From Vulnerability Formalization to Finitely Additive Probability Monads

Dissertation

zur Erlangung des Grades
eines Doktors der Naturwissenschaften (Dr. rer. nat.)
am Fachbereich Mathematik & Informatik
der Freien Universität Berlin

vorgelegt von
Sarah Wolf

Berlin, im Juli 2009

Betreuer:
Prof. Dr. Rupert Klein, *Institut für Mathematik, Freie Universität Berlin*

Zweitgutachter:
Prof. Dr. Carlo C. Jaeger, *Potsdam-Institut für Klimafolgenforschung und Wirtschafts- und Sozialwissenschaftliche Fakultät, Universität Potsdam*

Datum der Disputation: 12.02.2010

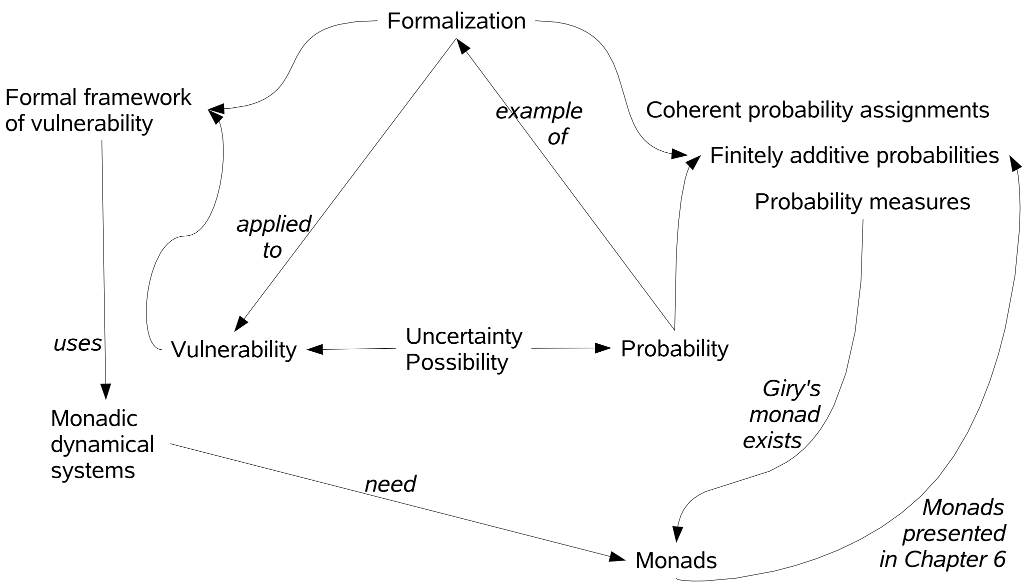
Contents

Chapter 1

Introduction

They say one should begin a dissertation with a sentence of the kind "this work does that" stating the result up front. For the second part (Chapter 6), such a sentence reads: this work presents monads of finitely additive probabilities. It generalizes a result by Giry [1981] which proved the existence of a probability monad, using Kolmogorov's axioms of probability. Finitely additive probability generalizes these by dispensing with the axiom of countable additivity. Much less used in probability theory, finitely additive probability has been studied, but in connection with category it seems not to have been explored so far.

For the first part, which is a work about – rather than in – mathematics, such a sentence would have to be very long. Instead, the following picture provides an impression of how this work links the method formalization, the concepts vulnerability and probability, and mathematical models of these concepts.



A formal framework of vulnerability, developed by the FAVAIA group at the Potsdam Institute for Climate Impact Research within the European ADAM Project, motivates this dissertation. The framework itself is motivated as follows.

In 1992 the United Nations Framework Convention on Climate Change (UNFCCC) set out to prevent dangerous anthropogenic interference with the climate system [UNFCCC, 1992]. In the subsequent scientific efforts to understand how climate change might affect natural and social systems, and to identify and evaluate options to respond to these effects, 'vulnerability' emerged as a central concept. According to the UNFCCC, the needs of

"Parties that are particularly vulnerable to the adverse effects of climate change" should be given full consideration, and they should be assisted in meeting costs of adaptation [UNFCCC, 1992].

Hence, immediate questions in climate change related research are which parties are particularly vulnerable and, first of all, what is meant by 'vulnerable'. The Intergovernmental Panel on Climate Change (IPCC) defines vulnerability as

> the degree to which a system is susceptible to, and unable to cope with adverse effects of climate change, including climate variability and extremes. Vulnerability is a function of the character, magnitude, and rate of climate change and variation to which a system is exposed, its sensitivity, and its adaptive capacity. [Parry et al., 2007]

While prominent in the climate change community, this definition is not the only one; on the contrary, there seem to be almost as many definitions of vulnerability as studies with this topic. The glossary by Thywissen [2006] collects 35 of them. In addition, many other concepts are used in vulnerability related research. The list 'susceptibility, sensitivity, exposure, resilience, adaptive capacity, coping range, adaptation baseline, risk, hazard, ... ' is far from complete and each of these concepts is again defined in various ways.

Studies of vulnerability to climate change are generally interdisciplinary, and researchers from such diverse fields as climate science, development studies, disaster management, health, social science, and economics have no common scientific language to resort to. Therefore, the language used in an assessment is often generated ad hoc and is specific to the case under consideration. It is thus not surprising that the terminology around vulnerability has been compared to the Babylonian confusion. A large amount of conceptual literature, including glossaries and frameworks, has been compiled over the last decades. However, up to now, the much desired common understanding has not been attained, and the search for a coherent, flexible and transparent common language is ongoing.

The formal framework of vulnerability proposes to build this language based on mathematics. It has been developed using the method of formalization, in short, translation into mathematics. A more detailed account of this method and of some benefits it is hoped to yield is given in Chapter 2. The framework itself is presented in Chapter 5 with a focus on the task of concept clarification. It is applied to this task for scientific definitions of vulnerability as well as some conceptual work on it and further to risk assessment in the context of natural hazards.

Different presentation styles that have been used in communicating the framework to different audiences are introduced. The communication aspect is important when working in interdisciplinary groups, and especially the task of communicating mathematics to non-mathematicians was an interesting enterprise accompanying the framework development. This work is certainly influenced by that enterprise. Translating concepts used in a field rooted in the social sciences into mathematical concepts expressed by formulae and then re-translating into diagrams for an audience that prefers these to formulae, the question what exactly is 'mathematics' in such a formerly formal framework arises. Chapter 2 also deals with this question.

In the picture, three routes starting at the formal framework can be taken to introduce the concept probability into the work at this point. The first route is via formalization: probability, a concept used to describe information about uncertainty, can be found as a prime example for illustrating benefits of formalization in the literature. The concept has a history of several hundred years in which different interpretations and mathematical models appear. When the classical calculus of probability became insufficient to function as foundations for probability theory, Kolmogorov's axiomatization of probability was perceived as greatly clarifying matters. Chapter 4 introduces these two mathematical models of probability, and two others, finitely additive probability and coherent probability assignments. Also, the concept itself and some interpretations of it are discussed. The so-called frequentist and subjective interpretations are important here. They show different ranges of applicability of the concept: while the frequentist interpretation presupposes repeatable events, the use of subjective probability, defined as a degree of belief, can be justified for much more general cases. There is no one-to-one correspondence between mathematical models and interpretations of probability. The two purposes of the chapter can be summed up as providing the mathematical basis about probability needed in Chapter 6, and providing some insights on formalization by means of the example probability.

The second route is via vulnerability: uncertainty is inherent in the concept vulnerability, which describes a future possibility of being harmed. Mathematically, uncertainty is often described by means of probability. Concerning the use of probability in vulnerability questions, a subjective interpretation of probability becomes relevant, because the application of the frequentist interpretation may not be warranted in the climate change context. Also, the much older conceptual discussion of probability suggests some points which may be relevant for vulnerability.

The third route is a technical one: the formal framework contains mathematical concepts from category theory to represent descriptions of an uncertain future in a very general fashion. Probability measures defined according to Kolmogorov's axioms are an instance of these concepts, a functor and a monad of probability measures exist. This mathematical model of probability thus can be set to work in the formal framework of vulnerability. The question arises whether this is also true for the more general mathematical models of probability, finitely additive probability and coherent probability assignments. It constitutes the motivation for the results in Chapter 6.

A short remark is due before getting into the matter. Residing on the borders between mathematics, social sciences (for vulnerability) and the humanities (for the rather philosophical questions about concepts and interpretations), three different conventions concerning citations were available. Mathematics, which can generally do without literal quotes, offers the plainest style. In this work, a hybrid between the other two alternatives is chosen. Literal quotes are more frequently present than usual in the vulnerability literature because, as common in the humanities, we deem the original words of authors (or at most translations) important in conceptual questions. The citation style, however, corresponds to the standards in the vulnerability literature, that is, works are cited in brackets in the text, not in footnotes.

Chapter 2

Formalization and mathematics

> *...it demeans mathematics to justify it by appeals to work, to getting and spending. Mathematics is above that – far, far above. Can you recall why you fell in love with mathematics? It was not, I think, because of its usefulness in controlling inventories. Was it not instead because of the delight, the feelings of power and satisfaction it gave; the theorems that inspired awe, or jubilation, or amazement; the wonder and glory of what I think is the human race's supreme intellectual achievement? Mathematics is more important than jobs. It transcends them, it does not need them.*
> *Is mathematics necessary? No. But it* is *sufficient. [Dudley, 1997, p. 364]*

Formalization is one of the topics linking vulnerability and probability in this work. This section introduces the vocabulary needed, without however going into the respective theories of linguistics etc., and the method of formalization, translation into mathematics (Sections 2.1-2.3). Mathematics as a language, and how via the characteristics of this language, formalization can yield benefits, are considered in Section 2.4. We then locate the concepts of interest with respect to formalization: probability, having been variously formalized, can serve as an example concept. Vulnerability, on the other hand, seems to suggest itself for being formalized; the confusion in the terminology is set out in some detail (Section 2.5). We conclude the chapter by considering some difficulties of communication in the language mathematics.

2.1 Concepts and theoretical definitions

In this work, different types of languages are distinguished: *natural language* comprises *ordinary language* used in everyday communication and *scientific language*, used in scientific fields, containing technical terms and specific phrases etc. Languages have as their building blocks *concepts*, that is, words together with their meaning. When a word of a scientific language is considered without the attached meaning it is called a *term* here. An *expression* may be an ordinary language word or a scientific language term. A distinction between concepts and expressions is not always made explicitly, for example in the conceptual literature on vulnerability 'term' and 'concept' are sometimes used interchangeably. The scientific language around a certain concept will also be called its *terminology*.

In ordinary language, the meaning(s) attached to a word can be looked up in a dictionary, however, very many words are intuitively clear to the speakers of a language. It is a characteristic of ordinary language that concepts are not precisely defined.

In scientific language, technical terms have a more specific and more precise meaning than words in ordinary language. Hinkel [2008, see p. 16] outlines the introduction of a scientific language as using some *basic concepts*, or *primitives* and defining more abstract concepts upon them. The primitives are undefined but must be intuitively clear to the users of the language. When all definitions use only primitives and other already defined concepts, every concept can be defined in terms of the primitives alone by substituting in the definition any already defined concept by its definition. This is the case for the introduction of mathematical concepts. Moreover, different definitions of the same concept, such as 'a function', given for example by different authors, can be considered the same definition. Minor differences in the formulation of the definition do not matter. However, scientific concepts are not always introduced in this way. It will be shown (see Sections 2.5.2 and 5.2) that problems in the vulnerability terminology derive among other things from the fact that it is introduced in a less consistent manner.

The same term may be associated with different meanings, for example, in the scientific languages used in different disciplines. This occurs especially for scientific concepts which are based on a concept from ordinary language. The ordinary language word is chosen as a technical term because the meaning of the ordinary language concept conveys a basic idea that the scientific concept is supposed to convey. Then, the meaning is refined, for example, by a definition that attaches a more precise meaning to the technical term, producing a scientific concept. When the same concept is refined in different ways, multiple meanings for the same term are created. Different refinements of the same ordinary language concept are sometimes referred to as different scientific concepts. Considering the underlying meaning from ordinary language important, we prefer here to speak of *one concept*, such as 'the concept vulnerability' and 'the concept probability'. The different refined meanings attached to the same term will be referred to as different *interpretations*[1]. When an ordinary language concept has different aspects, different interpretations may for example arise from the fact that more emphasis is put on a certain aspect in one interpretation and on another one in another interpretation.

In the literature, one also finds the terms conceptualization or definition for interpretations in this sense. The term *definition* is in this work reserved for the statements made to attach a meaning to a term. In particular, *theoretical definitions* state the meaning of a new concept in terms of other concepts.[2] A theoretical definition decomposes the defined concept into the defining concepts. If a scientific language is introduced as Hinkel [2008] describes, more abstract concepts are decomposed into less abstract concepts, which are a step closer to the primitives. The defining concepts are also referred to as *components* of a concept in this work.

The relations between a concept, its interpretations and its definitions are not always clear. It would stand to reason that an interpretation is at an intermediate level of generality between the concept and the definitions, that is, for one concept, there may be several interpretations and for each interpretation there may be several definitions. This, however, need not be the case. For vulnerability, it will be seen that similar definitions are proposed with rather different interpretations. An explanation lies in the emphasis on different aspects of a concept in different interpretations: a theoretical definition tries to capture the concept with all its aspects, and, being a short statement, it generally does not reveal which aspect the emphasis is on in the given interpretation.

2.2 Measurement and operational definitions

Vulnerability and probability are concepts that are being assessed or measured. Studies are undertaken to attach a number to the probability of an event, and a number, several numbers, or some qualitative but comparative value (like the colour of a pixel in a map, where red is 'worse than' green) to the vulnerability of an entity. A case study that uses

[1] The term interpretation is not without controversy. For example, with reference to probability, Hàjek [2007] notes:

> 'Interpreting probability' is a commonly used but misleading name for a worthy enterprise. The so-called 'interpretations of probability' would be better called 'analyses of various concepts of probability', and 'interpreting probability' is the task of providing such analyses. [p. 1].

However, he then follows "common usage" and discusses "different interpretations of probability". Here, the choice is explicitly against considering these as different concepts, but for one concept with different interpretations.

[2] We distinguish theoretical and operational definitions, the latter are introduced later.

measurements to obtain 'the vulnerability of an entity' is called a *vulnerability assessment*. The configuration of methods used in an assessment is referred to as a *methodology* [e.g. Hinkel, 2008].

In the measurement context, we encounter *operational definitions*, which state the meaning of a concept by giving a rule how to measure it. Concerning probability, de Finetti is a vehement advocator of these:

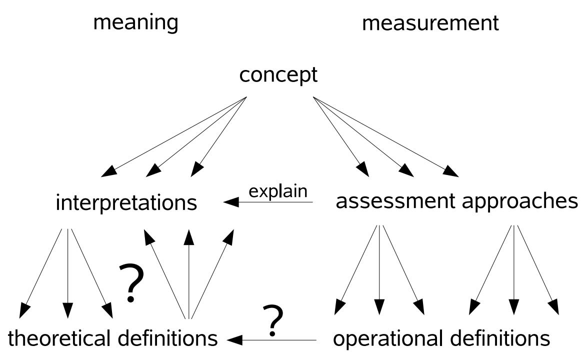
> In order to give an effective meaning to a notion – and not merely an appearance of such in a metaphysical-verbalistic sense – an operational definition is required. By this we mean a definition based on a criterion which allows us to measure it. [de Finetti, 1974, p. 76]

Any vulnerability assessment can be considered an operational definition, because by measuring vulnerability it shows (one way) how to measure it, for the special case under consideration. In most cases, operational definitions are more precise than theoretical ones, however, they are less general. One of the merits of the mathematical definitions in the formal framework presented in Chapter 5 will be that they allow to be general and precise at the same time.

In the vulnerability literature, theoretical definitions are often referred to simply as 'definitions', while the operational definitions used in an assessment are called 'operationalizations' of the definitions. One also finds this distinction referred to as 'definiton' and 'use/usage' of the concept. In this work, when the term definition is used without a qualifying adjective, it refers to a theoretical definition.

For operational definitions of vulnerability, one can observe similarities between some assessments. Assessments have been classified according to such similarities in the literature. A class of similar assessments will be referred to as an *assessment approach*. Here, the obvious nesting in terms of generality is given: the one concept vulnerability is assessed using different approaches and for each approach there are many operational definitions, differing in minor points but sharing a common structure in the methodologies used.

Assessment approaches can be seen as the operational counterpart of interpretations as the diagram illustrates. While the relations between theoretical definitions and interpretations are not necessarily clear, the operational definition does provide information about the assessment approach used. Hence, it generally reveals the interpretation of a concept more clearly than the theoretical definition. De Finetti's statement about operational definitions giving "an effective meaning" is thus confirmed.



We conclude this section with a remark on the measurement of probability and vulnerability in general. Measurements in physics and other experimental sciences are often undertaken to confirm or refute a given theory. Henshaw [2006] gives the example of Newton's laws of motion, which due to experimental confirmation involving measurement have become the generally accepted theory. Given such a theory, indirect measurements can be made. Henshaw's example is the mass of the earth, which cannot be measured directly but needs to be inferred "from the earth's 'performance' (in particular as it relates to the gravitational forces the Earth exerts)." [p. 94]. Henshaw contrasts this with the measurement of intelligence. While the mass of the Earth can be deduced based on Newton's theory, "deducing intelligence from observed performance is more difficult - there is no universally accepted theory of intelligence analogous to Newton's theories of motion." [p. 94]. Vulnerability assessments can be considered similar to intelligence measurement in that they do not follow one generally accepted theory of vulnerability, and cannot, simply because no such theory exists. This and other aspects of the measurement of vulnerability will be considered in Section 5.4.4.

Further, both concepts probability and vulnerability are related to uncertainty: probability describes further information about possibilities, vulnerability describes a possibility of future harm as will be amply discussed in the respective chapters. The concept possibility has implications for measuring probability and vulnerability and validating measurements, deriving from the fact that an ex-post observation does not determine the ex-ante possibility (see, for example, Section 5.1.1).

2.3 The method of formalization

Formalization is the translation of natural language concepts or statements into mathematics, or into a formal language. Ionescu and Botta [2008] discuss different uses of the term formalization. In a strict sense, which is used in logic, formalizing something means setting up a formal system together with an interpretation to that thing. Formal systems consist of so called valid objects and rules for producing other valid objects. Since these rules can be implemented on a computer, and therefore can be checked mechanically, formal systems provide a firm standard of proof. In a weaker sense, formalization means expressing an idea mathematically. Ionescu and Botta refer to this as semi-formalization or mathematization. It is commonly used for example in the natural sciences, where formalization means the translation of ideas from the discipline into mathematics. Even in mathematics itself, one mostly encounters semi-formalization rather than the strict version of formalization. It is considered sufficient that it is clear how one could express the concepts by a formal system, if necessary. Formal proofs are rarely found outside logic in mathematics, because they do not portray the ideas behind the proofs.

