amount of information that is known to the regulator and that can be used to increase nuclear safety.

Bibliography

Aarset, M. V. (1987). How to identify a bathtub hazard rate. *IEEE Transactions on Reliability*, 36(1):106–108.

Adler, M. D. and Sanchirico, C. (2006). Inequality and uncertainty: Theory and legal applications. *University of Pennsylvania Law Review*, (155):279–377.

Adrangi, B., Chow, G., and Raffiee, K. (1997). Airline deregulation, safety and profitability. *Transportation Journal*, 36(4):44–52.

Alary, D., Gollier, C., and Treich, N. (2013). The effect of ambiguity aversion on insurance and self-protection. *The Economic Journal*, 123(573):1188–1202.

Allon, G. and Bassamboo, A. (2011). Buying from the babbling retailer? the impact of availability information on customer behavior. *Management Science*, 57(4):713–726.

Allon, G., Bassamboo, A., and Gurvich, I. (2011). "We will be right with you": Managing customer expectations with vague promises and cheap talk. *Operations research*, 59(6):1382–1394.

Alt, F. L. (1972). Archaelogy of computers: Reminiscences, 1945-1947. *Communications of the ACM*, 15(7):693–694.

Angrist, J. and Pischke, J. (2009a). A Note on Bias in Just Identified IV with Weak Instruments (referering to Harmless Econometrics: An Empiricist's Companion).

Angrist, J. and Pischke, J. (2009b). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Antic, N. and Persico, N. (2016). Communication among shareholders. *Working Paper*.

Armstrong, G. A. (1981). Consumer safety and the regulation of industry. *Managerial and Decision Economics*, 2:67–73.

Arrow, K. J., Cropper, M. L., Eads, G. C., Hahn, R. W., et al. (1996). Is there a role for benefit-cost analysis in environmental, health, and safety regulation? *Science*, 272(5259):221.

Athanassoglou, S. and Xepapadeas, A. (2012). Pollution control with uncertain stock dynamics: when, and how, to be precautious. *Journal of Environmental Economics and Management*, 63(3):304–320.

Banerjee, P. and Shogren, J. F. (2010). Regulation, reputation, and environmental risk. *Economics Letters*, 106(1):45–47.

Barham, B. L., Chavas, J.-P., Fitz, D., Salas, V. R., and Schechter, L. (2014). The roles of risk and ambiguity in technology adoption. *Journal of Economic Behavior & Organization*, 97:204–218.

Barke, R. P. and Jenkins-Smith, H. C. (1993). Politics and scientific expertise: scientists, risk perception, and nuclear waste policy. *Risk Analysis*, 13(4):425–439.

Bastide, S., Moatti, J.-P., Fagnani, F., et al. (1989). Risk perception and social acceptability of technologies: the french case. *Risk analysis*, 9(2):215–223.

Becker, G. S. (1974). Crime and punishment: An economic approach. In *Essays in the Economics of Crime and Punishment*, pages 1–54. NBER.

Bell, A. and Jones, K. (2013). The impossibility of separating age, period and cohort effects. *Social science & medicine*, 93:163–165.

Bell, A. and Jones, K. (2014). Another futile quest? a simulation study of yang and land's hierarchical age-period-cohort model. *Demographic Research*, 30:333.

Bell, A. and Jones, K. (2016). The hierarchical age-period-cohort model: Why does it find the results that it finds? *Quality & Quantity*, pages 1–17.

Berger, L. (2015). The impact of ambiguity prudence on insurance and prevention. *FEEM Nota di Lavoro, 15.2015.*

Berger, L. and Bosetti, V. (2016). Ellsberg re-revisited: An experiment disentangling model uncertainty and risk aversion. *FEEM Nota di Lavoro, 37.2016.*

Berger, L., Emmerling, J., and Tavoni, M. (2016). Managing catastrophic climate risks under model uncertainty aversion. *Management Science*.

Berthélemy, M. and Rangel, L. E. (2015). Nuclear reactors' construction costs: The role of lead-time, standardization and technological progress. *Energy Policy*, 82:118–130.

Bewley, T. F. (2003). Knightian decision theory, part i. *Decisions in Economics and Finance*, 25:79–110.

Biais, B., Mariotti, T., Rochet, J.-C., and Villeneuve, S. (2010). Large risks, limited liability, and dynamic moral hazard. *Econometrica*, 78(1):73–118.

Bizet, R. and Lévêque, F. (2017). The economic assessment of the cost of nuclear accidents. In *Resilience: A New Paradigm of Nuclear Safety*, pages 79–95. Springer.

Boccard, N. (2014). The cost of nuclear electricity: France after fukushima. *Energy Policy*, 66:450–461.

Boer, J., Wouter Botzen, W., and Terpstra, T. (2014). Improving flood risk communication by focusing on prevention-focused motivation. *Risk analysis*, 34(2):309–322.

Bommier, A. and Zuber, S. (2008). Can preferences for catastrophe avoidance reconcile social discounting with intergenerational equity? *Social Choice and Welfare*, 31(3):415–434.

Borenstein, S. and Zimmerman, M. B. (1988). Market incentives for safe commercial airline operation. *The American Economic Review*, 78(5):913–935.

Bose, S. and Renou, L. (2014). Mechanism design with ambiguous communication devices. *Econometrica*, 82(5):1853–1872.

Bovens, L. and Fleurbaey, M. (2012). Evaluating life or death prospects. *Economics & Philosophy*, 28(2):217–249.

Boyer, M. and Laffont, J.-J. (1997). Environmental risks and bank liability. *European Economic Review*, 41(8):1427–1459.

Boyer, M. and Laffont, J.-J. (1999). Toward a political theory of the emergence of environmental incentive regulation. *The RAND Journal of Economics*, pages 137–157.

Brandt, U. S. (2014). The implication of extreme events on policy responses. *Journal of Risk Research*, 17(2):221–240.

Bressoux, P., Kramarz, F., and Prost, C. (2009). Teachers' training, class size and students' outcomes: Learning from administrative forecasting mistakes. *Economic Journal*, 119(536):540–561.

Brown, J. P. (1973). Toward an economic theory of liability. *The Journal of Legal Studies*, 2:323–349.

Brunette, M., Cabantous, L., Couture, S., and Stenger, A. (2013). The impact of governmental assistance on insurance demand under ambiguity: a theoretical model and an experimental test. *Theory and decision*, pages 1–22.

Buchanan, J. M. (1975). *Altruism, Morality and Economic Theory*, chapter The Samaritan's Dilemma. New York: Russel Sage Foundation.

Burgherr, P. and Hirshberg, S. (2014). Comparative risk assessment of severe accidents in the energy sector. *Energy Policy*, pages S46–S57.

Butler, D. (2011). Reactors, residents and risk. *Nature*, 474:36.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression analysis of count data*, volume 53. Cambridge university press.

Carlsson, F., Daruvala, D., and Jaldell, H. (2012). Do administrators have the same priorities for risk reductions as the general public? *Journal of Risk and Uncertainty*, 45(1):79–95.

Cason, T. N., Friesen, L., and Gangadharan, L. (2016). Regulatory performance of audit tournaments and compliance observability. *European Economic Review*, 85:288–306.

CBO (2008). Nuclear power's role in generating electricity. Technical report, U.S. Congressional Budget Office.

Chateauneuf, A., Kast, R., and Lapied, A. (1996). Choquet pricing for financial markets with frictions. *Mathematical Finance*, 6(3):323–330.

Chen, Y., Kartik, N., and Sobel, J. (2008). Selecting cheap-talk equilibria. *Econometrica*, 76(1):117–136.

Chen, Y., Wang, Z., Qiu, J., Zheng, B., and Huang, H.-Z. (2011). Adaptive bathtub hazard rate curve modelling via transformed radial basis functions. In *2011 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering (ICQR2MSE)*, pages 110–114. IEEE.

Chichilnisky, G. (2000). An axiomatic approach to choice under uncertainty with catastrophic risks. *Resource and Energy Economics*, 22(3):221–231.

Coase, R. H. (1960). The problem of social cost. *Journal of Law and Economics*, 3:1–44.

Coate, S. (1995). Altruism, the samaritan's dilemma, and government transfer policy. *The American Economic Review*, pages 46–57.

Cochran, T. B. (2012). Global implications of the fukushima disaster for nuclear power. In *International Seminar on Nuclear War and Planetary Emergencies: 44th Session: the Role of Science in the Third Millenium*, page 137. World Scientific.

Cohen, B. L. (1990). A test of the linear-no threshold theory of radiation carcinogenesis. *Environmental research*, 53(2):193–220.

Cohen, B. L. (1995). Test of the linear-no threshold theory of radiation carcinogenesis for inhaled radon decay products. *Health Physics*, 68(2):157–174.

Cour des Comptes (2014). Le coût de production de l'électricité nucléaire. Technical report, Communication à la commission d'enquête de l'assemblée nationale.

Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica*, pages 1431–1451.

Crès, H., Gilboa, I., and Vieille, N. (2011). Aggregation of multiple prior opinions. *Journal of Economic Theory*, (146):2563–2582.

Davis, L. W. (2012). Prospects for nuclear power. *The Journal of Economic Perspectives*, 26(1):49–65.

Davis, L. W. and Wolfram, C. (2012). Deregulation, consolidation, and efficiency: Evidence from US nuclear power. *American Economic Journal: Applied Economics*, 4(4):194–225.

Demsetz, H. (1972). When does the rule of liability matter? *Journal of Legal Studies*, 1:13–28.

D'Haeseleer, W. D. (2013). Synthesis on the economics of nuclear energy. Technical report, DG Energy.

Diamond, P. A. (1967). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility: Comment. *Journal of Political Economy*, (75):765–766.

Dionne, G., Gagné, R., Gagnon, F., and Vanasse, C. (1997). Debt, moral hazard and airline safety. an empirical evidence. *The Journal of Econometrics*, 79:379–402.

Doss, M. (2013). Linear no-threshold model vs. radiation hormesis. *Dose-response*, 11(4):dose–response.

Dow, J. and Werlang, S. R. d. C. (1992). Uncertainty aversion, risk aversion, and the optimal choice of portfolio. *Econometrica*, pages 197–204.

Downer, J. (2014). Disowning fukushima: Managing the credibility of nuclear reliability assessment in the wake of disaster. *Regulation & Governance*, 8(3):287–309.

Drakopoulos, S. A. and Theodossiou, I. (2016). Workers' risk underestimation and occupational health and safety regulation. *European Journal of Law and Economics*, 41(3):641–656.

Dreicer, M., Tort, V., and Manen, P. (1995). Externe: Externalities of energy vol. 5. nuclear. Technical report, European Commission.

Drottz-Sjöberg, B.-M. and Sjoberg, L. (1990). Risk perception and worries after the chernobyl accident. *Journal of Environmental Psychology*, 10(2):135–149.