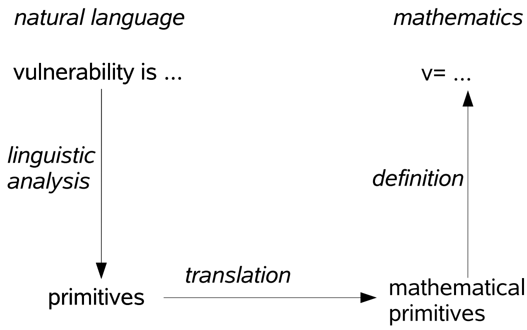
In this work, formalization is used in the latter sense of translating natural language into mathematics. The result of formalizing a concept is a mathematical model of it. We will use 'formalization' to refer to the process or the activity of translating into mathematics and 'mathematical model' (occasionally only 'model') for the result of this process. In the case of the concept probability, the term axiomatization also occurs frequently in the literature. While it also refers to expressing something mathematically, the emphasis is on the rules, the axioms, which some mathematical concepts have to fulfill rather than on the translation of natural language concepts into mathematical concepts.

In an idealized scheme, the activity of formalization can be described by the following three steps:

1. Analysis of the concept: the aim is to identify the primitives that the concept is defined upon and the relations between these in building the concept.

2. Translation of the primitives: the primitives are expressed by mathematical concepts.

3. Definition: the mathematical primitives obtained in Step 2 are used to mathematically define the concept, reproducing the relations between the primitives found in Step 1.



This procedure will be carried out in detail for the formalization of the ordinary language concept vulnerability in Section 5.1. For the scientific concept, the formalization is here presented without going into these steps in detail. When developing the framework, definitions of vulnerability were considered in Step 1 to make use of the decomposition of the concept already provided by these. Theoretical and operational definitions, as well as previous conceptual work were consulted.

2.4 Mathematics and benefits of formalization

The formalization process outlined contains a translation step. Any translation activity requires first of all to gain an understanding. While a rough understanding may do for many purposes, in order to reproduce something in another language, one needs to understand it well. Conversely, trying to reproduce something is generally a good criterion for checking whether one has understood it.

In formalization, the target language in the translation step is mathematics. Rényi [1965] explains how building a mathematical model leads to a better understanding of a situation. In a fictitious "Dialogue on the Applications of Mathematics" he has Archimedes say:

> even a crude mathematical model can help us to understand a practical situation better, because in trying to set up a mathematical model we are forced to think over all logical possibilities, define all notions unambiguously, and to distinguish between important and secondary factors. [p. 10].

This describes a benefit experienced by the person carrying out a formalization. More generally, Polya [2004] notes similar benefits of using mathematical notation:

> Mathematical notation appears as a sort of language, *une langue bien faite*, a language well adapted to its purpose, concise and precise, with rules which, unlike the rules of ordinary grammar, suffer no exception. [p. 135].

Here, 'mathematical notation' takes the place of a language. Mathematics has its own notation, or more correctly many notations, specific to the different branches of mathematics, that can easily be recognized due to the use of formulae. These consist of symbols which are mostly not part of our everyday alphabet.

However, we have described formalization as 'translation into mathematics', not as 'translation into mathematical notation'. In Chapter 5, the formal framework will also be presented in a form that does not contain mathematical notation. Formulae are not what constitutes mathematics, or "the uniqueness of the [mathematical] discourse is not only due to a privileged symbolic system" [Sarukkai, 2001]. An example is a theorem by Gale and Shapley [1962][3] with a proof which does not use any formula, nor even technical terms. Assuming that

> any mathematician will immediately recognize the argument as mathematical, while people without mathematical training will probably find difficulty in following the argument, though not because of unfamiliarity with the subject matter [p. 15],

they arrive at the old question 'what is mathematics?'.

This question has been asked and answered in many and very different ways, from not necessarily serious one-sentence-answers[4] to detailed philosophical analyses. Courant and Robbins [1941], in a book titled 'What is Mathematics', claim:

> For scholars and layman alike it is not philosophy but active experience in mathematics itself that alone can answer the question: What is mathematics? [p. xix]

Nevertheless, many authors have written in non-mathematical but popular scientific fashion on what mathematics is. Two examples are Beutelspacher [1996], starting with a chapter

[3]The article discusses matching in the context of college admission and as an example of one-to-one matching considers marriage.

[4]These can be searched in databases of mathematical quotes on the internet, e.g., the Furman University Mathematical Quotations Server at http://math.furman.edu/~mwoodard/mqs/mquot.shtml

titled precisely with this question, and Rényi [1967] who provides answers to this question in three fictitious dialogues, involving Socrates, Archimedes and Galilei, respectively.

A definition of mathematics seems hard to give, and in fact is often declined from the outset. Sometimes answers to the question what mathematics is not are given instead. Among Beutelspacher's examples are the statements that mathematics is not a natural science, not the art of solving logical puzzles, and not a belief. We should add that 'mathematics is not quantification', 'not just numbers' or as Gale and Shapley [1962] formulate "mathematics need not be concerned with figures, either numerical or geometrical" [p. 15]. While obvious to people doing mathematics, this is not always clear in the vulnerability context.

Like of all concepts discussed in this work, different interpretations of mathematics exist [see, e.g. Sarukkai, 2005]. Here, we take the point of view that mathematics is a *language*. This is of course convenient, given that by formalization we mean translation into mathematics, and one usually translates from one language into another. Moreover, it is not a new interpretation of mathematics. In fact, it goes back to Galilei's famous statement about the book of nature being written in the language of mathematics, "echoed by such figures as Newton, Einstein and Feynman" [Sarukkai, 2005] throughout the history of mathematics. Beutelspacher [1996] describes mathematics as a language for formulating problems, where a good description of the problem entails that the solution can also be formulated in this language. In the ideal case, the problem description even provides hints for the solution. Sarukkai [2005] sums up: "mathematics is a product of human imagination, is grounded in our experience with the world and functions like a language." [p. 417].

A point supporting the interpretation of mathematics as a language is its communicative aspect. Contrarily to the stereotype of the mathematician as a loner sitting over his piece of paper thinking all day, mathematics is a communicative discipline. Many concepts are invented and proofs discovered, to plagiarise Renyi's Socrates[5], by groups of people discussing a problem. Even in the case of the proverbial solitary mathematician compiling a proof all by himself, the validity of the proof can only be checked by having other people understand the proof. This presupposes communication, and explanation, of every detail and every step carried out. In mathematics, things are true only in being explained and can be understood only by being explained[6]. Except for the case of computer generated proofs (which are not accepted as proofs by some mathematicians), the criterion to establish whether a mathematical statement is true is whether other people assert that it is true. Others being able to follow and reproduce every detail is of much bigger importance in mathematics than for example in experimental sciences, where it is not necessary to reproduce a lab experiment to accept its results.

A famous example is Fermat's Last Theorem.[7] As in this case, it may happen that due to the complexity of a proof, very few people are actually able to follow it through and establish its validity. While such people are needed, the criterion is not a subjective one, in other words the verification of a proof does not ask for any subjective judgment. Rényi's Socrates confirms this:

> the mathematicians of different countries can usually agree about the truth, while about questions concerning the state, for example, the Persians and the Spartans have quite opposite views from ours in Athens ... [Rényi, 1964, p. 30].

And, Beutelspacher [1996] points out that mathematics treats its objects of study in an "objectively reproducible way" [p. 4]. Innermathematical questions are not normative ques-

[5] In the Socratic Dialogue, the mathematician is described as an "inventor" of concepts but at the same time as a "discoverer" of theorems [see Rényi, 1964, p. 30].

[6] As Günter Ziegler, at the time president of Deutsche Mathematiker Vereinigung, The German National Mathematical Society, stated in a TV discussion about mathematics in everyday life: "Fernseh-Talk 'Kluge Köpfe' im rbb-Fernsehen: Alles, was zählt! Mathematik im Alltag. Jörg Thadeusz diskutiert mit Vince Ebert (Physiker und Kabarettist), Martin Grötschel (Mathematikprofessor), Klaus Kinkel (Vorsitzender der Telekom-Stiftung), Brigitte Lutz-Westphal (Professorin für Mathematikdidaktik) und Günter M. Ziegler (Mathematikprofessor)". Broadcasted: RBB, 13 Nov 2008 at 22.35 and 15 Nov 2008 at 12.00.

[7] See Singh [1997] for an accessible account of the 358 year story of the proof of Pierre de Fermat's conjecture that no non-zero integer solutions to the equation $x^n = y^n + z^n$ exist for $n > 2$.

tions. These may arise, however, over which mathematical formulation is best suited to describe a real-world phenomenon, and may be fought over vehemently, as will be seen in the chapter on probability.

Objectivity is an important characteristic of mathematics, which Suppes [1968] includes among the benefits of formalization for being valuable in "areas of science where great controversy exists about even the most elementary concepts." [p. 655]. Vulnerability can certainly be seen as such an area of science, as discussed in Section 2.5. Objectitvity can be described with the following sentences by de Finetti:[8]

> statements have *objective* meaning if one can say [...] whether they are either TRUE or FALSE. [.... There is] one condition - do not *cheat*. To *cheat* means to leave in the statement sufficient confusion and vagueness to allow ambiguity, second-thoughts and equivocations in the ascertainment of this being TRUE or FALSE. This instead must always appear simple, neat and definitive. [de Finetti, 1974, p6]

The criteria "neat" and "definitive" as well as the exclusion of "ambiguity, second-thoughts and equivocations" reconnect back to characteristics of mathematics encountered before: unambiguously defined notions and precision occured in the quotes from Renyi and Polya on page 9.

Precision is considered a defining feature of mathematics by Gale and Shapley [1962]. Their answer to the old question is, in fact, "any argument which is carried out with sufficient precision is mathematical." [p. 15]. Precision arises from the way definitions are posed. While often using words from ordinary language, mathematical concepts avoid ambiguity by "restricting the semantic possibilities of some words used" [Sarukkai, 2001, p. 665]. Further, mathematical concepts are always defined before, or at the latest when[9], they are used for the first time. The introduction of mathematical concepts is well described by the process of introducing a scientific language, sketched in Section 2.1. The primitives are often standard mathematical concepts such as numbers and sets. In a system of concepts which are all defined upon (some of) the same primitives, the relations between concepts are clearly defined, as all concepts can be traced back to the primitives. This consistency is described and praised by Rényi's Socrates:

> I think the source of the success of the mathematicians lies in their method: the high standards of their logic, their striving without the least compromise to the full truth; their habit of *starting always from first principles, defining every notion used exactly, and avoiding self-contradictions.* [Rényi, 1964, p. 36, our emphasis].

Let us consider two further descriptions of mathematics from this dialogue. It is stated that mathematics studies non-existing things but finds out the full truth about them. This seeming paradox is then resolved by considering that mathematical objects are precisely what one thinks of them, unlike real-world objects which are never identical with the incomplete and approximate picture one has of them. An example illustrates the situation nicely: Socrates and his interlocutor cannot establish whether a (to them) real-world woman who is accused of murder is guilty, while they have no doubt about the guilt of Clytemnestra, "who is a character figuring in a play and who probably never existed" [p. 28]. Further, mathematics is described as a map of the real world, which aids orientation because it "does not show us every detail, but only the most important things." [p. 36].

In this sense, mathematical concepts are *abstractions*. A simple example is the concept of a set. A set has elements, but from the mathematical point of view it is of no importance what the elements of the set are. Consider in comparison an example from ordinary language: different words are used for sets of different animals. There is a flock of sheep and a flock of

[8]The remark is taken out of context: de Finetti here speaks not about innermathematical statements but about real-world events, for which it must be possible to "say, on the basis of a well-determined observation (which is at least conceptually possible), whether they are either TRUE or FALSE." However, the idea of *not cheating* can be applied equally well to mathematical statements.

[9]Strictly speaking, the definition may occur after the first use of a concept, however, then it immediately follows this use, usually in the form "where ... is defined as ...".

birds, but a murder of crows, a herd of cattle, a school of fish, a swarm of bees and a pack of wolves. The notion of a set incorporates what all these have in common. Of course, in such an abstraction some information is lost, for example, that the pack has a leader in the group. However, the extra information relating to the different animals does not always seem to be universal. Considering in comparison the German words for sets of animals (in the same order: eine Herde Schafe, ein Schwarm Vögel, ein Schwarm Krähen, eine Herde Rinder, ein Schwarm Fische, ein Schwarm Bienen, ein Rudel Wölfe), an interesting difference between the languages arises for birds. In German, the same word is used for fish, insects and birds (the underlying intuition being that of the group's movement) while a different word is used for sheep and cattle. The English 'flock' draws a connection between a group of birds and a group of sheep, in fact, birds are called a flock "especially when on the ground" as an online dictionary informs[10]. The word flock is used especially for geese, whom one sees on the ground as a group more often than smaller singing birds for example. The German word 'Schwarm' illustrates an intermediate degree of abstraction and generality. Much less general than a set, it describes a group of small animals usually seen in a certain kind of movement, abstracting away from the question whether these animals are birds, fish or insects. It is however more general than the three different words for these groups used in Italian, where 'stormo', 'sciame' and 'banco' refer to a group of birds, insects and fish, respectively.

Abstraction makes mathematical concepts universally applicable. Renyi's Socrates describes the benefit obtained by abstracting from "the number of sheep, ships or other existing things" to the numbers themselves:

> what the mathematicians discover to be true for pure numbers is valid for the number of existing things too [...] the mathematician finds that seventeen is a prime number. Does it not follow that you can not distribute seventeen living sheep among some persons so that each should get the same number in any other way than giving seventeen persons one sheep each? [p. 32].

The generality of mathematical concepts allows to reuse these to describe different real-world objects, bringing out commonalities. It is also among the benefits of formalization listed by Suppes [1968] because generality "often provides a means of seeing the forest in spite of the trees" [p. 654]. Mathematics not only uses general concepts, but also looks for general solutions. Euler's solution to puzzle of the "Seven Bridges of Königsberg"[11] is a prime example of the generality that mathematics obtains. The city of Königsberg has four parts, connected by seven bridges, and at Euler's time it was a pastime to try walking through the city crossing every bridge exactly once. Euler proved the impossibility of such a walk by abstracting from the map of Königsberg to a graph, thus laying the foundation of graph theory, and solving the problem for graphs. A walk of this type, now called Euler walk, is possible only when all parts of the city except two, which would be the starting and the end points, have an even number of bridges connecting them to other parts, which is not the case for Königsberg. The theory can now be applied to any other city, and by a simple counting of bridges for each part, one can immediately check whether an Euler walk through any city exists.

Having portrayed mathematics as a language, characterized by precision, abstraction, generality, and objectivity, let us compare it to natural languages. The differences due to precise definitions and notation are arguably not as large or as absolute as often stated. Sarukkai [2001] points out that, while the meaning of a term is restricted by a mathematical definition, the term itself could not be exchanged by other terms arbitrarily: although mathematically defined terms have a fixed meaning and function like names in sentences, they are not "devoid totally of meanings other than those derived through 'naming'." [p. 667]. On the contrary, the meaning known from natural language, with all its ambiguity, is important to get across an idea that underlies the mathematical concept. For example,

[10] http://www.thefreedictionary.com/flock+

[11] An accessible description for non-specialists can be found e.g. in Singh [1997].

"mathematics wants to hold on to the semantic image inspired by the word 'continuous' *as it occurs in non-mathematical talk*." [Sarukkai, 2001, p. 668]. Much as precision is a characteristic of mathematical concepts, an intuitive meaning associated to a concept, often based on its meaning in natural language and hence possibly ambiguous, is of considerable importance in doing mathematics.

The emphasis on intuition in mathematics is present also in some mathematical textbooks. De Finetti [1974], for example, prefers to make everything "simple, intuitive and informal" [p. 3]. He explicitly wants to avoid being "limited to lifeless manipulations in technical terms, or to heavy indigestible technical language."[p. 5]. Good [1977], in a review of de Finetti's Theory of Probability[12], appreciates:

> He also likes to give intuitive reasons for theorems, their "wherefore" (II, p. 218), which he rightly regards as more important than their formal proofs. It is a pity that this opinion is not universally taken for granted.

Even for the symbolic notation, Sarukkai [2001] argues a similarity between mathematics and natural language by saying that mathematical symbols "first and foremost refer to words in natural language" [p. 667]. This means the symbols function as abbreviations. The reader immediately translates these back into natural language, adapting if necessary the grammatical form. For example, the symbol $\in$ is translated as 'belongs to' or 'belonging to' depending on the grammatical construction of the sentence in which it is used.

However, the symbolic notation system does constitute an important difference between natural language and mathematics. Operations can be performed on the symbols, which cannot be done with natural language words. In some contexts the symbolic notation has a power of expression that natural language does not reach. A description of complex physical phenomena shows this power: for example, fluid dynamics would be very hard, if not impossible, to describe verbally to some detail, while mathematics provides means to not only describe but further to study the dynamics.[13]

The symbolic notation is very concise, slowing down communication. Consider the number of pages read in an hour of a mathematical and a non-mathematical text in comparison. To communicate mathematical results successfully (and not just give an overview) means that the listener or reader understands the argument. Understanding the operations carried out with the symbols in the symbolic notation in most cases means reproducing them, or at least following when seeing them produced.

Even in face to face communication of mathematics, writing has an important role. Sarukkai [2001] even states that "mathematics is writing" [p. 669].[14] In the age of slide presentations and beamers, mathematicians still use the old blackboard and chalk in order to explain their mathematics to others. A slide presentation with a slide full of formula manipulation would not decrease the time needed to go through each step, and one may just as well write the formulae in real time. Finding a balance between intuition and the precision of formula manipulation is a very important issue in communicating mathematics. As discussed in Section 5.6 such balancing acts occured in presenting the formal framework to non-mathematical audiences.

A further benefit of formalization, immediately linked to the possibility of performing operations on symbols in formulae, arises in the computational context. All computer programs "presuppose a series of models which act as a bridge between the dull precision of the computer and the real world, ever changing and hard to define." [Todesco, 2009, p. 163]. To formalize means to provide such bridges. Many vulnerability assessments make use of a computer to run models, evaluate measurements, aggregate information and the like. A general mathematical formulation of concepts, as provided by formalization, can be reused in different assessments and filled in with the specifics of the assessment under consideration.

[12]Both volumes, de Finetti [1974, 1975].

[13]A book illustrating an expressive power of mathematics and images, that natural language does not have, is Markowich [2007].

[14]Euler, who was still a very productive mathematician after having gone blind, is certainly an exception.

At some point in such an assessment, a translation of concepts into a language that the computer can work with has to occur, and a mathematical model resulting from formalization is a step closer to a language the computer understands than natural language.

Benefits of formalization have been discussed in the literature in more general fashion. A prominent work is "The Desirability of Formalization in Science" by Suppes [1968]. The author considers formalization as a tool to "clarify conceptual problems and to make explicit the foundational assumptions of each scientific discipline."[p. 653]. We will consider some further benefits listed by Suppes in relation to vulnerability in Section 2.5.4. Let us therefore turn to the concepts considered, probability and vulnerability.