Du, Y. and Parsons, J. E. (2009). Update on the cost of nuclear power. *MIT Center for Energy and Environmental Policy Research Working Paper 09-004*.

Dubin, J. A. and Rothwell, G. S. (1990). Subsidy to nuclear power through price-anderson liability limit. *Contemporary Policy Issues*, VIII:73–79.

Duflo, E., Greenstone, M., Pande, R., and Ryan, N. (2013). Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. *The Quarterly Journal of Economics*.

Ecofys (2014). Subsidies and costs of eu energy. Technical report, Report for the European Commission.

Eeckhoudt, L., Schieber, C., and Schneider, T. (2000). Risk aversion and the external cost of a nuclear accident. *Journal of Environmental Management*, pages 109–117.

Ehrlich, I. and Becker, G. S. (1972). Market insurance, self-insurance, and self-protection. *The Journal of Political Economy*, 80:623–648.

EIA (2016). Capital cost estimates for utility scale electricity generating plants. Technical report, U.S. Energy Information Administration.

Eichberger, J., Grant, S., Kelsey, D., and Koshevoy, G. A. (2011). The α-meu model: A comment. *Journal of Economic Theory*, 146(4):1684–1698.

Ellsberg, D. (1961). Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, 75:643–669.

Engel, K., Frerks, G., Velotti, L., Warner, J., and Weijs, B. (2014). Flood disaster subcultures in the netherlands: the parishes of borgharen and itteren. *Natural Hazards*, 73(2):859–882.

Epstein, L. and Schneider, M. (2003). Recursive multiple-priors. *Journal of Economic Theory*, 113(1):1–31.

Epstein, L. G. and Schneider, M. (2008). Ambiguity, information quality and asset pricing. *The Journal of Finance*, 63(1):197–228.

Epstein, L. G. and Wang, T. (1994). Intertemporal asset pricing under knightian uncertainty. *Econometrica: Journal of the Econometric Society*, pages 283–322.

Epstein, R. A. (1996). Catastrophic responses to catastrophic risks. *Journal of Risk and Uncertainty*, 12(2-3):287–308.

Evans, M. F., Gilpatric, S. M., and Liu, L. (2009). Regulation with direct benefits of information disclosure and imperfect monitoring. *Journal of Environmental Economics and Management*, 57(3):284–292.

Ewers, H.-J. and Rennings, K. (1991). Die volkswirtschatlichen kosten eine super-gau's in biblis. *Zeitschrift für Umweltpolitik und Umweltrecht*, pages 379–396.

Ewers, H.-J. and Rennings, K. (1992). *Abschätzung der Schäden durch einen sogennanten Super-GAU*. Prognos, Bâle.

Fairlie, I. and Summer, D. (2006). The other report on chernobyl. Technical report, Berlin.

Farrell, J. and Gibbons, R. (1989). Cheap talk with two audiences. *The American Economic Review*, 79(5):1214–1223.

Faure, M. G. and Skogh, G. (1992). Compensation for damages caused by nuclear accidents: A convention as insurance. *The Geneva Papers on Risk and Insurance*, 17:499–513.

Feinstein, J. S. (1989). The Safety Regulation of U.S. Nuclear Power Plants: Violations, Inspections and Abnormal Occurences. *Journal of Political Economy*, 97:115–154.

Figueroa, P. M. (2013). Risk communication surrounding the fukushima nuclear disaster: an anthropological approach. *Asia Europe Journal*, 11(1):53–64.

Filer, R. and Golbe, D. L. (2003). Debt, operating margin and workplace safety. *Journal of Industrial Organization*, 51(3):359–381.

Fischhoff, B., Slovic, P., and Lichtenstein, S. (1983). "the public" vs."the experts": Perceived vs. actual disagreements about risks of nuclear power. In *The analysis of actual versus perceived risks*, pages 235–249. Springer.

Fleurbaey, M. (2010). Assessing risky social situations. *Journal of Political Economy*, 118(4):649–680.

Fleurbaey, M. and Zuber, S. (2017). Fair management of social risk. *Journal of Economic Theory*, 169:666–706.

Gajdos, T., Tallon, J.-M., and Vergneaud, J.-C. (2008). Representation and aggregation of preferences under uncertainty. *Journal of Economic Theory*, (141):68–99.

Ghirardato, P. (1994). Agency theory with non-additive uncertainty. *Unpublished working paper, Cal. Tech.*

Ghirardato, P., Maccheroni, F., and Marinacci, M. (2004). Differentiating ambiguity and ambiguity attitude. *Journal of Economic Theory*, 118:133–173.

Gilboa, I. (2004). *Uncertainty in Economic Theory - Essays in honor of David Schmeidler's 65th birthday*. Routledge - Taylor and Francis Group, London.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18:141–153.

Gilpatric, S. M., Vossler, C. A., and McKee, M. (2011). Regulatory enforcement with competitive endogenous audit mechanisms. *The RAND Journal of Economics*, 42(2):292–312.

Giraud, R. and Thomas, L. (2017). Ambiguity, optimism, and pessimism in adverse selection models. *Journal of Economic Theory*.

Glenn, N. D. (2005). *Cohort analysis*, volume 5. Sage.

Gonzalez, F. (2008). Precautionary principle and robustness for a stock pollutant with multiplicative risk. *Environmental and Resource Economics*, 41(1):25–46.