2.5 Formalization of probability and vulnerability

Probability has been variously formalized over the last few hundred years. In mathematics, it is nevertheless considered a young discipline because the mathematical foundations for probability theory that are in use today were laid only at the beginning of the twentieth century. In many mathematical disciplines, mathematical models are studied rather detached from what they represent. In probability theory, ordinary language concepts like probability, likelihood and independence play an important role. The focus on how the mathematics connects to what it represents entails an interest in the topic of formalization itself. This is found for example in probability textbooks, like in Loeve's introduction to his "Probability Theory":

> Probability theory is concerned with the mathematical analysis of the intuitive notion of "chance" or "randomness", which, like all notions, is born of experience. The quantitative idea of randomness first took form at the gaming tables, and probability theory began, with Pascal and Fermat (1654), as a theory of games of chance. ...
> A theory becomes mathematical when it sets up a mathematical model of the phenomena with which it is concerned, that is, when, to describe the phenomena, it uses a collection of well-defined symbols and operations on the symbols. As the number of phenomena, together with their known properties, increases, the mathematical model evolves from the early crude notions upon which our intuition was built in the direction of higher generality and abstractness.
> ...the inner consistency of the model of random phenomena became doubtful, and this forced a rebuilding of the whole structure in the second quarter of this century, starting with a formulation in terms of axioms and definitions. Thus, there appeared a branch of pure mathematics - probability theory - concerned with the construction and investigation *per se* of the mathematical model of randomness. [Loeve, 1977, p. 1]

Concepts such as randomness, chance, or probability itself, used in relation with uncertainty in everyday situations, are not easy to grasp. This may be a reason for the interest in the connections between these concepts and their mathematical models. Suppes [1968] considers the formalization of a family of concepts "one way of bringing out their meaning in an explicit fashion" and names probability as "a good example of what can be hoped for in this direction" [p. 654]. His further statements that

> the formalization did not end discussion and philosophical analysis of the concept of probability. Rather, it helped to raise the discussion to a new level. [Suppes, 1968, p. 654].

will be discussed in Chapter 4.

In the context of vulnerability, formalization is a rather new idea. The formal framework of vulnerability to climate change (presented and discussed in Chapter 5) contains two mathematical models of vulnerability, a basic model of the ordinary language concept, and a refined version for the scientific concept. The remainder of this section introduces the vulnerability terminology with its problems, and the state of the art in conceptual work to define the scope of this formalization and identify possible benefits of formalization in this particular case.

2.5.1 Origins of 'vulnerability to climate change'

Occuring in the UN Framework Convention on Climate Change, the concept vulnerability has been of importance from the beginning of the internationally institutionalized consideration of climate change. The choice of a technical term that is familiar from ordinary language suggests that the scientific concept is based on the working understanding people have of 'vulnerability'. There is a "long history of vulnerability assessments developed in other contexts, such as food security, livelihoods, natural disasters, and risk management in general" [Füssel and Klein, 2006, p. 302]. Cutter [1996] lists for example global change and environment and development studies, geography, political ecology, and human ecology as areas in which vulnerability is studied.

In the climate change context, the introduction of the concept vulnerability is usually described along the following lines: climate change related research and policy efforts have been divided between the two goals of *mitigation*, that is, the reduction of greenhouse gas emissions to avoid climate change, and *adaptation* to the "remaining climate change". While mitigation is targeted at the global climate system, adaptation actions are mostly of a local or regional scope and deal with the smaller scale effects of global climate change. In a first moment, mitigation questions were the main focus of climate change related research. Füssel and Klein [2006] provide reasons for the traditionally greater attention towards mitigation in climate change research and policy, for example, the "polluter pays principle", the relative ease of monitoring greenhouse gas emissions and the global scope of mitigation actions. Impact assessments were being conducted to set mitigation targets based on the investigation of consequences of (unmitigated) climate change.

Then, a shift from impact assessment to vulnerability assessment is observed (for example in the focus of the IPCC's second working group) and is commonly attributed to the recognition that climate change is already taking place, which increases the importantance of adaptation. This implies a changing purpose in the assessments made in climate change research, from informing mitigation decisions at the international level to aiding decision-makers at national and local levels in adaptation questions. In fact, Füssel and Klein [2006] who distinguish four prototypical assessment stages, note a development from *impact assessments* to *vulnerability assessments*, and later to *adaptation policy assessments*, which serve different purposes. The general development of assessments is one from a narrow focus on estimating potential impacts of climate change to considering climate change in a broader context. Including economic, political and social factors, assessments become more and more interdisciplinary. Concepts used in other research fields, such as food security, disaster risk reduction, poverty and livelihoods are introduced into climate vulnerability research. That is, with the shift from impact to vulnerability assessments, disciplines that have been using the concept vulnerability before gain influence in climate change research.

The interdisciplinarity of studies of vulnerability to climate change implies that in most cases researchers from the different disciplines do not have a common technical language to resort to. Therefore, the language used in an assessment is often generated ad hoc and is specific to the case under consideration.[15] (Re-)defininining concepts anew for each case study implies an enormous production of concepts and definitions. In fact, the abundance of concepts, and of definitions for each concept, in the vulnerability terminology is one of its problems.

2.5.2 (Too) many definitions

There seem to be almost as many theoretical definitions of 'vulnerability to climate change' as studies assessing it. These definitions generally decompose the concept vulnerability

[15]See e.g. Hinkel [2008] on transdisciplinary assessments in the climate change vulnerability context.

into several other concepts, for which again there are many definitions each. Brooks [2003] observes a "sometimes bewildering array of terms" [p. 2].

While in Section 2.1 the introduction of a scientific language was described as a procedure of defining all concepts upon the intuitively clear primitives and other already defined concepts only, in definitions of vulnerability and related concepts, this procedure is generally not followed. Vulnerability assessments state theoretical definitions for the most important concepts used, decomposing them into other concepts. Most articles presenting assessment results contain a section on definitions of the concepts used. However, the defining concepts are not necessarily themselves defined or intuitively clear. Many concepts from ordinary language are used in definitions without having restricted their meaning so far as to make them unambiguous. Thus, the definitions presented generally leave quite a bit of room for interpretation.

Similarly, definition efforts of more general scope do not base all definitions on primitives and previously defined concepts. An example is the IPCC glossary of terms proposed with the Third Assessment Report[16]: an analysis of the concept vulnerability and its defining concepts in this glossary is presented by Hinkel [2008]. The most basic concepts found are not intuitively clear, as for example "ability to adjust", "statistical reference distribution", "ability to cope" and "significant climate variations". What is the difference between the ability to cope and that to adjust, one might ask, or when is climate variation to be judged 'significant'?

Theoretical definitions decompose vulnerability to different levels of detail, and in general the level of detail increases from theory to practice with the purpose of the definition. Definitions which may be considered somewhere between theoretical and operational are given by guidelines to vulnerability assessment [see, e.g. Buckle et al., 2001; Twigg, 2004]. These practically oriented texts sometimes explicitly avoid theoretical definitions, follow "common usage" [e.g., Twigg, 2004] or keep definitions "broad" [e.g., Buckle et al., 2001] and as general as possible. For example:

> The following definitions are offered as a starting point, to clarify the discussion below. Please avoid getting concerned about the particular definitions used, and instead concentrate on their general meaning. [Kuban and MacKenzie-Carey, 2001].

Then, these guidelines present detailed lists of factors contributing to vulnerability which should be taken into account in assessments. These (theoretical) generalized rules for assessment make a step towards operational definitions without however giving precise rules how to practically measure vulnerability.

Another hybrid between the theoretical and the operational level, which also defines vulnerability, consists of 'frameworks for vulnerability assessment'. These propose a theoretical foundation for the practical task of assessing vulnerability. Often, they also provide an intermediate level of detail, as for example the "framework for vulnerability analysis in sustainability science" of Turner II et al. [2003] that decomposes the concept vulnerability into "broad classes of components and linkages that comprise a coupled system's vulnerability to hazards." [p. 8076].

Finally, there are operational definitions, given for example by any vulnerability assessment, as explained in Section 2.2. Each of these definitions is specific to the case under consideration. It is common to find different operational definitions accompanied by the same theoretical definitions of concepts, such as the frequently cited definitions from the IPCC glossary. The prominent IPCC definition of vulnerability quoted in the Introduction is an example that has been operationalized in many ways. This confirms that the definitions leave room for interpretation: each assessment uses it own "reading" of the definition [see also Ionescu, 2009, who provides an analysis of studies that use the IPCC definition].

[16] It can be found under http://www.ipcc.ch/glossary/index.htm. For the Fourth Assessment Report, there are three glossaries by the different working groups. Each intended for use with the corresponding chapter, they differ in which terms are defined and do not necessarily propose the same definitions for the same terms.

Thus, the terminology is characterized by a multitude of concepts, each equipped with a host of definitions of different kinds, from the general but imprecise theoretical to the precise but case-specific operational ones. Some of these definitions may be inconsistent with one another, and in any case definitions are not easily compared. To provide a random example, what is the relation between "vulnerability, the degree to which a system is susceptible to a hazard" and "vulnerability, the exposure of a system to a hazard, combined with its capacity to react"? Is there an essential difference? What exactly is meant by susceptibility? Can it be interpreted as a combination of exposure and capacity? Are the definitions not so different after all? Often, the concern that "one definition does not fit all cases" is expressed, or the question whether *one single definition of vulnerability* is possible, and whether it is needed or should be avoided is posed.

Answers shall be provided with the help of formalization in Chapter 5. In the spirit of de Finetti's statement that "mathematics is interesting as an instrument for condensing concepts, which by their own right, without mathematics would still be what they are" [de Finetti, 2008, p. 1], the framework will provide general but precise mathematical definitions of vulnerability and related concepts, "condensed" out of the many definitions given in the literature.

2.5.3 State of the art in work about the concept

The just described situation in the vulnerability terminology has been referred to as 'Babylonian confusion' and has been the motivation of much conceptual work from a meta-perspective. Thus, in addition to conducting vulnerability studies, researchers have studied the concepts used in these studies [see Hinkel, 2008].

As early as 1981, Timmerman starts "A Review of Models and Possible Climatic Applications" from the premise that "it is hard to say just what 'vulnerability' and 'resilience' are". More than 20 years later, the literature still asserts problems: "...the fact remains that the word 'vulnerability' means different things to different researchers" [O'Brien et al., 2007]. The relations between the many concepts used are not clear, and for example Gallopín [2006] warns: "If care is not used, the field of human dimensions research can become epistemologically very messy." [p. 301].

Lists of concepts related to vulnerability become longer: for example, Füssel and Klein [2006] add "exposure, sensitivity, coping capacity, and criticality" to the list they quote from Liverman [1990] that already contains "concepts such as resilience, marginality, susceptibility, adaptability, fragility, and risk". Likewise, the lists of previous conceptual work lengthen in each new work. Füssel [2007], who claims to "provide the much-needed conceptual clarity", refers the reader to eleven works for "general reviews of the conceptualization of vulnerability" and mentions eight further publications "focussing on the conceptualization of vulnerability in climate change research".

A complete list of conceptual vulnerability literature would probably be be so long as to be no longer useful. Here, we concentrate on some examples of what conceptual work has to offer: Füssel and Klein [2006] analyse the evolution of conceptual thinking in climate change vulnerability assessments; O'Brien et al. [2007] discuss "how two interpretations of vulnerability in the climate change literature are manifestations of different discourses and framings of the climate change problem" and provide a "diagnostic tool for distinguishing the two interpretations" [p. 73]; Adger [2006] "reviews research traditions of vulnerability to environmental change" [p. 268]; and Gallopín [2006] uses "a systemic perspective to identify and analyze the conceptual relations among vulnerability, resilience, and adaptive capacity within socio-ecological systems" [p. 293]. Glossaries of terms have been assembled [e.g. Thywissen, 2006], and different frameworks[17] related to vulnerability describe assessment approaches or propose how vulnerability should be assessed.

The conceptual work has not shown the effect of clarification in the terminology, but

[17] Again, lists could be very long. As a selection, let us cite Cutter [1996]; Jones [2001]; Brooks [2003]; Turner II et al. [2003]; Luers [2005].

hope has not been given up. Still, in the 2006 special issue of Global Environmental Change on "Resilience, vulnerability and adaptation" the editors "experienced a Tower of Babel in hearing the diverse definitions made of core concepts" [Janssen and Ostrom, 2006, p. 237] but state that "efforts should be made to develop clear (and hopefully, mutually compatible) specifications of the concepts for use in abstract and field studies of ecological and social systems." [p. 238]

Introducing one more meta-level, it has been discussed why the conceptual work did not succeed. Hinkel [2008] states that "the meta-perspective activities themselves suffer from conceptual difficulties" [p. 13], for example, in that *meta-concepts* for speaking about different definitions are themselves left undefined. An example here could be Füssel [2007], who uses 'interpretation of vulnerability', 'different vulnerability concepts', 'terminologies and classifications of vulnerability' and 'conceptualization of vulnerability' in the range of 5 sentences [p. 159] without having made explicit how these relate to each other. The difference between interpretations, concepts and conceptualizations, if any, is unclear.

In particular, most conceptual work does not distinguish between theoretical and operational definitions, nor between interpretations and assessment approaches. Interpretations can often most clearly be observed from assessment approaches but theoretical and operational definitions are not as closely related. One important source of the confusion, the gap between these, is left out of consideration. Glossaries, which consider only theoretical definitions, lack the perspective on the operational side of the concept.

Finally, formulations in conceptual work are not always clear. At least to a mathematically minded, it seems that, for consistency, in the statement

> The word 'vulnerability' is used in all three discourses; thus it is often difficult to 'place' a study within a discourse and framing based on the presence or absence of this word. [O'Brien et al., 2007, p. 78],

the words 'often difficult' should be replaced with 'impossible'.

2.5.4 Why formalize vulnerability

Having seen some general benefits of formalization and the situation in the vulnerability terminology, characterized by an over-abundance of definitions of all kinds, the use of formalization is easy to motivate. Brooks [2003] expresses the aim that researchers from different backgrounds

> must develop a common language so that vulnerability and adaptation research can move forward in a way that integrates [the] different traditions in a *coherent yet flexible* fashion, allowing researchers to assess vulnerability and the potential for adaptation in a wide variety of different contexts, and in a manner that is *transparent* to their colleagues. [p. 2, our emphasis].

An argument why mathematics can produce transparency is explicitness. Not only does the translation step require to state assumptions clearly. Also when already given, the mathematical formulations, unlike their natural language counterparts, do not carry implicit connotations which differ according to the disciplinary background of the users. The desired coherence can be provided by mathematics due to the consistent introduction of precise concepts and the absence of exceptions (see the quote by Polya on page 9). Flexibility is supplied by generality: general mathematical concepts can be filled with different content according to the situations in vulnerability assessments. Janssen and Ostrom [2006] add "clear (and hopefully, mutually compatible) specifications of the concepts" [p. 238] to the list of desiderata. Clarity and compatibility also follow from consistently introducing precise definitions. Recalling the characteristics of mathematics described in Section 2.4, these two quotes almost seem to call for a mathematical definition of concepts.

The multitude of definitions for the same concepts suggests that a necessary first step is understanding what these have in common, proverbially speaking, see the forest in spite of the trees. This is the goal of the analysis step: uncover the common structure of the many similar definitions by identifying the primitives. According to Beutelspacher, mathematics is an attempt at discovering logical structures.[18] Taking this literally, even the analysis step in the formalization process, when applied to many definitions of the same concept, can be considered mathematics. Formalizing several concepts from one domain, some primitives recur and some concepts might be the primitives of other concepts. Tracing back all concepts of interest to one set of primitives renders obvious the relations between the concepts, uncovers the structures between them.

Moreover, having mathematized a given definition of a concept, analysing further definitions becomes easier by comparing whether the same primitives and relations between them can be found. Thus, formalization helps understand the meaning of concepts for which confusingly many similar definitions are found. The general but yet precise mathematical definition serves as the foundation upon which different definitions can be compared, for example, by expressing them as special instances of it. Formalization can therefore be expected to yield benefits in the confused terminology around vulnerability.

The question whether one single definition is needed will be answered "yes and no" for definitions at different levels of detail. We consider an agreed basic understanding a necessary foundation for unambiguous communication about the details of each case [see also Hinkel and Klein, 2007]. Without such a common starting point, successful communication seems rather impossible. The benefit to be expected from formalization lies in the very general mathematical definition for the ordinary language concept vulnerability (in Section 5.1): it can serve as this starting point.[19]

One further benefit discussed by Suppes [1968] is not immediately obvious for the formalization of vulnerability. Standardization, that should "make communication easier across different scientific disciplines" [p. 654] would be an important aim for vulnerability research because of the many disciplines involved. While this may work between natural sciences, communication in mathematical language cannot be taken for granted in the context of vulnerability studies. Difficulties relating to the communication of mathematics are considered in more detail in the following section. However, via a retranslation of the mathematical framework into a graphical framework as presented in Section 5.5, the structures found in the analysis can still be presented in a clearer fashion than encountered in much previous conceptual work. Independent of the use of formulae, the framework remains an abstract and general description of concepts, obtained by a mathematical way of thinking, and aiming at enhancing clarity.

2.6 Communication difficulty

The non-formal representation of the formal framework in Chapter 5 serves a simple purpose: making the framework readable for non-mathematicians, and increasing its acceptance by researchers from the vulnerability community who are often social scientists, not trained to use mathematical notation. A very first step sometimes needed is to clear up the misunderstanding that "translation into mathematics means quantification". Having experienced a certain hostility towards a framework presented in terms of formulae, a less and less mathematical representation of the framework has been developed, and is discussed in Section 5.6 of the chapter on vulnerability.

Problems of communicating mathematics to non-mathematicians are well known and both mathematicians and non-mathematicians have written about them. For example, Enzensberger [1999] describes mathematics as a cultural anathema and investigates reasons for the tacit consensus in the public's attitude toward mathematics, where hatred and "no talent" are stated "with a remarkable blend of defiance and pride" [p. 9]. Mathematics has been described as the ivory tower science par excellence, as for example in this summary by Sarukkai [2003], who later however challenges this view.

[18]The German original says "Mathematik ist der Versuch, logische Strukturen zu entdecken. [Beutelspacher, 1996, p. 5].

[19]The answer "no" will be given in Section 5.2.

> In creating its image, mathematics has turned out to be very successful. By projecting itself as an extremely hard discipline, mathematics has built a fortress around it, inaccessible to intrusions from unfriendly people and disciplines. The process of seclusion is strengthened by the contention that mathematics is a formal, axiomatic, deductive and logical system, which is concerned only with a platonic world and whose connection with our real world is only incidental and perhaps even mysterious. It has also been consistently claimed that mathematics is universal, acultural and ahistorical. Further, this image serves to guard mathematics against the perceived vagaries of verbal languages. [p. 3648].

Teaching is in part held responsible for the rejection of mathematics by many people. Enzensberger [1999] complains that rather than mathematics only arithmetics is taught in the first years of school. An instruction based on "mind-numbing routines" and a "cookbook of formulas", not only has nothing to do with mathematical thinking, it "ultimately fosters mathematical illiteracy." [see p. 35,37].