Gottardi, P., Tallon, J. M., and Ghirardato, P. (2017). Flexible contracts. *Games and Economic Behavior*, 103:145–167.

Grabowski, H. G. and Vernon, J. M. (1978). Consumer product safety regulation. *The American Economic Review*, 68:284–289.

Gray, W. B. (1987). The cost of regulation: Osha, epa and the productivity slowdown. *The American Economic Review*, 77(5):998–1006.

Gray, W. B. and Shimshack, J. P. (2011). The effectiveness of environmental monitoring and enforcement: A review of the empirical evidence. *Review of Environmental Economics and Policy*, 5(1):3–24.

Grepperud, S. (2014). Optimal safety standards when accident prevention depends upon both the firm and worker effort. *European Journal of Law and Economics*, pages 1–17.

Grubler, A. (2010). The costs of the french nuclear scale-up: A case of negative learning by doing. *Energy Policy*, 38(9):5174–5188.

Grüne-Yanoff, T. and Rosencrantz, H. (2011). Beneficial safety decreases. *Theory and decision*, 70(2):195–213.

Hammitt, J. K. (2013). Positive versus normative justifications for benefit-cost analysis: Implications for interpretation and policy. *Review of Environmental Economics and Policy*, 7(2):199–218.

Hansen, L. P. and Sargent, T. J. (2001). Robust control and model uncertainty. *The American Economic Review*, 91(2):60–66.

Hansson, I. and Skogh, G. (1987). Moral hazard and safety regulation. *The Geneva Papers on Risk and Insurance*, 12:132–144.

Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, 63(4):309–321.

Hasegawa, R. (2013). Disaster evacuation from japan's 2011 tsunami disaster and the fukushima nuclear accident. *Studies*.

Hausman, C. (2014). Corporate incentives and nuclear safety. *American Economic Journal: Economic Policy*, 6(3):178–206.

Hayashi, F. (2000). *Econometrics*. Princeton University Press.

Hayashi, M. and Hughes, L. (2013). The fukushima nuclear accident and its effect on global energy security. *Energy Policy*, pages 102–111.

Haynes, K., Barclay, J., and Pidgeon, N. (2008). The issue of trust and its influence on risk communication during a volcanic crisis. *Bulletin of Volcanology*, 70(5):605–621.

Heckman, J. J., Urzua, S., and Vytlacil, E. (2006). Understanding Instrumental Variables in Models with Essential Heterogeneity. *The Review of Economics and Statistics*, 88(3):389–432.

Henry, C. and Henry, M. (2002). Formalization and applications of the precautionary principles. Columbia Discussion Papers Series.

Heyes, A. and Heyes, C. (2000). An empirical analysis of the nuclear liability act (1970) in canada. *Resource and Energy Economics*, 22:91–101.

Heyes, A. and Liston-Heyes, C. (2000). Capping environmental liability: The case of north american nuclear power. *The Geneva Papers on Risk and Insurance*, 25:196–202.

Heyes, A. G. and Liston-Heyes, C. (1998). Subsidy to nuclear power through price-anderson liability limit: comment. *Contemporary Economic Policy*, XVI:122–124.

Hiriart, Y. and Martimort, D. (2012). How much discretion for risk regulators? *The RAND Journal of Economics*, 43(2):283–314.

Hiriart, Y., Martimort, D., and Pouyet, J. (2004). On the optimal use of ex ante regulation and ex post liability. *Economics Letters*, 84(2):231–235.

Hiriart, Y., Martimort, D., and Pouyet, J. (2010). The public management of risk: Separating ex ante and ex post monitors. *Journal of Public Economics*, 94:1008–1019.

Hirschberg, S., Burgherr, P., Spiekerman, G., and Dones, R. (2004). Severe accidents in the energy sector: comparative perspective. *Journal of Hazardous Materials*, 111(1):57–65.

Hofert, M. and Wüthrich, M. V. (2011). Statistical review of nuclear power accidents. *Asia-Pacific Journal of Risk and Insurance*, 7:1–13.

Hohmeyer, O. (1990). Latest results of the international discussion on the social costs of energy - how does wind compare today? Technical report, Madrid.

Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics*, pages 311–324.

Holford, T. R. (1991). Understanding the effects of age, period, and cohort on incidence and mortality rates. *Annual review of public health*, 12(1):425–457.

Hooker, A. M., Bhat, M., Day, T. K., Lane, J. M., Swinburne, S. J., Morley, A. A., and Sykes, P. J. (2004). The linear no-threshold model does not hold for low-dose ionizing radiation. *Radiation research*, 162(4):447–452.

HSE (2011). Step 4: Probabilistic Safety Analysis Assessment of the EDF and AREVA UK EPR Reactor. Assessment report: Onr-gda-ar-11-019, Office for Nuclear regulation.

Huang, L., Zhou, Y., Han, Y., Hammitt, J. K., Bi, J., and Liu, Y. (2013). Effect of the fukushima nuclear accident on the risk perception of residents near a nuclear power plant in china. *Proceedings of the National Academy of Sciences*, 110(49):19742–19747.

Huhtala, A. and Remes, P. (2017). Quantifying the social costs of nuclear energy: Perceived risk of accident at nuclear power plants. *Energy Policy*, 105:320–331.

IAEA (2006). *Chernobyl's Legacy: Health Environmental and Socio-Economic Impacts and Recommendations to the Governments of Belarus, the Russian Federation and Ukraine*. The Chernobyl Forum: 2003-2005, Vienna, second revised edition edition.