There are by now a large number of efforts to change the public image of mathematics portrayed above and to improve the teaching of mathematics, such as a "Year of Mathematics"[20] in 2008 in Germany; 11 years of conferences "Mathematics and Culture"[21] in Venice, Italy; projects to improve mathematics teaching like the Germany wide "SINUS-Transfer"[22]; as well as many popular scientific publications playing with mathematics[23], just to mention a selection.

The literature has also seen many refutations of an overly formal description of mathematics, for example by Courant and Robbins [1941] who plainly state that if mathematics really was merely a system of deductions as described above, "mathematics could not attract any intelligent person" [p. xvii]. That there is a connection with the real world in mathematics is pointed out via a comparison with poetry by Renyi's Socrates: mathematics is not at the arbitrary choice of the mathematician because in that case "there would be as many mathematics as mathematicians", while "mathematicians living far from each other and having no contact discover the same truths independently. I never heard of two poets who wrote the same poem." [Rényi, 1964, p. 30].

The fact that, as described above, communication of mathematics is only successful when the reader/listener understands the mathematics presented, constitutes a difficulty for its acceptance. Independently of the notation, understanding mathematics generally asks a high level of concentration and patience. Mathematics teaches humility because hardly ever are all problems solved by the end of the day.[24] Gale and Shapley [1962], who expect the non-mathematical reader to have difficulty in following their formula-free proof, state (somewhat unfriendly) that

> the reason that your friends and ours cannot understand mathematics is not because they have no head for figures, but because they are unable to achieve the degree of concentration required to follow a moderately involved sequence of inferences. [p. 15].

Enzensberger attributes "a certain isolation", the "*déformation professionelle*" of the mathematician, for one part to the fact that the mathematician's "occupation demands above all intense and sustained concentration" [p. 15].

A superficial reading of mathematics does not do. The fact that some people may not have the time or be willing to engage into a deeper reading creates a barrier to communicating mathematics independently of the notational system chosen. In fact, concerning the formal framework of vulnerability presented later, avoiding formulae is rather a psychological choice, to not scare people off from the beginning. The diagrams used present a

[20]http://www.jahr-der-mathematik.de/

[21]http://www.mat.uniroma1.it/venezia2009/

[22]http://sinus-transfer.uni-bayreuth.de/

[23]A particularly pleasant book is the analysis of "La Mathematique du Chat" by Justens [2008], investigating what the Belgian comic strip cat by Philippe Geluck has to say about mathematics.

[24]As Martin Grötschel said in the previously mentioned TV discussion on mathematics, see Footnote 6.

simplified version of the framework, which contains only mathematical concepts at the level of high school mathematics. Thus, those who engage into reading and understanding the diagram presentation would most probably also understand the corresponding presentation in formulae (see Section 5.6).

2.7 Conclusions

This chapter has introduced the method of formalization as translation into mathematics. Mathematics itself is here considered as a language, which is more precise, general and abstract then natural languages are. Its specific notation has the advantage that it allows to carry out operations on symbols. Results can be deduced and computers may be used.

The current situation in the terminology of vulnerability and conceptual work on it was outlined. The main finding is confusion: a multitude of different kinds of definitions of many concepts can be found. The relations between different concepts and different definitions are unclear. In short, general definitions in the vulnerability terminology, that is, theoretical definitions, remain vague or ambiguous (as shall be further seen in Section 5.2). Precise definitions, that is, operational definitions, are case specific. Since mathematical definitions are general and precise at the same time, some clarity can be expected from a translation into mathematics in this situation. Further, such a translation forces one to make assumptions explicit and the mathematical definitions produced do not carry implicit connotations differing with the users' backgrounds. The stage has thus been set for formalizing vulnerability in Chapter 5.

It was also pointed out that mathematical notation, but also the precision of mathematics, can be (perceived as) a barrier to communication. The formalization of vulnerability will include balancing acts to raise accessibility of the formal framework for the non-mathematically trained.

Probability, having a history of formalizations, can be an example of what formalization achieves and was seen to be even considered a paradigmatic example. What it tells about formalization will be seen in the next but one chapter, because first some mathematical preliminaries have to be provided.

Chapter 3

Mathematical preliminaries

This chapter establishes notation for mathematical concepts used and states some results needed at different places throughout this work. Selected basics from measure theory and category theory are dedicated a section each. Parts of this chapter can be considered a work of reference, for example, Section 3.1 can be skipped without losing content in reading this work. The measure theory section also outlines the problem of the extension of measures which causes a main mathematical difference between probability measures and finitely additive probabilities in the following chapter. The category theory section, apart from basics, provides the definition and some discussion of the monadic dynamical systems intoduced by Ionescu [2009] for the formal framework of vulnerability.

3.1 Sets, functions, algebras ...

Sets will be denoted with capital letters, A, B etc., in the probability context often Ω, Ω', sets of states also S. For sets of sets, mostly calligraphic letters are used. For a non-empty set Ω, its powerset, that is the set of all its subsets, is denoted $\mathfrak{P}(\Omega)$. If Ω is finite, so is its powerset: with $|\Omega| = n$, where $|.|$ denotes the cardinality of a set, one has $|\mathfrak{P}(\Omega)| = 2^n$. $\mathbb{R}$ denotes the set of real numbers, $\bar{\mathbb{R}}$ the so called extended real numbers, $\mathbb{R} \cup \{-\infty, \infty\}$.

Functions will be denoted by small letters or, in the vulnerability formalization by words starting with small letters. The notation $f : A \longrightarrow B$ means that the function f is defined on A and takes values in B, where both A and B are sets. When f is applied to an element a of A, it outputs an element of B, denoted ... differently in the two branches of mathematics brought together in this work, probability and category theory, or rather in category theory and the rest of mathematics. The choice is between $f(a)$, commonly used in mathematics, and $f\ a$, more usual in category theory. The choice in this work tries a compromise between legibility issues. In the chapter on probability, we stick to "the good old notation" used in probability theory. For functions on sets and probabilities, that is functions on sets of subsets of a set, the classical notation with brackets will be used. For the category theoretical concept of a functor (and some others), brackets will be omitted to avoid too many of them in a row. These notations are explained at first use. Function composition is denoted $\circ$, the dot $\cdot$ denotes multiplication.

In category theory, the set A is referred to as the *source*, B as the *target* of f. Together, the information $f : A \longrightarrow B$ is called the *type of* f. We also refer to A as the domain of definition of a function, as is common use in mathematics.

Dealing with probability and measures, sets of sets with the following structures are going to be used.

Definition 3.1.1. $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$ is an *algebra (of subsets of* Ω*)* or *algebra over* Ω, if it contains

Ω and is closed under finite intersection and complementation. In formulae,

$$\Omega \in \mathcal{A}$$
$$A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$$
$$A, B \in \mathcal{A} \Rightarrow A \cap B \in \mathcal{A}$$

It follows by induction that the intersection of finitely many elements of the algebra is again one of its elements, however, this is generally not true for infinite intersections. From $A \cup B = (A^c \cap B^c)^c$, one concludes that an algebra contains also finite unions of its elements.

Similarly, a σ-algebra is defined requiring closure under countable intersections:

Definition 3.1.2. $\mathcal{F} \subseteq \mathfrak{P}(\Omega)$ is a *σ-algebra (of subsets of Ω)* or *σ-algebra over* Ω, if

$$\Omega \in \mathcal{F}$$
$$A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$$
$$A_1, A_2, \ldots \in \mathcal{F} \Rightarrow \bigcap_{i=1}^{\infty} A_i \in \mathcal{F}$$

Any σ-algebra is an algebra but the converse is not true. The powerset $\mathfrak{P}(\Omega)$ is a σ-algebra over Ω, and hence also an algebra.

Given a non-empty set Ω, a *partition* is a set of subsets $\{A_1, A_2, \ldots\}$ of Ω with two properties: any two sets A_i and A_j with $i \neq j$ are disjoint, that is $A_i \cap A_j = \emptyset$, and their union $\bigcup_{i=1}^{\infty} A_i = \Omega$ is the whole set. For further reference:

Definition 3.1.3. Given a non-empty set Ω and an algebra $\mathcal{A}$ on Ω, a finite *partition of* Ω *in* $\mathcal{A}$ is a partition of Ω into finitely many sets, say $\{A_1, A_2, \ldots, A_n\}$, where all sets of the partition belong to $\mathcal{A}$, that is, $A_i \in \mathcal{A}$ for all $i = 1, \ldots, n$ (and analogously for σ-algebras).

For a set of subsets $\mathcal{E} \subseteq \mathfrak{P}(\Omega)$ which is not itself an algebra, an important concept is the algebra generated by $\mathcal{E}$.

Definition 3.1.4. Given $\mathcal{E} \subseteq \mathfrak{P}(\Omega)$, *the algebra generated by* $\mathcal{E}$, denoted $\mathfrak{a}(\mathcal{E})$, is the smallest algebra over Ω which contains $\mathcal{E}$.

The intersection of (arbitrarily many) algebras is an algebra, wherefore one can intersect all algebras which contain $\mathcal{E}$ to obtain $\mathfrak{a}(\mathcal{E})$. At least one algebra which contains $\mathcal{E}$ exists for any $\mathcal{E} \subseteq \mathfrak{P}(\Omega)$, because the powerset itself is an algebra. Hence $\mathfrak{a}(\mathcal{E})$ exists for any $\mathcal{E}$.

Any set in $\mathfrak{a}(\mathcal{E})$ can be constructed from sets in $\mathcal{E}$ using three simple steps. Define

$$\mathcal{F}_1 = \{A \subseteq \Omega | A = \emptyset \text{ or } A = \Omega \text{ or } A \in \mathcal{E} \text{ or } A^c \in \mathcal{E}\}$$
$$\mathcal{F}_2 = \{\cap_{j=1}^n A_j | n \geq 1, A_1, \ldots A_n \in \mathcal{F}_1\}$$
$$\mathcal{F}_3 = \{\cup_{j=1}^n B_j | n \geq 1, B_1, \ldots B_n \in \mathcal{F}_2 \text{ disjoint}\},$$

then $\mathcal{F}_3 = \mathfrak{a}(\mathcal{E})$ [see e.g. Bogachev, 2006, p. 5].

On the interval $[0, 1]$, the algebra generated by the intervals (a, b) with $a, b \in [0, 1]$ and $a \leq b$ will be of importance in Chapter 6. It is denoted[1] by $\mathcal{J}[0, 1]$. It contains all open, closed and half-open intervals and points: open intervals by definition, closed intervals via complements, half-open intervals as intersections of open and closed ones, points simply by taking $a = b$. One can also use the following set of intervals as generator for this algebra:

$$\{[0, a], [0, a) | a \in [0, 1]\}.$$

This generates all open intervals: an open interval $[a, b]$ can be obtained as $[0, b) \backslash [0, a]$.

A concept that will be important in Section 4.5.4 is the set of constituents of a finite $\mathcal{E}$.

[1]Its importance in the construction of the so-called Jordan-measure suggests the initial to be used.

Definition 3.1.5. Given $\mathcal{E} = \{E_1, \ldots, E_n\}$, the *constituents* of $\mathcal{E}$ are all sets of the form $E_1^* \cap \ldots \cap E_n^* \neq \emptyset$ with $E^* \in \{E, E^c\}$. The set of constituents is denoted $\mathcal{C}$, the constituents $C_1, \ldots, C_m$.

The constituents partition Ω and each $E_i \in \mathcal{E}$ can be written as a disjoint union of constituents. There are $m \leq 2^n$ constituents.

Remark 3.1.6. The set of constituents $\mathcal{C}$ of $\mathcal{E}$ generates the same algebra as $\mathcal{E}$ does.

Similar to the generated algebra, one can consider the σ-algebra generated by a set of subsets $\mathcal{E} \subseteq \mathfrak{P}(\Omega)$.

Definition 3.1.7. Given $\mathcal{E} \subset \mathfrak{P}(\Omega)$, *the σ-algebra generated by* $\mathcal{E}$, denoted $\sigma(\mathcal{E})$, is the smallest σ-algebra over Ω which contains $\mathcal{E}$.

It is the intersection of all σ-algebras over Ω which contain $\mathcal{E}$. In general, it is not possible to construct its elements in countably many steps from the elements of $\mathcal{E}$ [see e.g. Billingsley, 1979, p. 26].

The commonly used σ-algebra on $\mathbb{R}$ is the *Borel-σ-algebra*, denoted $\mathcal{B}(\mathbb{R})$ here, which is the σ-algebra generated by the open subsets of $\mathbb{R}$. It can be shown that this σ-algebra is strictly smaller than the powerset of the real numbers, that is, non-Borel-measurable sets exist, though "every subset of $\mathbb{R}$ which you meet in everyday use" [Williams, 1991, p. 17] is a Borel set. All intervals are contained in the Borel-σ-algebra, and it can also be shown that the set of intervals $\{(-\infty, x] | x \in \mathbb{R}\}$ generates the Borel-σ-algebra on $\mathbb{R}$. The Borel-σ-algbra on $\bar{\mathbb{R}}$ is denoted by $\bar{\mathcal{B}}$, that on the interval $[0,1]$ by $\mathcal{B}[0,1]$.

Given sets that are equipped with algebras, and functions that have sets as source and target, it will be of interest what functions can do with the algebras. Here, the following concepts are needed. Consider a function $f : \Omega \longrightarrow \Omega'$ and sets $A \subseteq \Omega$, $A' \subseteq \Omega'$. The *image* of A under f is the set

$$f(A) = \{f(\omega) | \omega \in A\} \subseteq \Omega'.$$

The *inverse image* of A' under f is the set of all points whose image is in A':

$$f^{-1}(A') = \{\omega \in \Omega | f(\omega) \in A'\} \subseteq \Omega.$$

One can consider the inverse image as a mapping between the powersets, $f^{-1} : \mathfrak{P}(\Omega') \longrightarrow \mathfrak{P}(\Omega)$. It exists for any function f and is not to be confused with an inverse function. With a slight (but in measure theory commonly used) abuse of notation, we denote the image of a set of sets $\mathcal{B} \subseteq \mathfrak{P}(\Omega')$ under the inverse image mapping by

$$f^{-1}(\mathcal{B}) = \{f^{-1}(B) | B \in \mathcal{B}\}.$$

The inverse image has the very important property that it preserves the set operations union, intersection and complementation.

$$f^{-1}\Big(\bigcup_{\iota \in I} B_\iota\Big) = \bigcup_{\iota \in I} f^{-1}(B_\iota)$$

for an arbitrary index set I and $B_\iota \subseteq \Omega'$ and similarly for intersections and complements. These properties are meant when the inverse image is referred to as "well-behaved". Due to these properties, the inverse image maps algebras on Ω' to algebras on Ω, and moreover one has the following

Lemma 3.1.8. *Given $f : \Omega \longrightarrow \Omega'$ and an algebra $\mathcal{A} = \mathfrak{a}(\mathcal{E})$ on Ω', the inverse image of $\mathcal{E}$ generates the algebra $f^{-1}(\mathcal{A})$ on Ω, that is*

$$\mathfrak{a}(f^{-1}(\mathcal{E})) = f^{-1}(\mathfrak{a}(\mathcal{E})).$$

Remark 3.1.9. In proofs in Chapter 6, this lemma will often be used to simplify the task of showing that the inverse image of an algebra $\mathcal{B}$ on Ω' is contained in a given algebra $\mathcal{A}$ on Ω. If $\mathcal{B} = \mathfrak{a}(\mathcal{E})$, then for showing that $f^{-1}(B) \in \mathcal{A}$ for any $B \in \mathcal{B}$, it suffices to show that this is the case for the sets in the generator $\mathcal{E}$. When $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{E}$, also their intersections and complements are in $\mathcal{A}$ because this is an algebra. Therefore, $\mathfrak{a}(f^{-1}(\mathcal{E})) \subseteq \mathcal{A}$. But this means that $f^{-1}(\mathcal{B}) = f^{-1}(\mathfrak{a}(\mathcal{E})) = \mathfrak{a}(f^{-1}(\mathcal{E})) \subseteq \mathcal{A}$.

The analogous lemma for the generated σ-algebra is also true. The lemma can be used to equip a set Ω with a (σ-)algebra when a function $f : \Omega \longrightarrow \Omega'$ and a (σ-)algebra on Ω' is given. One uses the inverse images of all sets in that (σ-)algebra. This is also the idea underlying the *initial σ-algebra* on a set which is not equipped with a σ-algebra a priori, which will be needed in Chapter 6.

Definition 3.1.10. Consider given a set Ω, an index set I, sets with algebras $(\Omega_\iota, \mathcal{F}_\iota)$ and functions $f_\iota : \Omega \longrightarrow \Omega_\iota$ for $\iota \in I$. The *initial σ-algebra* on Ω with respect to the family $(f_\iota)_\iota$ is the σ-algebra

$$\mathcal{F} = \sigma\left(\bigcup_{\iota \in I} f_\iota^{-1}(\mathcal{F}_\iota)\right)$$

Anticipating Definition 3.2.3 of a measurable function from the next section, the initial σ-algebra is the smallest σ-algebra $\mathcal{F}$ on Ω that makes all $f_\iota : (\Omega, \mathcal{F}) \longrightarrow (\Omega_\iota, \mathcal{F}_\iota)$ measurable. If the $\sigma-$algebras $\mathcal{F}_\iota$ have generators $\mathcal{E}_\iota$ respectively, then

$$\mathcal{E} = \bigcup_{\iota \in I} f_\iota^{-1}(\mathcal{E}_\iota)$$

generates $\mathcal{F}$ [see Elstrodt, 1999, p. 112].

3.2 Measure theory

The mathematical concept of a measure generalizes such everyday measures as length, area, volume etc. We will here just sketch the results needed later, details and proofs can be found in textbooks on measure theory. We additionally mention Elstrodt [1999], who provides a historical account of the problems encountered, and Ulam [1943], who nicely summarizes the developments in measure theory in an almost formula-free description.

The idea behind the definition of a measure is described by Ulam [1943] as follows:

> There is given a collection (class) of sets situated in the Euclidean space. This class contains the elementary figures and is large enough to include all sets that can be obtained by the processes usually employed in analysis; in particular the complement of a set that is in the class also belongs to the class, and the union of any denumerable number of sets in the class yields sets belonging to this class (the denumerable union of sets corresponds to the process of summing infinite series). One has to attach to every set in the class a non-negative, real number, called its measure, so that the following postulates will be satisfied:
>
> (*i*) All sets of a specific subclass should have measure, for example, all sets consisting of a single point.
>
> (*ii*) The measure of a set should coincide with its ordinary value in the case when the set is an elementary figure.
>
> (*iii*) The postulate of additivity: two forms, a weaker and a stronger form are possible. This requires that the measure of a finite (or in a stronger form, denumerably infinite) sum of mutually disjoint sets should be equal to the numerical sum of the measures of the individual sets.

(*iv*) The invariance or congruence postulate: This requires that sets congruent in the sense of elementary geometry, should have equal measure. [p. 598]

We will here use the term "countable" instead of "denumerable", later, "σ-additivity" also refers to countable additivity. The "sum" of sets referred to in axiom (*iii*) is their union. The finite and the countable additivity condition will be referred to as (*iiia*) and (*iiib*), respectively.

Lebesgue constructed a measure for $\mathbb{R}^n$, now carrying his name, that satisfies conditions (*i*) where each point gets measure zero,(*ii*) in that to a "cube" in $\mathbb{R}^n$ it associates the product of the lengths of its sides, (*iiib*), and (*iv*). Lebesgue measure is denoted as Λ later. The class of subsets for which this measure is defined is referred to as the set of *Lebesgue measureable sets*. It contains the important σ-algebra of the Borel sets, but, as was shown first by Vitali in 1905, it is impossible to define this measure on *all* subsets of $\mathbb{R}^n$ [for a proof see e.g. Elstrodt, 1999, p. 96].

Several generalizations of this non-existence result, but also some existence results for "measures" on the powerset of the real numbers exist. A fundamental difference is seen between requiring finite and countable additivity, that is, between (*iiia*) and (*iiib*). A complete overview over these results seems not easy to get, because measure theory textbooks often state the non-existence result only for $\mathbb{R}^n$ with the congruence postulate. For example, the results given by Elstrodt [1999] are [adapted, see pp. 4-6]:

- "Das Inhaltsproblem": problem statement: find a function $m : \mathfrak{P}(\mathbb{R}^n) \longrightarrow [0, \infty]$ that satisfies conditions (*iiia*), (*iv*), and a norming condition corresponding to (*ii*).
- This problem cannot be solved for $\mathbb{R}^n$ if $n \geq 3$ (Theorem by Hausdorff, 1914).
- This problem has a (non-unique) solution for $n = 1$ and $n = 2$ (Theorem by Banach, 1923).
- The so called Banach-Tarski paradox "dramatically" illustrates that the problem cannot be solved for $n \geq 3$: given any two bounded sets with non-empty interior in $\mathbb{R}^n$ with $n \geq 3$, it is possible to "cut the first set up" into finitely many subsets, move these subsets and reassemble them into the other set.
- The seeming absurdity of this result is remarked by quoting Galilei's statement that when a body is divided into finitely many pieces, it is doubtlessly impossible to reassemble these in such a way that they should occupy more space than before.
- An explanation for the counterintuitive result by Banach and Tarski is that the pieces into which the first set is cut can be very complicated. They are constructed by means of the axiom of choice.
- "Das Maßproblem": The corresponding problem statement with condition (*iiib*) instead of (*iiia*).
- Vitali's result: this problem cannot be solved.
- The corresponding Banach-Tarski paradox: in this case one may have to cut the first set into countably many pieces.
- The conclusion drawn is that it does not make sense to require measures to be defined on the whole powerset of $\mathbb{R}^n$.
- A remark is added that the restriction to $\mathbb{R}^n$ has turned out unneccessary. Using a general set is useful for such branches of mathematics as functional analysis and probability theory.

Then, measure theory is developed for general sets instead of $\mathbb{R}^n$, and supposing the stronger additivity condition (*iiib*) implying the use of a σ-algebra as the domain of definition of a measure. Algebras and finitely additive functions are simply not further considered, being discarded as "not rich enough" [Elstrodt, 1999].[2]

Generally, *measure*, *measure space* and a *measurable space* are defined along the lines of the following definition that will be used in this work.

Definition 3.2.1. Let $\mathcal{F}$ be a σ-algebra over Ω. A function $m : \mathcal{F} \to [0, \infty]$ is a *measure on* $(\Omega, \mathcal{F})$ if it satisfies

(*i*) $m(\emptyset) = 0$

(*ii*) $m(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} m(A_i)$ for disjoint sets $A_i \in \mathcal{F}$.

In this case $(\Omega, \mathcal{F}, m)$ is called a *measure space*. The pair $(\Omega, \mathcal{F})$ alone is also referred to as a *measurable space*. A measure is said to be *normed*, when it takes values in a bounded interval $[0, a] \subset [0, \infty]$.

A case of special importance in probability theory is that of $a = 1$, as will be seen in Section 4.3. When a fixed measure space is considered, sets which are assigned the value 0 by the measure m are called null sets and are often considered to be of no importance. The expressions 'almost everywhere' and 'almost surely' are used to say 'except for a set of measure zero'. For example, apart from the usual notions of convergence, such as pointwise or uniform convergence, in measure theory *almost sure convergence* is of interest, meaning pointwise convergence except on a null set.

Returning to the more general setting, there are further results scattered in the literature. For example, the conditions (*ii*) and (*iv*) have been dropped from the problem statements above. Banach and Kuratowski [1929] show the non-existence of a measure on the whole powerset, where the set considered is the interval $(0, 1)$, the function is supposed to assign positive measure to some set, but measure zero to any single point, and further conditions are used. Banach [1930] generalizes this (again under further conditions) to arbitrary sets of the same cardinality and measures that can take negative values. Ulam [1930] weakens the further conditions, to give some examples.

The version of Vitali's result stated by de Finetti [1972] is probably best adapted for this work, in terms of generality. Translated to our terminology, it states that

Theorem 3.2.2. *A countably additive measure* $m : \mathcal{E} \longrightarrow [0, 1]$ *(that is, normed to 1) does not exist when* $\mathcal{E}$ *is the powerset of an infinite set, unless it is concentrated in a finite number or a countably infinite number of points in* Ω*, with values summing to* 1.

Finitely additive functions defined on algebras can be found in the literature under the name of *charges*, see for example Bhaskara Rao and Bhaskara Rao [1983], who, however, do not discuss the problem of extending charges to the powerset of the underlying set. The term *content* is also to be found for these functions. In this work, a result on the extendability of finitely additive functions will be stated in the context of de Finetti's mathematization of probability in Section 4.5.

Having defined measures only on certain σ-algebras of subsets of a given set, a restriction is introduced for the functions that are taken into consideration in measure theory. Functions from a measurable space to another need to be compatible with the σ-algebras in the following sense:

Definition 3.2.3. A function between the two measurable spaces $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$, that is $f : \Omega \longrightarrow \Omega'$, is *measurable*, or more precisely $\mathcal{F}$-$\mathcal{F}'$-measurable, if

$$f^{-1}(A') \in \mathcal{F} \text{ for all } A' \in \mathcal{F}'.$$

[2] "Für den Aufbau einer fruchtbaren Maßtheorie erweisen sich Ringe und Algebren als nicht reichhaltig genug, da sie nur bez. der Bildung endlicher Vereinigungen abgeschlossen sind." [p. 13].

We denote measurable functions $f : \Omega \longrightarrow \Omega'$ or $f : (\Omega, \mathcal{F}) \longrightarrow (\Omega', \mathcal{F}')$ depending on whether the emphasis is on the function itself or on its measurability.

Measurable functions can be used to construct a measure on the target measurable space from a given measure on the source measurable space:

Definition 3.2.4. Given a measure space $(\Omega, \mathcal{F}, m)$ together with a measurable function $f : (\Omega, \mathcal{F}) \longrightarrow (\Omega', \mathcal{F}')$ to a measurable space, the *induced measure* m^f on $(\Omega', \mathcal{F}')$ is defined by

$$m^f(A') = m(f^{-1}(A')) \text{ for all } A' \in \mathcal{F}'.$$

This construction will be important for the monads in Chapter 6.

Measurable functions from an $(\Omega, \mathcal{F})$ to the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ are of special importance, for example, with regard to integration. We list some properties:

Lemma 3.2.5. *Let $f_1, f_2, \ldots$ be $(\mathcal{F}$-$\mathcal{B}(\mathbb{R}))$-measurable functions from Ω to $\mathbb{R}$.*

(*i*) *$f_1 + f_2$ is a measurable function.*

(*ii*) *For any $\alpha \in \mathbb{R}$, the function αf_1 is measurable.*

(*iii*) *If $h : \mathbb{R} \to \mathbb{R}$ is a continuous function, then $h \circ f_1$ is measurable. In particular, $|f_1|$ is measurable.*

(*iv*) *$\sup_{n\in\mathbb{N}} f_n$ and $\inf_{n\in\mathbb{N}} f_n$ are measurable functions.*

(*v*) *If $f_n \to f$ (pointwise convergence) for a function $f : \Omega \longrightarrow \mathbb{R}$, then f is measurable.*

An important class of measurable functions are the so called simple functions. To denote these, we need the notation for the *characteristic* or *indicator function* of a set $A \subseteq \Omega$.

$$\mathbb{1}_A : \Omega \longrightarrow \{0, 1\}$$
$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \in A^c. \end{cases}$$

Definition 3.2.6. Given $(\Omega, \mathcal{F})$, a function $f : \Omega \longrightarrow \mathbb{R}$ is called *simple* if it can be written in the form $f = \sum_{i=1}^{n} c_i \mathbb{1}_{A_i}$ for some $n \in \mathbb{N}$, $c_1, \ldots, c_n \in \mathbb{R}$ and a partition $\{A_1, \ldots, A_n\}$ of Ω in $\mathcal{F}$.

Their importance derives from the fact that

Lemma 3.2.7. *Any non-negative measurable function $f : \Omega \longrightarrow \mathbb{R}$ is a monotone pointwise limit of simple functions.*

This fact is used in Lebesgue's integral construction, which, roughly speaking, works as follows: The integral of a simple function (in the above notation) with respect to a measure m is defined as

$$\int f dm = \sum_{i=1}^{n} c_i m(A_i). \tag{3.1}$$

First, non-negative measurable functions are considered. For such a function f, the integral is defined as the limit of the integrals for a sequence of simple functions converging to f. It can be shown that these integrals of simple functions converge to a value which does, moreover, not depend on the particular sequence chosen, so that this defines the integral of a non-negative measurable function unambiguously as a real number or the value ∞. For general measurable functions, the positive and negative parts are integrated separately: $f = f^+ - f^-$ with $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$, so that $\int f dm = \int f^+ dm - \int f^- dm$. If both terms in this last sum are infinite, the function is said to be *not integrable*. The *Lebesgue integral* is the integral obtained from this construction when the measure used is the Lebesgue measure.

Two convergence results are used in many proofs in measure theory:

Theorem 3.2.8 (MON, monotone convergence theorem, or Theorem of B. Levi). *For a monotonically increasing sequence of non-negative measurable functions* $f_n : (\Omega, \mathcal{F}) \longrightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$*:*

$$\int \left(\lim_{n\to\infty} f_n \right) dm = \lim_{n\to\infty} \int f_n dm$$

Theorem 3.2.9 (DOM, dominated convergence theorem, or Lebesgue's convergence theorem). *Let* $f_n : (\Omega, \mathcal{F}) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ *be measurable functions, which converge almost surely to the function* f*. Let* $g : (\Omega, \mathcal{F}) \longrightarrow (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ *be an integrable function with finite integral, and let* $|f_n| \leq g$ *almost surely (under Lebesgue measure), for all* n*.*
Then, f *is integrable with finite integral value and*

$$\int f dm = \int \left(\lim_{n\to\infty} f_n \right) dm = \lim_{n\to\infty} \int f_n dm$$

The procedure of *measure theoretic induction* uses either of these two. Often a result which involves showing the equality of two integrals is easy to show for simple functions. It is then extended to measurable functions by writing the measurable function as the limit of a sequence of simple functions and then switching limit and integral using MON or DOM.

3.3 Category theory

Category theory provides very general descriptions of mathematical concepts. The formal framework of vulnerability presented in Chapter 5 makes use of this generality. This section introduces the concepts needed. The first definitions given here are taken or adapted from Bird and de Moor [1997, see pages 25-35], who say that "one does not so much learn category theory as absorb it over a period of time" [p. 25].

Definition 3.3.1. A *category*, denoted $\mathbb{C}$, is an algebraic structure consisting of a class of *objects* and a class of *arrows* (also called morphisms) with three total operations (source, target and identity) and a partial operation (composition) which satisfy the rules below. Objects are denoted A, B and so on, arrows f, g and so on.

The operations *source* and *target* each assign an object to an arrow: denoted $f : A \longrightarrow B$, and read "f is an arrow from A to B", here A is the source of f, B is its target. The notation $Hom_{\mathbb{C}}(A, B)$ refers to the set of all arrows from A to B in the category $\mathbb{C}$. The third total operation takes an object A to an arrow $id_A : A \longrightarrow A$, the *identity (arrow) on* A.

The partial operation, composition, takes two arrows $f : A \longrightarrow B$ and $g : B \longrightarrow C$ to an arrow $g \circ f : A \longrightarrow C$, read "$g$ after f". Composition

- is associative, that is, for $f : A \longrightarrow B$, $g : B \longrightarrow C$ and $h : C \longrightarrow D$ one has $(h \circ g) \circ f = h \circ (g \circ f)$, and
- has the identity arrows as units: given $f : A \longrightarrow B$, one has $id_B \circ f = f = f \circ id_A$.

The category $\mathbb{F}un$ of sets and functions This is the most well-known example: *Objects* in the category $\mathbb{F}un$ are sets. We denote these by Ω, Ω' and so on as usual in the probability context. An *arrow* $f : \Omega \longrightarrow \Omega'$ is simply a function. Its source is Ω, its target Ω'. Each object has an *identity arrow*, the identity function on the set, that is $id_\Omega : \Omega \longrightarrow \Omega$ defined by $id_\Omega(\omega) = \omega$. *Composition* is the usual composition of functions. Take $f : \Omega \longrightarrow \Omega'$ and $g : \Omega' \longrightarrow \Omega''$, then $g \circ f : \Omega \to \Omega''$ is given by $(g \circ f)(\omega) = g(f(\omega))$. Since composition is associative and it has the identity arrows as units, $\mathbb{F}un$ is a category.

Further, the category $\mathbb{M}es$ of measurable spaces and measurable functions is used in this work. It is introduced in Section 6.3.

Definition 3.3.2. A category $\mathbb{C}$ is a *subcategory* of another category $\mathbb{D}$ if all objects of $\mathbb{C}$ are also objects of $\mathbb{D}$ and similarly all arrows of $\mathbb{C}$ are also arrows of $\mathbb{D}$. We call $\mathbb{C}$ a *full subcategory* of $\mathbb{D}$, if given two objects A and B in $\mathbb{C}$, all arrows between them in $\mathbb{D}$ are already arrows in $\mathbb{C}$.

The formal framework uses the concept of a functor to describe and work with sets equipped with a structure containing further information (such as probabilities). These structures will serve to express uncertainty: all elements of a set are considered possible, the structure gives further information about the possibilities, for example their probabilities. "Work with" more specifically means apply functions to such sets in a way that maintains the extra information structure. Given some structure over a set, one wants to apply a function to the elements of the set, so that the structure is preserved and the application of the function occurs "inside" it, as in the case of an induced probability measure, for example.

Definition 3.3.3. A *functor* is a mapping from a category $\mathbb{C}$ to a category $\mathbb{D}$, consisting of a mapping for objects and one for arrows. Both are denoted by the same F. As rather common for category theory, we will not use brackets to denote functor application.[3] For example, $F\,A$ is the object of $\mathbb{D}$ obtained by applying the functor to the object A of $\mathbb{C}$. A functor has to satisfy $F\,f : F\,A \longrightarrow F\,B$ whenever $f : A \longrightarrow B$. Further, it needs to preserve identities and composition, in formulae

$F\,id_A = id_{F\,A}$ for every object A of $\mathbb{C}$, and

$F\,(g \circ f) = F\,g \circ F\,f$ for arrows $f : A \longrightarrow B$ and $g : B \longrightarrow C$ of the category $\mathbb{C}$.

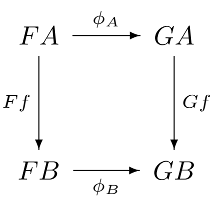
A functor from a category to itself is called an *endofunctor*. The identity is a functor on every category, denoted $Id_{\mathbb{C}}$, or Id only, if the category is clear. The composition of two functors $F : \mathbb{C} \longrightarrow \mathbb{D}$ and $G : \mathbb{D} \longrightarrow \mathbb{E}$ is a functor itself, denoted $G \circ F$.

Example 3.3.4. *An example which is important for this work is the* powerset functor*, denoted* $\mathfrak{P}$ *here. It associates to a set the set of all its subsets and lifts a function* $f : A \longrightarrow B$*, so that* $\mathfrak{P}f : \mathfrak{P}A \longrightarrow \mathfrak{P}B$*, by the definition:* $\mathfrak{P}fA = \{f(a)|a \in A\}$.

The importance of this functor derives from the fact that sets can be used to represent uncertainty. In the so called *non-deterministic* description of uncertainty, a set of possibilities is given without any further information[4].

There are further mappings between functors which we need: transformations, and in particular, natural transformations which are in some sense compatible with the functors.

Definition 3.3.5. Given two functors between the same categories, $F : \mathbb{C} \longrightarrow \mathbb{D}$ and $G : \mathbb{C} \longrightarrow \mathbb{D}$, a transformation from F to G is a collection of arrows $\phi_A : FA \longrightarrow GA$, one for each object A of the two functors' source category $\mathbb{C}$. The transformation ϕ is *natural* if for every arrow $f : A \longrightarrow B$ of $\mathbb{C}$, one has $Gf \circ \phi_A = \phi_B \circ Ff$, as expressed by commutativity of the diagram:



Again, the identity is a natural transformation from any functor F to itself, which we denote id_F. Also, the composition of two natural transformations $\phi : F \longrightarrow G$ and $\eta : G \longrightarrow H$ is a natural transformation itself, denoted $\eta \circ \phi$. A natural transformation $\phi : F \longrightarrow G$ can

[3] Strictly speaking, this causes a clash of notation with the common notation of multiplication, where no symbol is used. While it should be clear from the context, whether functor application or a multiplication of numbers in meant, whenever there might be a doubt, multiplication will therefore be denoted with the explicit $\cdot$.

[4] See Section 4.5 for such a description of uncertainty, which preceedes the assignment of probabilities.

be applied to compositions of functors: for a functor H of the appropriate type, we denote by $\phi_H : F \circ H \longrightarrow G \circ H$ the corresponding transformation. Vice versa, a functor can be used to "lift" a transformation by applying it to each arrow in the transformation. Denoted $H\,\phi$, this yields a transformation from $H \circ F$ to $H \circ G$.

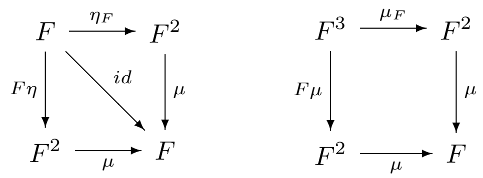
Next, we need certain functors with natural transformations which allow to get back and forth between several "levels" of functor application, F, F^2 etc., that is, F is composed with itself.