IAEA (2013). Periodic safety review for nuclear power plants, specific safety guide ssg-25. Technical report, International Atomic Energy Agency.

IEA/NEA (2015). Technology roadmap: Nuclear energy. Technical report, International Energy Agency.

IER (2013). Die risiken der kernenergie in deutschland im vergleich mit risiken anderer stromerzeugungstechnologien. Technical report, Preiss, P and Wissel, S and Fahl, U and Friedrich, R and Voß, A.

IRSN (2013). Méthodologie appliquée par l'IRSN pour l'estimation des coûts d'accidents nucléaires en France. Technical report.

Ju, N. and Miao, J. (2012). Ambiguity, learning, and asset returns. *Econometrica*, 80(2):559–591.

Karni, E. (2009). A reformulation of the maxmin expected utility model with application to agency theory. *Journal of Mathematical Economics*, 45(1):97–112.

Kasperson, R. E., Renn, O., Slovic, P., Brown, H. S., Emel, J., Goble, R., Kasperson, J. X., and Ratick, S. (1988). The social amplification of risk: A conceptual framework. *Risk Analysis*, 8:177–187.

Kathren, R. L. (1996). Pathway to a paradigm: The linear nonthreshold dose-response model in historical context. *Health Physics*.

Katzman, M. T. (1988). Pollution liability insurance and catastrophic environmental risk. *the Journal of Risk and Insurance*, 55:75–100.

Kawamura, K. (2011). A model of public consultation: why is binary communication so common? *The Economic Journal*, 121(553):819–842.

Keeler, T. E. (1994). Highway safety, economic behavior, and driving enforcement. *The American Economic Review*, 84(3):684–693.

Kellner, C. (2015). Tournaments as a response to ambiguity aversion in incentive contracts. *Journal of Economic Theory*, 159:627–655.

Kellner, C. and Le Quement, M. (2015). Endogenous ambiguity in cheap talk. *Discussion Papers in Economics and Econometrics*, (1701).

Keyes, K. M., Grant, B. F., and Hasin, D. S. (2008). Evidence for a closing gender gap in alcohol use, abuse, and dependence in the united states population. *Drug and alcohol dependence*, 93(1):21–29.

Keyes, K. M., Utz, R. L., Robinson, W., and Li, G. (2010). What is a cohort effect? comparison of three statistical methods for modeling cohort effects in obesity prevalence in the united states, 1971–2006. *Social science & medicine*, 70(7):1100–1108.

Kivimäki, M. and Kalimo, R. (1993). Risk perception among nuclear power plant personnel: a survey. *Risk Analysis*, 13(4):421–424.

Klibanoff, P., Marinacci, M., and Mukerji, S. (2005). A smooth model of decision making under ambiguity. *Econometrica*, 73:1849–1892.

Klibanoff, P., Marinacci, M., and Mukerji, S. (2009). Recursive smooth ambiguity preferences. *Journal of Economic Theory*, 144(3):930–976.

Knight, F. H. (1921). *Risk, uncertainty and profit*. Hart, Schaffner and Marx, New York.

Kolstad, C. D., Ulen, T. S., and Johnson, G. V. (1990). Ex post liability for harm vs ex ante safety regulation: Substitutes or complements. *The American Economic Review*, 80:888–901.

Kreps, D. M. (1979). Preference for flexibility. *Econometrica*, 47(3):565–577.

Kunreuther, H. (1984). Causes of underinsurance against natural disasters. *Geneva Papers on Risk and Insurance*, pages 206–220.

Kunreuther, H. (1996). Mitigating disaster losses through insurance. *Journal of risk and Uncertainty*, 12(2):171–187.

Kunreuther, H. (1997). Rethinking society's management of catastrophic risks. *Geneva Papers on Risk and Insurance. Issues and Practice*, pages 151–176.

Kunreuther, H. (2006). Disaster mitigation and insurance: Learning from Katrina. *The Annals of the American Academy of Political and Social Science*, 604(1):208–227.

Kunreuther, H., Meyer, R., and Michel-Kerjan, E. (2013). *The Behavioral Foundation of Policy*, chapter Overcoming Decision Biases to Reduce Losses from Natural Catastrophes, pages 398–413. Princeton University Press.

Kydland, F. E. and Prescott, E. C. (1977). Rules rather than discretion: The inconsistency of optimal plans. *Journal of political economy*, 85(3):473–491.

Laffont, J.-J. (1995). Regulation, moral hazard and insurance of environmental risks. *Journal of Public Economics*, pages 319–336.

Lemoine, D. M. and Traeger, C. P. (2012). Tipping points and ambiguity in the economics of climate change. Technical report, National Bureau of Economic Research.

Lewis, T. and Nickerson, D. (1989). Self-insurance against natural disasters. *Journal of Environmental Economics and Management*, pages 209–223.

Lewis-Beck, M. S. and Alford, J. (1980). Can government regulate safety? the coal mine example. *The American Political Science Review*, 74:745–756.

Lin, L. (2013). Enforcement of pollution levies in china. *Journal of Public Economics*, 98:32–43.

Lovering, J. R., Yip, A., and Nordhaus, T. (2016). Historical construction costs of global nuclear power reactors. *Energy Policy*, 91:371–382.

Lévêque, F. (2014). *The Economics and Uncertainties of Nuclear Power*. Cambridge University Press, Cambridge.

Lévêque, F. (2015a). *The Economics and Uncertainties of Nuclear Power*. Cambridge University Press, Cambridge.