Definition 3.3.6. A *monad* on a category $\mathbb{C}$ is a triple (F, η, μ) where F is an endofunctor on $\mathbb{C}$, and the natural transformations $\eta : Id \longrightarrow F$ and $\mu : F \circ F \longrightarrow F$ satisfy the following conditions:

$$\mu \circ \eta_F = id_F = \mu \circ F\eta$$

$$\mu \circ \mu_F = \mu \circ F\mu$$

These conditions can again be expressed by commutative diagrams:



Monads are interesting for the formal framework of vulnerability because they can be used to produce dynamical systems which combine the possibility of iteration known from classical dynamical systems with the possibility of representing uncertainty given by functors. Ionescu [2009] introduces a *monadic dynamical system* to describe in one general model different types of systems, such as non-deterministic and probabilistic systems. The details can be found in Chapter 5 of that work. Here, a sketch of the idea, an adapted definition and a brief description of trajectories of the system must suffice. The monadic dynamical system is the motivation for our monads in Chapter 6, however, for the presentation of the formal framework as a tool for literature analysis, the rather involved technical work of Ionescu is not so relevant. Hence, restating it would not make sense here.

Consider the simplest example of a dynamical system: given a set S called the set of states (or state space) and an endofunction on it, $f : S \longrightarrow S$ called the transition function, one can take a starting state $s \in S$ and compute the trajectory of the system starting in that state by repeated application of f. The system evolution is given by this *trajectory*, $s, f(s), f(f(s)), \ldots$.

Now, to represent uncertainty, one would like the transition function to output not a single next state but, for example, a set of possible next states or a probability distribution over possible next states. The general description would be given by a functor. The transition function would be of type $S \longrightarrow FS$. However, it is not obvious how to iterate such a function. A possibility would be to apply the lifted function to the result of a first application, that is, consider a trajectory of the form $s, f(s), F\,f(f(s)), \ldots$. However, the output would accumulate levels of functor application, which is not desired. For example, in the case of probability, one would like to obtain a probability over states in every iteration, not a probability over states in the first step, a probability over probabilities over states in the second step and so on. Here, the monad can help.

Associated to a monad on a category $\mathbb{C}$, there is the so called Kleisli-category of the monad.

Definition 3.3.7. The *Kleisli-category* $\mathbb{K}_F$ of a monad F on a category $\mathbb{C}$ is constructed as follows:

- Objects are the objects of $\mathbb{C}$.

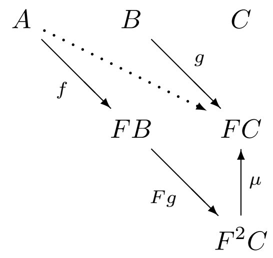
- Arrows $f : A \longrightarrow B$ in $\mathbb{K}_F$ are arrows of the form $f : A \longrightarrow FB$ in $\mathbb{C}$. These are also referred to as Kleisli-arrows. $Hom_{\mathbb{K}}(A, B)$ abbreviates $Hom_{\mathbb{K}_F}(A, B)$ when the monad is clear.

- For any object A, the identity arrow in $\mathbb{K}_F$ is given by η_A.

- Kleisli-composition, denoted $\diamond$, works as follows. Take arrows $f : A \longrightarrow B$ and $g : B \longrightarrow C$ in $\mathbb{K}_F$. That means, there are the corresponding arrows $f : A \longrightarrow FB$ and $g : B \longrightarrow FC$ in $\mathbb{C}$. Define $g \diamond f$ in $\mathbb{K}_F$ by the following operations in $\mathbb{C}$:

$$g \diamond f = \mu \circ Fg \circ f$$

The illustration by diagram, with $g \diamond f$ given by the dotted arrow, is



In $\mathbb{C}$, the composed arrow is of type $A \longrightarrow FC$, thus, in $\mathbb{K}_F$, we have $g \diamond f : A \longrightarrow C$.

Considering an arrow of type $A \longrightarrow FA$ in $\mathbb{C}$, Kleisli-composition allows the repeated application desired for the dynamical system. Thus, having a functor for uncertainty description which is also a monad, one can take as transition function a Kleisli-arrow. Fortunately, the powerset functor seen above forms a monad, and there is a monad of probability measures. That this is also the case for finitely additive probabilities is shown in Chapter 6.

The description given here was simpler than the definition by Ionescu [2009], because only discrete time dynamical systems were considered. The definition generalizing this, for example to continuous time, involves a monoid, that is, a set together with an associative operation and a distinguished element from the set that works as unit for the operation. In the more general dynamical system, a monoid $(T, +, 0)$ represents time. The Kleisli-category construction yields another monoid: for a fixed object A, the set $Hom_{\mathbb{C}}(A, FA)$, or $Hom_{\mathbb{K}}(A, A)$, together with the operation Kleisli-composition and the unit η_A forms a monoid because Kleisli-composition is associative and has η as unit. Consider given an object S of a (locally small[5]) category, fixed as the state space, and a monad with the corresponding Kleisli-category. Then,

Definition 3.3.8. A *monadic dynamical system* with state space S is a monoid morphism sys from a monoid $(T, +, 0)$ to the monoid $(Hom_{\mathbb{K}}(S, S), \diamond, \eta_S)$.

In the discrete time case, the time monoid consists of the natural numbers with addition and 0, and sys is determined by a transition function: denoting $sys(1) = f$, one has $sys(n) = f^n$ for all $n \in \mathbb{N}$. Here, $f : S \longrightarrow FS$ in $\mathbb{C}$. Let us consider the trajectories of this system. Given a starting state $s \in S$, an application of the transition function yields a result of type FS. As discussed above, this can represent uncertainty over the next state. In the cases of interest here, it is a set of possible next states, possibly together with some information about these, such as their probability.

Making two steps means applying $f \diamond f : S \longrightarrow FS$. The result is again of type FS, and still represents uncertainty as viewed from the starting state s, but now referring to the possible states after two steps. The sequence of elements of type FS is called a *macro-trajectory* by Ionescu. One can also consider what he calls *micro-trajectories* by forming all

[5]Again, see Ionescu [2009] for details.

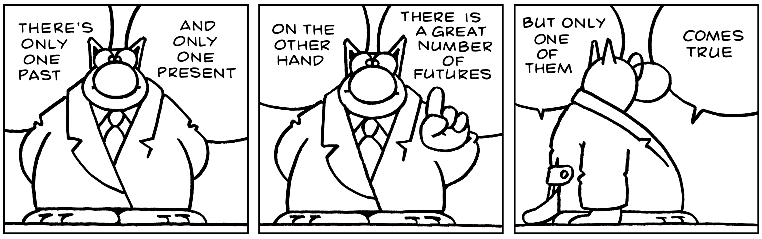
possible sequences of states that the system could go through given a macro-trajectory. Let us illustrate for the non-deterministic example[6]:

Consider two transitions of a non-deterministic system given by a transition function $f : S \longrightarrow \mathfrak{P}(S)$. The possible states after two steps are computed by first computing the possible states after one step and then applying one further step to each possible state. The union over all the sets obtained from the second step transitions is the result. System iteration thus means making new sets out of sets, without choosing a new starting state to "start anew" in the intermediate steps. This captures the idea that uncertainty increases as the time horizon shifts further into the future. The macro-trajectory here consists of two sets: the result of the first step and this latter union. The micro-trajectories are all sequences of the form s, s_1, s_2, where s is the given starting state, s_1 is an element of the set $f(s)$ and $s_2 \in f(s_1)$. Micro-trajectories can be seen as a description of the system evolution when time actually passes.

[6]The formal description by Ionescu [2009] is very computationally focussed and goes beyond the scope of this work.

Chapter 4

Probability



Geluck [2008][1]

This chapter on probability serves two purposes. First, it introduces the concept (Section 4.1) and several mathematizations of probability to provide the basis for the monads presented in Chapter 6. The mathematizations are: classical probability (Section 4.2), the well-known axiomatization proposed by Kolmogorov in his *Grundbegriffe der Wahrscheinlichkeitsrechnung* [Kolmogorov, 1933] in Section 4.3, finitely additive probability (Section 4.4), and a mathematical model of (subjective) probability based on work by de Finetti (Section 4.5). In this order, each mathematical model generalizes the previous one. The different models provide some insights on the method of formalization.

Second, different interpretations of the concept probability are sketched (Section 4.6). Together with the different mathematical models, these are discussed in relation to formalization (Section 4.7). The concept probability thus provides an example of what formalization can do and what it does not do. Essentially, the statement by Suppes [1968] quoted in Section 2.4 will be confirmed: formalization does not end the debate but raises the discussion to another level. Finally, Section 4.8 considers probability in the context of climate change research as an example of how the concept is used (or avoided) before Section 4.9 concludes.

4.1 The concept probability

Probability is defined as "the quality or state of being probable; the extent to which something is likely to happen or be the case" [Soanes and Stevenson, 2005] by the Oxford Dictionary of English. In ordinary language, we use phrases like 'something is probable' or 'more probable than another thing'. In "our 'unconscious reasoning' in making decisions under uncertainty" Diaconis [2007a], probability serves to express additional information about different possibilities, as described by de Finetti [1974]:

> There are no degrees of possibility: it is possible (equally possible) that it snows on a winter or summer day. [. . .] Faced with uncertainty, one feels . . . a more or less strong propensity to expect that certain alternatives rather than others will turn out to be true. [p. 71].

[1] With kind permission of the author.

A basic idea underlying the concept probability is that in the long run more probable things happen more often. This aspect of the concept is referred to when probability is linked to statistical regularities, as for example by Botts [1969]:

> Probability theory is the study of mathematical models for random phenomena. A random phenomenon is an empirical phenomenon whose observation under given circumstances leads to various different outcomes. [...] the various possible outcomes appear to occur with what is called "statistical regularity". This means that the relative frequencies of occurence of the various possible outcomes appear to approach definite limiting values as the number of independent trials of the phenomenon increases indefinitely. [p. 123]

A standard example to explain what is meant by probability is that of tossing a coin. Persi Diaconis, professor of statistics and mathematics as well as magician[2], has somewhat invalidated this example by showing that, if thrown under the same conditions, a coin always lands on the same side, and by, moreover, learning to toss the coin deterministically by hand [see Landhuis, 2004]. However, assuming a less well-trained tosser, one for example asserts that "a coin is fair" and expresses this by saying that in one toss the probability of heads and that of tails are each 1/2. It is unclear what exactly this means for a single toss, but in a long series of repetitions of the coin toss, one expects both sides to appear with more or less equal frequency. The 'more or less' is where the everyday use of the concept is hard to make precise. It seems "built into the very notion of probability" that "misleading results", as when a fair coin lands heads 9 out of 10 times, can occur [see Hàjek, 2007, p. 16]. Alleged coincidences [see, e.g. Spizzichino, 2009] and debates such as the one about the famous Monty Hall problem illustrate that intuitions about probability are often contrary to long-run observations; basic concepts such as 'chance' are unclear for example to students confronted with probability theory [see Sill, 1993]. Also, psychological studies show that people commonly violate basic rules of mathematical probability [see Kahneman et al., 1982]. The concept thus seems far from clear in its everyday use.

Literature on the history of probability tries to explain the late emergence of the concept in the seventeenth century.[3] Hacking [1975], who traces the "emergence of probability", expresses a duality of the "Janus-faced" concept probability as follows.

> On the one side it is statistical, concerning itself with stochastic laws of chance processes. On the other side it is epistemological, dedicated to assessing reasonable degrees of belief in propositions quite devoid of statistical background. [Hacking, 1975, p. 12].

For 'statistical' he also uses the terms 'aleatory' and later 'physical', we will use the terms *frequency* and *belief* to refer to the two aspects of probability.

Hacking lists a number of attempts at separating these aspects into two concepts[4] and claims that "there is something wrong about" the duality of probability: "Why all these frantic gropings for a terminology to make distinctions?"[p. 12]. Let us sketch his description of how probability emerged. He first of all points out an old meaning of the concept probable, deriving from Latin 'probabilis', which corresponds to 'approvable' or 'worthy of approbation'. The dictionary entry quoted above similarly gives the origin of the word probable as: "late Middle English (in the sense of 'worthy of belief')" [Soanes and Stevenson, 2005]. Then, Hacking describes a Medieval distinction between knowledge and opinion, which treat different objects: knowledge, aspired in the high sciences, concerns universal truths or can be

[2]The following biographical hints can be found for example in Aldous and Diaconis [1986]: Persi Diaconis left High School at an early age to earn a living as a magician and gambler, only later to become interested in mathematics and earn a Ph.D. at Harvard. After a spell at Bell Labs, he is now Professor in the Statisitcs Department at Stanford. ...

[3]There is literature on early ideas related to probability which we will not go into. For example, Franklin [2001] investigates "The science of conjecture, evidence and probability before Pascal", and Maĭstrov [1974] dedicates a chapter to the "prehistory of probability theory".

[4]Carnap's *probability*$_1$ and *probability*$_2$ are probably best known; others have suggested to distinguish using for example the French words *chance* and *probabilité* or *facilité* and *motiv de croire*.

established by demonstration, whereas opinion, occuring in the low sciences, "tends to refer to belief which results from by reflection, argument, or disputation" [p. 21] by alchemists, astrologers, miners and physicians. Such opinions are "probable when they are approved by authority, when they are testified to, supported by ancient books." [Hacking, 1975, p. 30].

Hacking further discusses the lack of the concept of evidence as used today and explains the concept of sign that was in use instead. Signs could be "anything by which we may make a prognosis" [Hacking, 1975, p. 28], for example, in a medical diagnosis not only the symptoms observed, but just as well the constellation of the planets. The concept of evidence emerges when a distinction is being made between natural and conventional signs, such as signs of language.

> A new kind of testimony is accepted: the testimony of nature which, like authority, was to be read. Nature now could confer evidence, not, it seemed in some new way, but in the old way of reading and authority. Thus: to call something probable was still to invite the recitation of authority. But: since the authority was founded on natural signs, it was usually of a sort that was only 'often to be trusted'. Probability was communicated by what we should now call law-like regularities and frequencies. Thus the connection of probability, namely testimony, with stable law-like frequencies is a result of the way in which the new concept of internal evidence came into being. [Hacking, 1975, p. 44].

In conclusion, the two aspects, degree of belief and frequency, are inherently linked in the concept probability, or, the other way around,

> Aleatory probabilities have to do with the physical state of coins or mortal humans. Epistemic probabilities concern our knowledge. [...] the word 'probability' was annexed to this pair of concepts only after 1662. [Hacking, 1975, p. 123].

While there are previous works, which from today's point of view can be related to the concept probability, the beginnings of *probability theory* are generally dated with the famous correspondence between Pascal (1623-1663) and Fermat (1601-1665) in 1654 (published in 1679) about questions related to gambling. One finds gambling problems in earlier works, for example in Cardano's 'De ludo aleae', written around 1550, but printed only in 1663, where the concept proclivity and the 'ability' to throw various combinations with several dice occur. Galilei wrote about dice games[5], observing that in throwing several dice "some numbers are more easily and more frequently made than others" [as cited by Hacking, 1975, p. 52].[6] Further, in the work of Hobbes, which discusses *conjectural* signs, with more or less *assurance* "according as they have often or seldom failed", in 1650, probability "has emerged in all but name" [Hacking, 1975, p. 48].

The term probability itself (respectively an adequate translation) occurs neither in the letters between Pascal and Fermat, nor in the first printed textbook of probability by Huygens [1657], which is inspired at the problems discussed by Pascal and Fermat. Huygens uses the concept expectation for the fair price of a lottery.

A theoretical definition of probability is given by Jakob Bernoulli (1654-1705) in *Ars Conjectandi* that was published posthumously in 1713: "Probabilitas enim est gradus certitudinis, & ab hac differt ut pars á toto." – "Probability is a degree of certainty, and differs from certainty as a part from a whole" [as cited and translated by Shafer, 1996, p. 7]. Bernoulli also describes the "art of conjecturing" itself as

> the art of measuring the probability of things as exactly as possible, to be able always to choose what will be found the best, the more satisfactory, serene and reasonable for our judgements and actions. [Bernoulli, 2005].

That is, the aim of this art is to support decision-making. Reasoning about uncertainty without a calculus was perceived to be difficult. Bernoulli, for example, points out "a mistake into which it would have been easy to fall, *nisi nos calculus aliud docuisset*" [Todhunter,

[5] Sopra le Scoperte dei Dadi.
[6] See also Chapter 5 in Maĭstrov [1974].

1949, p. 63]. The calculus of probability is directly referred to by the original name of the mathematical discipline in several languages, e.g. the German 'Wahrscheinlichkeitsrechnung' for which Kolmogorov introduced his foundations. It has been dubbed the 'logic of uncertainty', suggesting the comparison with logic, a "suitable superstructure" to support intuition in reasoning [see de Finetti, 1974, p. 25].

The original mathematical superstructure, or model, to support reasoning with uncertainty is today referred to as *classical probability.* This calculus, developed and used from Pascal and Fermat's times on[7], shall be considered in the following section, because it gives some insight into the concept probability itself (at least for simple cases). It thus confirms the hypothesis that formalization can clarify the meaning of concepts.

4.2 The classical calculus of probability

A historical description of the *theory of chance* can be found in the work of Laplace[8]:

> The theory of chance consists in reducing all the events of the same kind to a certain number of cases equally possible, that is to say, to such as we may be equally undecided about in regard to their existence, and in determining the number of cases favorable to the event whose probability is sought. The ratio of this number to that of all the cases possible is the measure of this probability, which is thus simply a fraction whose numerator is the number of favorable cases and whose denominator is the number of all the cases possible.

Probability as taught in high school today comes close to the classical calculus of probability. The basic idea is to consider a *Fundamental Probability Set* [Neyman, 1950], a finite set of equally probable possibilities, here Ω. With $|\Omega| = n$, each possibility is assigned a probability of $\frac{1}{n}$. Events are subsets of the set of possibilities. To all events $A \subseteq \Omega$ a probability is associated by the rule

$$P(A) = \frac{\text{number of favourable cases}}{\text{number of possible cases}} = \frac{|A|}{n}. \tag{4.1}$$

Laplace's formulation above has been read as a (theoretical) definition of probability and accused of circularity due to the "equally possible" cases:

> The notion of "equally possible" cases faces the charge of either being a category mistake (for 'possibility' does not come in degrees), or circular (for what is meant is really 'equally probable'). [Hàjek, 2007].

Recall[9] the quote from de Finetti [1974] stating that there are no degrees of possibility and asserting an equal possibility of snow on a winter and a summer day. This is not what is meant by 'equally possible' in classical probability. In fact, most people would not describe snow on a winter and on a summer day as 'equally possible' in the sense of Laplace's quote, where 'equally probable' is meant.

The challenge of circularity is however invalidated by Rényi [1969], partly in disguise as Pascal[10]. While 'equally possible' and 'equally probable' means the same thing, a circularity of definition is avoided if one considers the statement as a rule to establish the numerical value of probabilities in particular cases only, that is, an operational, and not a theoretical definition. In fact, Laplace defines "the measure of this probability" here, not the meaning of the concept probability. Rényi considers this as given by Bernoulli's definition, probability as the degree of certainty. To show that there is no circularity of definition, he lets Pascal

[7] A detailed history of probability from the times of Pascal to that of Laplace has been published as early as 1865 by Todhunter.

[8] English translation as cited by Hàjek [2007, p. 7].

[9] See page 35.

[10] Rényi [1969], titled "Letters on Probability", presents four fictitious letters by Pascal addressed to Fermat that provide a semi-popular and semi-historical account of classical probability. The argument here is given by "Pascal" in the second letter as well as by Rényi in a letter to the reader concluding the book.

conclude equal probability of all faces of a die by an argument of symmetry: it would not be noticed if the faces were secretly renumbered. For this conclusion, it is not necessary to measure numerical values of the probabilities beforehand, just as one can decide whether two items are equally heavy by weighing them with a double pan balance without measuring the weight in grams. Having established that the cases are equally probable, the measurement of the numerical value of the probabilities reduces to a computation, the simple division of 1 by the number of cases, so that each face of the die has the degree of certainty $\frac{1}{6}$.

In this simple setting, the assignment of equal numbers, adding up to one, represents the intuition that exactly one of the possibilities will turn out true while all possibilities are considered equivalent or "equally possible". This intuition can be considered a "clear interpretation" of the concept:

> What does probability mean? The earliest clear interpretations of probability [were] simple gambling problems where symmetry or the open physical process of drawing from an urn make the basic equally likely outcomes intuitively clear. [Diaconis, 2007b].

Probability here occurs as a primitive: in the simple case 'equal probability' is intuitively clear. The mathematical definition (4.1) makes the degree precise, and hence, in this simple case, the measured probabilities clarify the meaning of the concept. Once given numbers, probabilities of events can be compared also where the symmetry argument does not apply: consider, for example, events from two different games. Within this setting, mathematical conclusions can be and have been drawn for several centuries. As just one relatively simple example consider the famous question treated in the Pascal-Fermat correspondence: "In throwing two dice, how many tosses are needed to have at least an even chance of getting double-six?"[as in Hacking, 1975, p. 75]. Answers to such questions can be considered support for reasoning with uncertainty for the people who intend to play these games.

However, the great interest in gambling problems, rather than truly practical, seems to be motivated from theoretical considerations: games, and probably only games, provide situations with intuitively equivalent possibilities. Of course, the assumption that, for example, a die is fair is extramathematical.[11] Classical probability has been criticized for example because it is generally unclear how to obtain the equivalent possibilities to which the concept can be applied in situations other that simple games. Possibilities are judged equivalent given symmetrically balanced evidence or in the absence of evidence. The latter "principle of indifference", assuming the equality of probabilities for lack of information, has been accused of "extracting information from ignorance" [Hàjek, 2007, p. 10] and can be used in incompatible ways.[12] Gnedenko rejects the principle of indifference for a non-symmetrical coin.

> Clearly one should not say that in the first, or second, or any toss of the coin the probability of a certain side is $\frac{1}{2}$, it is simply unknown. Its determination or estimation should be made not by doubtful means which deprive the notion of probability of its objective role as a numerical characteristic of definite real phenomena. [As cited by Maĭstrov, 1974, p. 145].

Nevertheless, it seems easy enough to accept the basic assumption of a given Fundamental Probability Set in order to do some mathematics. The slightly more general mathematical model where different probabilities may be assigned to the finitely many given possibilities has also been studied extensively, generally without however making explicit an operational definition that would yield such probabilities. The very inventors/discoveres of classical probability see a connection between the games studied and a theory of probability. While Huygens merely hopes his reader notices the "foundations of a new theory" underlying his

[11] At first sight, the example of an urn from which coloured tickets are drawn seems clearer: looking at the number of tickets of each colour in the urn, probabilities seem more intuitive than those of certain faces of a die, which might be invisibly loaded. However, the assumption that any ticket is drawn from the urn with equal probability is just as basic as that of the fairness of the die.

[12] See, for example, Hàjek's [2007] comments on Bertrand-style paradoxes.

treatment of games[13], Bernoulli explicitly states that the games are purposefully constructed in such a way that classical probability applies:

> We thus see that for correctly conjecturing about some thing, nothing else is required than both precisely calculating the number of cases and finding out how much more easily can some of them occur than the others. Here, however, we apparently meet with an obstacle since this only extremely seldom succeeds, and hardly ever anywhere excepting games of chance which their first inventors, desiring to make them fair, took pains to establish in such a way that the number of cases involving winning or losing were determined with certainty and known and the cases themselves occurred with the same facility. [Bernoulli, 2005]

At the beginning of the eighteenth century, Montmort even points out that games are a pretext, having in mind "the enjoyment of the mathematicians and not the advantages of the players; it is our opinion that those who waste time on games fully deserve to lose their money as well." [As cited by Maĭstrov, 1974, p. 77].

While games are considered the most intuitive examples, even here the Fundamental Probability Sets are not necessarily obvious. Again, the example of tossing several dice is illustrative and long since considered. For tosses of three dice, Galilei explains that "some numbers are more easily and more frequently made than others, which depends on their being able to be made up with more variety of numbers" [as cited by Hacking, 1975, p. 52]. A distinction has to be made between partitions and permutations: throwing a sum of 3 or 4 with three dice, both numbers can be made up by one partition, the 1-1-1 for 3 and the 1-1-2 for 4. However, the latter arises from three permuations, 1-1-2, 1-2-1, and 2-1-1, while there is only one permutation making up 1-1-1. It seems obvious today that the equally probable alternatives here are the permutations, but even "Leibniz once made the mistake of supposing that with dice partitions (rather than permutations) form the Fundamental Probability Set." [Hacking, 1975, p. 52]. For example, for photons and electrons this set has to be obtained differently. The case for dice is merely a well established example by now:

> It is not obvious that permutations rather than partitions are equally probable. Indeed it is arguably an empirical fact about dice that can only be learned from observation. [Hacking, 1975, p. 51].

In fact, Galilei's explanation is motivated by the fact that "it is known from long observation that dice players consider 10 and 11 to be more advantageous than 9 and 12." [as cited by Hacking, 1975, p. 51]. Here, the frequency aspect of probability becomes explicit. This aspect is not immediate in classical probability. There is, for example, no obvious Fundamental Probability Set for a coin that is not 'fair'. With the generalization that different probabilities may be attached to the two possibilities 'heads' and 'tails', a model of an unfair coin suggests itself, however, how to choose the numbers is then a question that cannot be answered within the mathematical model. That this question is still up to debate will be seen in Section 4.6.

Nevertheless, applications of classical probability to statistical questions were carried out right from the beginning. Huygens was the first to apply methods of probability theory to demographic statistics in 1669, and Euler in the 1760s posed and solved many problems that later became the basis of mathematical demography [see Maĭstrov, 1974]. Hacking [1975] states that already at Galilei's times "the probability calculus is a calculus of frequencies." [p. 53] and indicates the years 1675/76 as the time by which "it seems, problems about dicing and about mortality rate had been subsumed under one problem area" [p. 109].

A first mathematical result that explicitly clarifies the connection between classical probability and the frequency aspect of probability is Bernoulli's law of large numbers stating that, when given probabilities, relative frequencies converge to these in a simple setting.

[13] "I would like to believe that in considering these matters closely, the reader will observe that we are dealing not only with games but rather with the foundations of a new theory." [As cited by Maĭstrov, 1974, p. 58].

Bernoulli calls upon common sense to make this connection work the other way around: he wants to measure probability values from relative frequencies in long sequences:

> what we are not given to derive a priori, we at least can obtain *a posteriori*, that is, can extract it from a repeated observation of the results of similar examples. Because it should be assumed that each phenomenon can occur and not occur in the same number of cases in which, under similar circumstances, it was previously observed to happen and not to happen. [...] This empirical method of determining the number of cases by experiment is not new or unusual. [...] note also that it is not enough to take one or another observation for such a reasoning about an event, but that a large number of them are needed. Because, even the most stupid person, all by himself and without any preliminary instruction, being guided by some natural instinct (which is extremely miraculous) feels sure that the more such observations are taken into account, the less is the danger of straying from the goal. [Bernoulli, 2005, p. 18]

This is an operational definition of probability of frequentist type as will be seen in Section 4.6.1. Problems with this a posteriori measurement of an a priori probability are also discussed there. The search for a means to

> determine the probability of a future or unknown event not on the basis of the number of possible combinations resulting in this event or in its complementary event, but only on the basis of the knowledge of order of familiar previous events of this kind [Condorcet, 1785, as cited by Todhunter, 1949, p. 377]

is ongoing. As soon as one leaves the situation of equally probable alternatives, the measurement question for probability is not answered by the mathematical model. It is, however, considered fundamental, for example, by Boldrighini [2009] who points out that the question how to obtain numbers from a situation of uncertainty seems to be implied by the Italian and French names of the discipline ("calcolo delle probabilità", respectively "calcul des probabilités") which literally translate to "calculation of probabilities" [see p. 256]. Hajek[14] uses 'ascertainability', the requirement that there be some method for obtaining values of probabilities, that is, measure them, as a criterion in the discussion of interpretations of probability.

At the same time it is considered essential to distinguish the meaning from methods of measurement, for example, by Coletti and Scozzafava [2002], who claim that

> ignoring this distinction would be analogous to identifying the concept of temperature with the numbers shown by a thermometer, [so that one would not be] entitled to speak of temperature in a room without a thermometer [p. 24].

The questions of meaning and measurement were seen to be closely related for classical probability. In the vulnerability context, it will be mostly measurement that clarifies meaning. We will return to these questions and to the development from classical probability to today's standard probability theory in Section 4.7. Here, we jump directly into the mathematics in the following section.

4.3 "Standard" mathematical probability

This section presents some basics of what is today found in most probability textbooks.[15] The axioms go back to Kolmogorov's formalization of probability, which "has achieved the status of orthodoxy"[Hàjek, 2007].

4.3.1 Kolmogorov's axioms

Kolmogorov's axioms embed probability into measure theory. Since the functions associating probabilities to events in this model are measures, the term *probability measure* will be used.

[14]Based on Salmon [1966].
[15]The selection of presented material is driven by what is required in Chapter 6. It does not aspire completeness in any sense.

Definition 4.3.1. A probability measure is defined as a measure with unit mass, that is, a function $P : \mathcal{F} \to [0, 1]$ defined on a measurable space $(\Omega, \mathcal{F})$, with

(*i*) $P(\Omega) = 1$ and

(*ii*) For disjoint sets $A_1, A_2 \in \mathcal{F}$, $P(A_1 \cup A_2) = P(A_1) + P(A_2)$. (*additivity* or *finite addivity*[16])

(*iii*) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ for disjoint $A_1, A_2, \ldots \in \mathcal{F}$ (σ-*additivity* or *countable addivity*).

The triple $(\Omega, \mathcal{F}, P)$ is called a *probability space*, Kolmogorov's term 'probability field' is not used anymore.

From the first two axioms one can immediately deduce that $P(\emptyset) = 0$, $P(A^c) = 1 - P(A)$ for all $A \in \mathcal{F}$, and $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$ for any finite number n of disjoint $A_1, \ldots, A_n \in \mathcal{F}$, in fact, *finite additivity* is often stated in this form. Another immediate property of (probability) measures is "monotonicity": if $A \subseteq B$ then $P(A) \leq P(B)$.

The mathematical concepts come with an interpretation along the following lines. The non-empty set Ω represents the possible outcomes of a random experiment, where 'experiment' is to be understood in a very general sense, it might be any situation with more than one possible outcome depending on chance. The $\omega \in \Omega$ are referred to as *elementary events*. Ω is also known as the *sample space* and its elements ω as *sample points*. Mathematically, the only restriction on Ω is that it is not empty. Otherwise, it may be arbitrary, in particular infinite and uncountably so. When representing an experiment, it is of no harm to include impossible outcomes in Ω, however, all possible outcomes should be included.

Events are represented by subsets of Ω. An event is a statement the validity of which depends only on the outcome of the experiment. The analogy between events and subsets derives from the fact that for each outcome one can decide whether an event has taken place or not, and likewise, for each ω, one can decide wether $\omega \in A \subseteq \Omega$, or $\omega \notin A$ for a given set.

The σ-algebra $\mathcal{F}$ over Ω consists of the events under consideration. Its use is mainly motivated from measure theory. For didactical reasons, Rényi [1970] introduces the σ-algebra as the set of observable events, pointing out that due to the degree of precision with which the outcome is measured "not every subset of Ω corresponds in this way to an event which can really be observed and which is of interest."[p. 2]. Closure under the set operations union and complement may be probabilistically motivated by the fact that if one can observe some event, one can certainly also observe that it has not occured, i.e. observe its complement, and for a sequence of events one can also observe whether "at least one" of them has occured, i.e. their union is an event as well.

The function P associates to each event its probability. That the certain event has probability 1 is a norming convention, one could use 100 and express probabilities in percent etc. However, that $P(\Omega)$ is finite is required to express a *degree*, fitting Bernoulli's theoretical definition of probability as a degree of certainty.

4.3.2 Basic properties

If the sample space is finite, say $\Omega = \{\omega_1, \ldots, \omega_n\}$, a probability measure is well-defined by assigning each $\omega_i \in \Omega$ a probability $P(\{\omega_i\}) = p_i \geq 0$ with the condition that $\sum_{i=1}^{n} p_i = 1$. The probability of any subset of Ω is defined as the sum of the probabilities of its elements

$$P(A) = \sum_{i:\omega_i \in A} p_i. \tag{4.2}$$

[16]Often, only the condition of σ-additivity is stated. It includes finite additivity because, except for a finite number of sets, all sets in the sequence can be the empty set. The latter is disjoint of itself, since, not having any elements, it does not share any elements with itself.

The σ-algebra on which P is defined can be (and by default is) chosen as the powerset of Ω. Additivity and the norming condition $P(\Omega) = 1$ follow immediately from the definition of P in (4.2), while countable additivity is vacuously satisfied. There are no (truly) infinite sequences of disjoint sets in $\mathfrak{P}(\Omega)$, which, Ω being finite, is itself finite. In fact, on a finite set Ω, all probabilities are of this form.[17] A special case is that of classical probability, where all ω_i are asigned equal probability, that is $\frac{1}{n}$. As was seen, computing the probability of a set reduces to counting its elements and dividing by the number of elements of the sample space. Many probability textbooks, and also Kolmogorov's *Grundbegriffe*, introduce probability for finite sample spaces first. In such treatments of "elementary probability theory" the axiom of finite additivity is given instead of the countable version.

Somewhat more generally, the same construction holds when the (σ-)algebra under consideration is finitely generated. Again, a probability measure (finitely additive, and vacuously countably additive) is fully determined by assigning probability values to the sets in the partition which generates the algebra, the constituents.

Remark 4.3.2. If $\mathcal{A}$ is generated by a finite partition $\{A_1, \ldots, A_n\}$ of Ω, then $P(A) = \sum_{A_i \subseteq A} P(A_i)$ for all $A \in \mathcal{A}$. That means, a probability measure on a (σ-) algebra that is generated by such a finite partition is completely determined by the values given to the elements A_i of the partition.

An analogous construction of probability measures can be carried out when the sample space Ω is countably infinite, say $\Omega = \{\omega_1, \omega_2, \ldots\}$ (or, again, the σ-algebra is generated by a countable partition). One assigns values p_i to the elements ω_i of Ω, now under the condition that $\sum_{i=1}^{\infty} p_i = 1$, and defines P on $\mathfrak{P}(\Omega)$ via formula (4.2). According to Theorem 3.2.2, this probability measure can be defined on the whole powerset. The resulting P is countably additive. Probability measures of the above types are called *discrete probability measures.*

When the set Ω is uncountable, for example, when $\Omega = \mathbb{R}$, measure theory starts playing a role. As seen in Section 3.2, probability measures on the whole powerset of Ω can be defined only when they concentrate the mass on at most countably many points in Ω. For all other cases, smaller σ-algebras are needed. On $\mathbb{R}$ it is common to work with the σ-algebra of the Borel-sets $\mathcal{B}(\mathbb{R})$.

A continuous, that is, not discrete, probability measure P on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ can be characterized by its *distribution function* $F : \mathbb{R} \to [0, 1]$, defined by $F(x) = P((-\infty, x])$ for all $x \in \mathbb{R}$. A probability measure is also referred to as a *probability distribution* in the literature, and sometimes the terms probability measure/distribution on the one hand and distribution function on the other are used interchangeably without considering that values in $[0, 1]$ are assigned in the first case to sets and in the second case to real numbers. A sloppy use of the terms can be somewhat excused by the following result.

Theorem 4.3.3. *The distribution function* $F : \mathbb{R} \to [0, 1]$ *of any probability measure* P *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *has the following properties:*

1. *it is increasing, that is, for* $x \leq y, F(x) \leq F(y)$*,*
2. *it is right-continuous, that is,* $\lim_{h \downarrow 0} F(x + h) = F(x)$*, and*
3. $\lim_{x \to -\infty} F(x) = 0$ *and* $\lim_{x \to \infty} F(x) = 1$.

Conversely, for any function $F : \mathbb{R} \to [0, 1]$ *with these properties there is a (unique) probability measure* P *on* $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *with* $F(x) = P((-\infty, x])$ *for all* $x \in \mathbb{R}$.

Here, the term *probability distribution* will be used in a rather non-technical sense to refer to a function assigning probabilities to events, where it is not specified whether a probability measure or a finitely additive probability is considered. The intuition is that of probability as a mass which is distributed over the different possibilities.

[17] Technically speaking, one could choose a smaller σ-algebra, but this is rarely or never seen.

4.3.3 Further concepts

This section introduces some mathematical concepts that are needed elsewhere. Mathematical detail is skipped where it is not required.

Concerning functions, contemporary probability theory uses a terminology and notation which is rather far from standards in many other branches of mathematics. This is probably due to the late introduction of Kolmogorov's axioms, which use functions for a concept that had long been studied but not precisely defined. Only in this axiomatization it came to be associated with functions defined on an underlying sample set: *random variables.* Nowadays, *random variables* are measurable functions with target[18] $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. The term "random variable" for a (non-random) function of a "random" argument may be unfortunate, but is widely accepted.[19] It is now common to denote random variables with capital letters $X, Y, \ldots$[20] unlike in other branches of mathematics where small letters are used to denote functions. We will stick to this convention here.

Measurability of a random variable $X : \Omega \to \mathbb{R}$ means that for all $x \in \mathbb{R}$, the set $X^{-1}((-\infty, x])$ is an element of $\mathcal{F}$, which in probability texts is denoted as $\{X \leq x\} \in \mathcal{F}$, abbreviating $\{\omega \in \Omega | X(\omega) \leq x\} \in \mathcal{F}$.

Because a random variable is a measurable function, it induces a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. This is referred to as the *distribution* of the random variable and denoted P^X. It is defined for Borel-sets $B \subseteq \mathbb{R}$ by

$$P^X(B) = P(X^{-1}(B)) = P(\{\omega | X(\omega) \in B\}) \text{ also written as } P(X \in B).$$

Its distribution function is denoted and defined $F_X(x) = P(X \leq x)$ for $x \in \mathbb{R}$.