Lévêque, F. (2015b). Vers une industrie nucléaire globalisée ? *The Conversation France*.

Macho-Stadler, I. and Pérez-Castrillo, D. (2006). Optimal enforcement policy and firms' emissions and compliance with environmental taxes. *Journal of Environmental Economics and Management*.

Maidl, E. and Buchecker, M. (2014). Raising risk preparedness through flood risk communication. *Natural Hazards and Earth System Sciences Discussions*, 2:167.

Marinacci, M. (2015). Model uncertainty. *Journal of the European Economic Association*, 13(6):1022–1100.

Mason, K. O., Mason, W. M., Winsborough, H. H., and Poole, W. K. (1973). Some methodological issues in cohort analysis of archival data. *American sociological review*, pages 242–258.

Matsuo, Y. (2016). A study on the estimation method of nuclear accident risk cost.

McKean, R. N. (1980). Enforcement costs in environmental and safety regulation. *Policy analysis*, 6:269–289.

Melkonyan, T. A. and Schubert, J. (2009). Food safety regulations under ambiguity. *American Journal of Agriculural Economics*, 91:1389–1396.

Michener, R. and Tighe, C. (1992). A poisson regression model of highway fatalities. *The American Economic Review*, 82(2):452–456. Papers and Proceeedings of the Hundred and Fourth Annual Meeting of the American Economic Association.

Millner, A., Dietz, S., and Heal, G. (2013). Scientific ambiguity and climate policy. *Environmental and Resource Economics*, 55(1):21–46.

MIT (2009). Update of the mit 2003 future of nuclear power. *Cambridge, Mass.: Report for Massachusetts Institute of Technology. Retrieved September*, 17:2009.

Moscarini, G. (2007). Competence implies credibility. *The American Economic Review*, 97(1):37–63.

Moses, L. N. and Savage, I. (1990). Aviation deregulation and safety: Theory and evidence. *Journal of Transport Economics and Policy*, 24(2):171–188.

Mudholkar, G. S. and Srivastava, D. K. (1993). Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*, 42(2):299–302.

Mukerji, S. (1998). Ambiguity aversion and incompleteness of contractual form. *American Economic Review*, pages 1207–1231.

Mukerji, S. (2003). Ambiguity aversion and cost-plus procurement contracts. *University of Oxford Department of Economics, Discussion paper series*, (171).

NEA (2003). Nuclear electricty production: What are the external costs. Technical report, OECD Nuclear Energy Agency.

Neumann, G. R. and Nelson, J. P. (1982). Safety regulation and firm size: Effects of the coal mine health and safety act of 1969. *Journal of Law and Economics*, 25:183–199.

Oestreich, A. M. (2015). Firms' emissions and self-reporting under competitive audit mechanisms. *Environmental Resource and Economics*, 62:949–978.

Ottinger, R. L., Wooley, D. R., Robinson, N. A., Hodas, D. R., and Babb, S. E. (1990). *Environmental Costs of Electricity.* Oceana Publications Inc, Pace University Center for Environmental Studies.

Paté-Cornell, E. (2002). Risk and uncertainty analysis in government safety decisions. *Risk Analysis*, 22(3):633–646.

Paté-Cornell, M. E. (1996). Uncertainties in risk analysis: Six levels of treatment. *Reliability Engineering & System Safety*, 54(2):95–111.

Perko, T. (2011). Importance of risk communication during and after a nuclear accident. *Integrated environmental assessment and management*, 7(3):388–392.

Peters, E. and Slovic, P. (1996). The role of affect and worldviews as orienting dispositions in the perception and acceptance of nuclear power. *Journal of applied social psychology*, 26(16):1427–1453.

Pidgeon, N., Kasperson, R. E., and Slovic, P. (2003). *The social amplification of risk*. Cambridge University Press.

Pljansek, K., Ferrer, M., De Groeve, T., and Clark, I. (2017). Science for disaster risk management 2017: Knowing better and losing less. Technical report, European Commission Joint Research Center - Disaster Risk Management Knowledge Centre.

Portney, P. R. (1992). Trouble in happyville. *Journal of Policy Analysis and Management*, 11(1):131–132.

Prati, G. and Zani, B. (2013). The effect of the fukushima nuclear accident on risk perception, antinuclear behavioral intentions, attitude, trust, environmental beliefs, and values. *Environment and behavior*, 45(6):782–798.

Rabl, A. and Rabl, V. A. (2013). External costs of nuclear: Greater or less than the alternatives? *Energy Policy*, pages 575–584.

Raghavan, S. and Rhoades, D. L. (2005). Revisiting the relationship between profitability and air carrier safety in the us airline industry. *Journal of Air Transport Management*, 11:283–290.

Ramana, M. (2011a). Beyond our imagination: Fukushima and the problem of assessing risk. *Bulletin of the atomic scientists*, 19:37–62.

Ramana, M. (2011b). Nuclear power and the public. *Bulletin of the Atomic Scientists*, 67(4):43–51.

Ramsay, K. W. (2011). Cheap talk diplomacy, voluntary negotiations, and variable bargaining power. *International Studies Quarterly*, 55(4):1003–1023.

Rangel, L. E. and Lévêque, F. (2015). Revisiting the cost escalation curse of nuclear power: New lessons from the french experience. econ. *Economics of Energy and Environmental Policy*, 4(2).

Rangel, L. E. and Lévêque, F. (2014). How Fukushima Dai-ichi core meltdown changed the probability of nuclear accidents ? *Safety Science*, 64:90–98.

Rasmussen, N. (1975). Wash-1400 (reactor safety study). *An assessment of accident risks in US Commercial Nuclear Power Plants*, 8.

Rheinberger, C. M. and Treich, N. (2016). Attitudes toward catastrophe. *Environmental and Resource Economics*, pages 1–28.

Rigotti, L. (1998). Imprecise beliefs in a principal agent model. *Tilburg University, CentER Working Paper*, (1998-128).

Robertson, L. S. (1981). Automobile safety regulations and death reductions in the united states. *American Journal of Public Health*, 71(8):818–822.

Rose, N. L. (1990). Profitability and product quality: Economic determinants of airline safety performance. *Journal of Political Economy*, 98(5):944–964.

Rothman, S. and Lichter, S. R. (1987). Elite ideology and risk perception in nuclear energy policy. *American Political Science Review*, 81(2):383–404.

Rothwell, G. (2002). Does the us subsidize nuclear power insurance. *Stanford Institute for Economic Policy Research - Policy Brief.*

Rothwell, G. (2016). *Economics of Nuclear Power.* Routledge, Taylor & Francis Group.

Salanié, F. and Treich, N. (2009). Regulation in happyville. *The Economic Journal*, 119(537):665–679.

Salop, S. (1976). Information and monopolistic competition. *American Economic Review, Papers and Proceedings*, pages 240–245.

Savage, L. J. (1954). *The Foundations of Statistics.* Dover Publications Inc., New York.

Schmeidler, D. (1989). Subjective probability and expected utility without additivity. *Economectrica*, 57:571–587.

Selvin, S. (2004). *Statistical analysis of epidemiologic data.* Oxford University Press.

Shavell, S. (1984a). Liability for harm versus regulation of safety. *The Journal of Legal Studies*, 13:357–374.

Shavell, S. (1984b). A model of the optimal use of liability and safety regulation. *Rand Journal of Economics*, 15:271–280.

Shimshack, J. P. (2014). The economics of environmental monitoring and enforcement. *Annu. Rev. Resour. Econ.*, 6(1):339–360.

Sjöberg, L. (1998). Worry and risk perception. *Risk analysis*, 18(1):85–93.

Sjöberg, L. (1999). Risk perception by the public and by experts: A dilemma in risk management. *Human Ecology Review*, pages 1–9.

Sjöberg, L. (2002). The allegedly simple structure of experts' risk perception: an urban legend in risk research. *Science, Technology, & Human Values*, 27(4):443–459.

Sjöberg, L. (2004). Explaining individual risk perception: the case of nuclear waste. *Risk Management*, 6(1):51–64.

Sjöberg, L. (2009). Precautionary attitudes and the acceptance of a local nuclear waste repository. *Safety Science*, 47(4):542–546.

Sjöberg, L. and Drottz-Sjöberg, B.-M. (1991). Knowledge and risk perception among nuclear power plant employees. *Risk analysis*, 11(4):607–618.

Sjoberg, L. and Drottz-Sjoberg, B.-M. (2009). Public risk perception of nuclear waste. *International Journal of Risk Assessment and Management*, 11(3-4):248–280.

Slovic, P. (1993). Perceived risk, trust, and democracy. *Risk analysis*, 13(6):675–682.

Slovic, P., Fischhoff, B., and Lichtenstein, S. (1980). Facts and fears: Understanding perceived risk. *Societal risk assessment: How safe is safe enough*, 4:181–214.

Slovic, P., Fischhoff, B., and Lichtenstein, S. (1981). Rating the risks. In *Risk/benefit analysis in water resources planning and management*, pages 193–217. Springer.

Slovic, P., Layman, M., and Flynn, J. H. (1991). Risk perception, trust, and nuclear waste: Lessons from yucca mountain. *Environment: Science and Policy for Sustainable Development*, 33(3):6–30.

Sovacool, B. K. (2008). The costs of failure: A preliminary assessment of major energy accidents,1907-2007. *Energy Policy*, 36(5):1802–1820.

Starr, C. (1969). Social benefit versus technological risk. *Science*, pages 1232–1238.

Stein, J. C. (1989). Cheap talk and the fed: A theory of imprecise policy announcements. *The American Economic Review*, pages 32–42.

Strand, J. (1994). Environmental accidents under moral hazard and limited firm liability. *Environmental and Resource Economics*, 4:495–509.

Suzuki, E. (2012). Time changes, so do people. *Social Science & Medicine*, 75(3):452–456.

Teh, T.-L. (2017). Insurance design in the presence of safety nets. *Journal of Public Economics*, 149:47–58.

Telle, K. (2013). Monitoring and enforcement of environmental regulations: Lessons from a natural field experiment in norway. *Journal of Public Economics*, 99:24–34.

Tengs, T. O., Adams, M. E., Pliskin, J. S., Safran, D. G., Siegel, J. E., Weinstein, M. C., and Graham, J. D. (1995). Five-hundred life-saving interventions and their cost-effectiveness. *Risk analysis*, 15(3):369–390.

The Japan Times (2013). Kan cites god's help in containing nuclear crisis. *The Japan Times*, 13 March 2013, Available: http://www.japantimes.co.jp/news/2013/03/12/national/kan-cites-gods-help-in-containing-nuclear-crisis/.

Tillio, A. d., Kos, N., and Messner, M. (2016). The design of ambiguous mechanisms. *The Review of Economic Studies*, 84(1):237–276.

Tollefson, J. (2016). Nuclear power plants prepare for old age. *Nature*, 537(7618):16–17.

Trager, R. F. (2010). Diplomatic calculus in anarchy: How communication matters. *American Political Science Review*, 104(2):347–368.

Treich, N. (2010). The value of a statistical life under ambiguity aversion. *Journal of Environmental Economics and Management*, 59(1):15–26.

Tubiana, M., Feinendegen, L. E., Yang, C., and Kaminski, J. M. (2009). The linear no-threshold relationship is inconsistent with radiation biologic and experimental data. *Radiology*, 251(1):13–22.

Tukey, J. W. (1977). Exploratory data analysis.

Tversky, A. and Kahneman, D. (1985). The framing of decisions and the psychology of choice. In *Environmental Impact assessment, technology assessment, and risk analysis*, pages 107–129. Springer.

Viscusi, K. W. (1979). The impact of occupational safety and health regulation. *The Bell Journal of Economics*, 10:117–140.

von Hippel, F. N. and Schoeppner, M. (2016). Reducing the danger from fires in spent fuel pools. *Science and Global Security*, 24(3):141–173.

Wahlberg, A. A. and Sjoberg, L. (2000). Risk perception and the media. *Journal of risk research*, 3(1):31–50.

Wheatley, S., Sovacool, B., and Sornette, D. (2017). Of disasters and dragon kings: a statistical analysis of nuclear power incidents and accidents. *Risk analysis*, 37(1):99–115.

Whitfield, S. C., Rosa, E. A., Dan, A., and Dietz, T. (2009). The future of nuclear power: Value orientations and risk perception. *Risk Analysis*, 29(3):425–437.

Wilson, A. J., Dean, M., and Higham, J. P. (2013). A game theoretic approach to multimodal communication. *Behavioral ecology and sociobiology*, 67(9):1399–1415.

Wittman, D. (1977). Prior regulation versus post liability: the choice between input and output monitoring. *The Journal of Legal Studies*, 6:193–211.

Wooldridge, J. (2002a). Econometric analysis of cross section and panel data. *Cambridge, MA: MIT Press.*

Wooldridge, J. M. (2002b). *Econometric analysis of cross section and panel data*. MIT press.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, 50(2):420–445.

Xie, M., Tang, Y., and Goh, T. N. (2002). A modified weibull extension with bathtub-shaped failure rate function. *Reliability Engineering & System Safety*, 76(3):279–285.

Yang, Y. and Land, K. C. (2006). A mixed models approach to the age-period-cohort analysis of repeated cross-section surveys, with an application to data on trends in verbal test scores. *Sociological methodology*, 36(1):75–97.

Zahran, S., Iverson, T., Weiler, S., and Underwood, A. (2014). Evidence that the accuracy of self-reported lead emissions data improved: A puzzle and discussion. *Journal of Risk and Uncertainty*, 49(3):235–257.

Zhang, H., Yan, W., Oba, A., and Zhang, W. (2014). Radiation-driven migration: the case of minamisoma city, fukushima, japan, after the fukushima nuclear accident. *International journal of environmental research and public health*, 11(9):9286–9305.

Résumé

Les quatre chapitres de cette thèse s'attachent à répondre à deux questions de recherche.
Dans un premier temps, je développe des outils théoriques et statistiques visant à mesurer la sûreté nucléaire malgré la rareté des accidents nucléaires majeurs. En particulier, j'applique des résultats de théorie de la décision afin de déterminer le coût social espéré d'un accident nucléaire majeur, en prenant en compte les attitudes individuelles envers les incertitudes qui le caractérisent. Ensuite, j'utilise des données récentes concernant des incidents de sûreté déclarés dans les réacteurs Français afin de mener une analyse statistique de l'évolution de la sûreté nucléaire au cours des 20 dernières années.
Dans la seconde partie de cette thèse, j'aborde la question de l'implémentation de réglementations de la sûreté nucléaire et des politiques post-accidentelles face à des risques rares et catastrophiques. En particulier, j'évalue empiriquement l'effet d'une politique publique française encadrant la surveillance des opérateurs nucléaires par des commissions locales sur le comportement déclaratif des opérateurs et sur leur conformité avec les réglementations existantes. Je propose ensuite une analyse par la théorie des jeux des problèmes de coordination qui existent entre les stratégies de communications de crises et les politiques publiques de prévention et de compensation post-accidentelles.

Mots Clés

Sûreté nucléaire, réglementation, incertitudes, conformité, désastres.

Abstract

The four chapters of this Ph.D. thesis follow two research axes.
First, I develop theoretical and statistical tools for the measurement of nuclear safety, when rare occurrences of accidents preclude the measurement of objective probabilities of incurring harm. In particular, using recent results from decision theory, I develop a framework for the assessment of the expected social cost of major nuclear accidents that accounts for the attitude of individuals towards the uncertainties that characterize their likelihood of occurrence. Next, I provide an empirical analysis of the French nuclear safety based on a novel dataset containing all the significant safety events reported in the currently-operated French reactors. Despite their minor consequences, I show how valuable information regarding safety can be drawn from this data.
In the second part of the thesis, I tackle the question of the implementation of safety regulations and disaster management strategies when risks are rare and catastrophic. I first focus on identifying the causal impact of an information-based incentive mechanism implemented in France on the levels of safety care and compliance exerted by nuclear plant managers. I then develop a cheap-talk model to analyse the coordination of disaster communication strategies with several preparedness and disaster response policies.

Keywords

Nuclear safety, regulation, uncertainty, compliance, disasters.