The *expected value* of a random variable is defined as an integral of Lebesgue type.

$$E(X) = \int X dP.$$

For discrete probability measures, this reduces to the weighted sum of values the random variable takes, where the weights are the probabilities. When the probability measure is to be emphasized, the notation $E_P(X)$ is chosen. The operator itself is also known as *expectation.*

Finally, for a simple model of vulnerability in Chapter 5, the concept of a Markov chain is needed. It will be introduced for a simple special case only.[21]

Definition 4.3.4. Consider for simplicity[22] a finite set S, called the *state space* of the Markov chain. Consider its elements numbered from 1 to n. The underlying idea is that of a dynamical system which can be in any of the n states and which, when in a given state at a certain point in time, makes its transition to a next state at the next time point according to a probability measure over the n states. The further assumption that this probability measure does not depend on the time point will for simplicity be used here, such Markov chains are called homogeneous in time.

A stochastic matrix, the *transition matrix* provides the transition function. Each row contains a probability measure over the n states, that is, all entries are elements of $[0, 1]$ and

---

[18] Sometimes all measurable functions are referred to as random variables, and the real valued ones as *real random variables.*

[19] Some probability textbooks provide pedestrian explanations to help getting used to the term, as e.g. Jacod and Protter [2000]:

> A random variable represents an unknown quantity (hence the term variable) that varies not as a variable in an algebraic relation (such as $x^2 - 9 = 0$), but rather varies with the outcome of a random event. Before the random event, we know which values X could possibly assume, but we do not know which one it will take until the random event happens. This is analogous to algebra when we know that x can take on *a priori* any real value, but we do not know which one (or ones) it will take on until we solve the equation $x^2 - 9 = 0$ (for example). [p. 21].

[20] Kolmogorov [1956] uses small x-es.

[21] The usual introduction of conditional probabilities and stochastic processes will be bluntly circumvented.

[22] A further advantage is avoiding the decision for or against countable additivity, see next section.

the matrix has column sums 1. The probability measure in row i gives the probabilities of next states when starting out in the state numbered i. The matrix entry p_{ij} in row i and column j is the transition probability from state i to state j. The probability of a certain sequence of states, or a trajectory of the system, can be computed from these values when given an initial probability measure, which assigns each state the probability of starting there. The probability of a given sequence $s_1, s_2, \ldots s_k$ is simply the probability of starting in state s_1, multiplied with the probability of transiting from s_1 to s_2, which is $p_{s_1 s_2}$, etc, multiplied with $p_{s_{k-1} s_k}$. A frequently used special case is that of a fixed starting state, then the initial probability measure is the concentrated probability in that state.

4.3.4 Remarks on formalization

Kolmogorov proposed his axioms for probability in a time where "probability was generally not considered to be mathematics, and many people thought it was only partly possible to mathematise probability" [Boldrighini, 2009, p. 255]. However, Kolmogorov claims:

> The theory of probability, as a mathematical discipline, can and should be developed from axioms in exactly the same way as Geometry and Algebra. [Kolmogorov, 1956, p. 1].

His focus is on probability *as a mathematical discipline*, and he states the task he had set himself as "putting in their natural place, among the general notions of modern mathematics, the basic concepts of probability theory - concepts which until recently were considered to be quite peculiar" [Kolmogorov, 1956, p. v]. This aim has been met, as e.g. Rényi [1969] comments: Kolmogorov not only provided logically sound foundations for probability, but also inserted it into the "blood circulation" of modern mathematics [see p. 84].

Kolmogorov aims at "the utmost simplicity both in the system of axioms and in the further development of the theory", wherefore he choses as the "most suitable" primitives the concepts of a random event and its probability [Kolmogorov, 1956, p. 2]. Using these as the undefined concepts implies that in this mathematical model the "question what probabilities are and how they are calculated is left open, or better, is placed inside something that is similar to a platonic archetype." [Boldrighini, 2009, p. 257]. In fact, Kolmogorov himself distinguishes the aim of his work from those formalizations[23] of probability in which

> the concept of probability is not treated as one of the basic concepts, but is itself expressed by means of other concepts. However, in that case, the aim is different, namely, to tie up as closely as possible the mathematical theory with the empirical development of the theory of probability. [Kolmogorov, 1956, p. 2].

Kolmogorov is interested in the mathematical theory that sets in once probabilities are given. This theory is developed in the framework of measure theory, based on "analogies between measure of a set and probability of an event" [Kolmogorov, 1956, p.v], which he states have been fully explained for the first time by Fréchet [see p. 8].

Kolmogorov is not concerned with justifying all axioms posed as necessary to describe the concept probability. On the contrary, he explicitly states that the choice of the axiom of countable additivity, "found expedient in researches of the most diverse sort", is an arbitrary choice, and that "it is almost impossible to elucidate its empirical meaning" [Kolmogorov, 1956, p. 15]. He also explicitly allows the mathematical theory to go beyond statements which can be directly interpreted in terms of the concept probability. Considering finite probability spaces an "image (idealized, however) of actual random events", yet, infinite spaces "will still remain merely a mathematical structure" [Kolmogorov, 1956, p. 17].

Kolmogorov's answer to the objection of having made probability theory a part of measure theory was that

> probability differs from measure theory in its intuitive-conceptual dimension, which is not irrelevant in that it determines how problems are posed, and which to a great extent derives from the old "logic of uncertainty". [Boldrighini, 2009, p. 257].

[23] He uses the term "postulational systems" on page 2.

However, the mathematical model is not as closely connected with the intuitive use of the concept probability as will be seen in de Finetti's model.

Answers to the questions what probability means and how it can be measured are provided elsewhere by Kolmogorov. For example, in the entry 'probability (*mathematical*)' in the Encyclopedia of Mathematics, he gives the following theoretical definition of the concept probability:

> A numerical characteristic expressing the degree to which some given event is likely to occur under certain given conditions which may recur an unlimited number of times. [Kolmogorov, 1991, p. 301].

This definition places the concept immediately into a mathematical setting by defining it to be a "numerical characteristic". Instead of one operational definition for measuring this "degree", he proposes several measurement rules for different cases:

> The numerical value of a probability may sometimes be obtained from its "classical" definition [...]. In other, more complicated cases, a statistical approach is needed to determine the numerical value of a probability. [Kolmogorov, 1991, p. 302].

Further, Kolmogorov distinguishes cases where a "definite probability" exists from those, where this is not the case:

> The assumption that a definite probability (i.e. a completely defined fraction of the number of occurrences of an event if the conditions are repeated a large number of times) in fact exists for a given event under given conditions is a hypothesis which must be verified or justified in each individual case. [Kolmogorov, 1991, p. 302].

However, he does not make more precise how to verify this hypothesis, that is, the domain of applicability of the concept probability remains vague.

Kolmogorov's probability axioms are differentially credited with some of the benefits of formalization discussed in Section 2.4. Kolmogorov himself sees mathematical probability as a tool that increases precision compared to the ordinary language use of the concept.

> The mathematical probability may serve as an estimate of the probability of an event in the ordinary, everyday-life sense, i.e. it may render more precise 'problematic' statements usually expressed by the words 'possibly' , 'probably', 'very probably', etc. [Kolmogorov, 1991, p. 302].

However, mathematical probability, according to Kolmogorov, is not intended to clarify the ordinary language concept as such, and is explicitly stated not to explain its meaning.

> ...neither [the] axioms nor the classical approach to probability nor the statistical approach fully explains the real meaning of the concept of "probability"; they are merely approximations to a more and more complete description. [Kolmogorov, 1991, p. 302].

De Finetti formulates rather radically that he does not concede to Kolmogorov's axioms, which "abstractly talk about relations among undefined objects" [de Finetti, 2008, p. 2], the benefit of clarifying the meaning of probability: "...some axioms for probability are introduced [...], without one being able to figure out, unless one knows it already, what 'probability' means." [de Finetti, 2008, p. 2].

On the other hand, Suppes [1968] refers to Kolmogorov's axioms of probability theory as a paradigmatic example of how formalization brings out the meaning of concepts in an explicit fashion, pointing out that previous to these axioms "there was much confusion about even the most elementary properties of probability". We will go into this in the discussion section (4.7).

For now, we conclude that Kolmogorov's axioms do not explain meaning and measurement of probability, because the concept is considered a primitive. Nevertheless, the axioms are perceived as providing some clarity in a situation of confusion. While this formalization of probability aims at providing the mathematical foundations for a mathematical theory, the formalization in Section 4.5 puts the focus between concept and mathematics rather differently. Before turning to this formalization, the concept of finitely additive probabilities, generalizing probability measures, will be introduced and compared to these.

4.4 Finitely additive probability

This section introduces *finitely additive probabilities.* These generalize Kolmogorov's probability measures and have been studied, for example by Dubins and Savage [1965], Dubins [1974], Purves and Sudderth [1976]. They are, however, generally not taught in courses of 'probability theory'. One reason may be that one cannot make use of convenient results from measure theory.

In this section, the term probability is used also to refer to the function that assigns probabilities to events. Since it will always be used together with the qualifyiers "finitely additive", there should not be any risk of confusion between the function and its values for some events. The term *probability measure* is reserved for a countably additive function defined on a σ-algebra throughout, which is also why finitely additive probabilities are not referred to as 'measures' here. In Section 4.5 on de Finetti's formalization of probability we will use *probability assignment* to avoid confusion.

Definition 4.4.1. A *finitely additive probability on* $(\Omega, \mathcal{A})$ is a function $P : \mathcal{A} \to [0, 1]$ from an algebra $\mathcal{A}$ over Ω to $[0, 1]$ that satisfies the two axioms:

A1 $P(\Omega) = 1$

A2 for any $A, B \in \mathcal{A}$ such that $A \cap B = \emptyset$, $P(A \cup B) = P(A) + P(B)$.

That is, comparing with Definition 4.3.1 of probability measures, an algebra substitutes the σ-algebra, and the axiom of countable additivity is dropped. Since any σ-algebra is an algebra and any countably additive function is finitely additive, probability measures are finitely additive probabilities, but not vice versa.

It was already seen that in the case of a finite set Ω or a finitely generated algebra on Ω the axiom of countable additivity is not needed, and hence in this case there is no difference between probability measures and finitely additive probabilities.

A countable set Ω is the simplest case where differences arise.[24] A discrete probability measure can be defined on the powerset by assigning values summing to 1 to all $\omega \in \Omega$. There are finitely additive probabilities for which the construction is not so simple. An illustrative example is the uniform distribution on the natural numbers, $\Omega = \mathbb{N}$. A probability measure giving equal probabilities to all natural numbers does not exist: if the value for each natural number were 0, the axiom of countable additivity would imply that $P(\mathbb{N}) = \sum_{n \in \mathbb{N}} 0 = 0 \neq 1$, which violates the norming condition. On the other hand, if the value were any positive ε, any set with more than $1/\varepsilon$ elements would be assigned a probability > 1, which equally violates the norming condition. De Finetti criticizes countable additivity as an axiom for probability theory, for example, by the following story about an individual who wishes to evaluate probabilities for the events in a countable partition.

> Someone tells him that in order to be coherent he can choose the p_i in any way he likes, so long as the sum = 1 (it is the same thing as in the finite case, anyway!)
>
> The same thing?!!! You must be joking, the other will answer. In the finite case, this condition allowed me to choose the probabilities to be all equal, or slightly different or very different; in short, I could express any opinion whatsoever. Here, on the other hand, the *content* of my judgements enter into the picture: I am allowed to express them only if they are unbalanced ... [de Finetti, 1974, p. 123].

[24]The same could be stated more generally for a countable partition of an arbitrary set Ω.

In fact, countable additivity requires probability measures on $\mathbb{N}$ which assign positive probability to all $n \in \mathbb{N}$ to be heavily unbalanced: the sum converging to 1 means that for any $\varepsilon > 0$ one finds a finite number of the n's, which together have a probability $> 1 - \varepsilon$. De Finetti comments in brackets:

> (In such circumstances, I am tempted to say that the events 'are not countably infinite' but 'a finite number – up to trifles'.) [de Finetti, 1974, p. 122].

Contrarily, in the finitely additive setting, one can consider a uniform distribution on $\mathbb{N}$ that assigns all $n \in \mathbb{N}$ the same probability. The probability of each n still has to be zero, however, the infinite sum is allowed to be zero in this case, because only finite additivity is assumed. The uniform distribution assigns to a set $A \subseteq \mathbb{N}$ its "relative frequency", for example, the probability of the even numbers is $\frac{1}{2}$, that of numbers divisible by 3 is $\frac{1}{3}$, and so on. More precisely, the probability of a set A is defined as follows: consider for each $n \in \mathbb{N}$ the set $\{1, \dots n\}$, that is, the first n natural numbers, and let a_n denote the number of elements of A contained in that set. Then, $P(A) = \lim_{n\to\infty}(a_n/n)$.

Especially being used to standard probability theory, it may first seem unintuitive that the positive probability of an infinite set (such as the even numbers) cannot be obtained by summing up the probabilities of its elements. Generally, in a countably infinite Ω, the probability of a countably infinite set $A \subseteq \Omega$ is

$$P(A) \geq \sum_{\omega \in A} P(\{\omega\})$$

while under the condition of countable additivity, equality would hold. However, a similar effect occurs for the case of uncountably infinte sets and countably additive measures: the probability of a set may not necessarily be deduced from those of its elements, it is bounded below by their "sum"[25] but may be strictly greater.

In fact, considering functions defined on a class of subsets of a given set Ω, one can roughly say that these reduce to functions defined on the elements of that set

- both under the finite and the countable additivity condition for finite Ω,
- not under finite but under countable additivity for countably infinite Ω,
- and neither under finite nor under countable additivity for uncountably infinite Ω.

Roughly speaking, the finitely additive setting makes the same difference between finite and infinite sample spaces that the countably additive setting makes between finite and countable sample spaces on the one hand and uncountably infinite ones on the other hand. This difference can also be stated as an answer to the question which kind of unions of events with probability zero may have positive probability. Finite unions of zero probability events always have zero probability for both additivity conditions, while countable unions of zero probability events may have positive probability for finitely additive probabilities but not for probability measures. Finally, uncountable unions of zero probability events may have positive probability in both cases. A critique of the axiom of countable additivity by de Finetti [1974] illustrates an interesting effect of countable additivity. In the context of the uniform distribution on a countable set, one finds the following sentences:

> ... we suspect 'Adhockery for mathematical convenience' – because the distinction between finite and infinite has without doubt a logical and philosophical relevance, whereas it might seem strange to draw the crucial distinction between finite and non-denumerable on the one hand, and countable on the other. [p. 121].

This distinction here is that between existence and non-existence of a uniform distribution, which can be defined under countable additivity for finite and uncountably infinite sets,

[25] Leaving out of consideration the problem of defining uncountable sums here, since the statement is used for intuitive illustration purposes only.

but not for countably additive sets. The questions considered above, which unions of zero events can have positive probability and whether a probability function is completely defined by assigning values to the elements of the sample space, drew a different distinction under countable additivity: finite and countably infinite versus uncountably infinite. This is found in the German translation: "während es eigenartig erscheinen würde, die ganz wesentliche Grenze zwischen endlich und abzählbar einerseits und nicht abzählbar andererseits zu setzen." [de Finetti, 1981, p. 151]. Both distinctions seem more of an adhockery than the simple distinction between finite and infinite. Which one was meant, was unfortunately not found out, because the original was not available.

As illustrated by the above example, a finitely additive probability is not fully determined by assigning values to the elements of Ω in the case where Ω is infinite. Assigning zero probability to all elements n of $\mathbb{N}$, one can conclude that the probability of any finite subset of $\mathbb{N}$ is zero, but no restrictions follow for infinite sets. This is similar for continuous probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, saying that each singleton set has probability zero does not give any further information about the probability measure, this statement is already contained in the qualifier "continuous". According to Theorem 4.3.3, these probability measures are fully described by distribution functions. This is not the case under finite additivity, as extensively discussed by de Finetti [1974, Chapter 6]. This is one of the points where mathematical complications arise for finitely additive probabilities as compared to probability measures. Convenient tools for probability measures do not fully describe a finitely additive probability. A corresponding tool for these does not seem to be available.

Further mathematical complications arise with respect to integration. Standard probability theory, being based on measure theory, can and does make use of Lebesgue's integral, yielding a unique well-defined integral value for all bounded measurable functions. For integration in the finitely additive setting, there is no correspondingly simply defined class of functions which yields a definite integral value. De Finetti claims that "the case for consigning the Riemann integral to the attic now that the Lebesgue integral is available has not itself been proved." [de Finetti, 1974, p. 231]. In many cases where the Riemann integral gives only an upper and a lower bound, the Lebesgue integral provides a well-defined value. De Finetti [1974] criticizes "a preconceived preference for that which yields a unique and elegant answer *even when the exact answer should instead be 'any value lying between these limits'*" [p. 231] and argues that the indetermincay of the Riemann integral "may very well have an essential meaning" [p. 231]. Integration with respect to finitely additive probabilities is considered in Section 6.4.

As stated in Section 3.2, countably additive measures cannot be extended to all subsets of an infinite set unless they are discrete measures. The domain of definition of measures is therefore restricted to a suitable σ-algebra of measureable sets. This measurability condition for sets and functions is not perceived as a problem in applications, in fact, Williams [1991] states that "although not all sets are measurable, it is always the case for probability theory that enough sets are measurable" [p. 15]. De Finetti's statement

> Never mind, it might be argued: measurable sets will suffice. But from what point of view? Practically speaking, the intervals themselves were perhaps sufficient. From a theoretical standpoint, however, is there any justification for this discrimination between sets of different *status*: the orthodox which we are permitted to consider, and the heretical which must be avoided at all costs? [de Finetti, 1974, p. 230]

could be seen as a direct response to this.

Unlike probability measures, finitely additive probabilities are extendable to the whole powerset of *any* set. While the "convenience" of countable additivity is often pointed out, this is one of the points where finite additivity proves to be more convenient. The extendability result and its discussion are delayed to Section 4.5.3.

Countable additivity can be viewed as a form of continuity, due to the following result. For a proof, see for example, Jacod and Protter [2000, p. 7].

Theorem 4.4.2. *For a finitely additive probability $P : \mathcal{F} \longrightarrow [0,1]$ defined on a σ-algebra $\mathcal{F}$, the following are equivalent.*

1. *The axiom of countable additivity, see Definition 4.3.1 (iii).*
2. *If $A_n \in \mathcal{F}$ and $A_n \downarrow \emptyset$, then $P(A_n) \downarrow 0$.*
3. *If $A_n \in \mathcal{F}$ and $A_n \downarrow A$, then $P(A_n) \downarrow P(A)$.*
4. *If $A_n \in \mathcal{F}$ and $A_n \uparrow \Omega$, then $P(A_n) \uparrow 1$.*
5. *If $A_n \in \mathcal{F}$ and $A_n \uparrow A$, then $P(A_n) \uparrow A$.*

where the notation $A_n \downarrow A$ means that $A_{n+1} \subseteq A_n$ for all $n \in \mathbb{N}$, and $\cap_{n=1}^{\infty} A_n = A$, and similarly, $A_n \uparrow A$ means $A_n \subseteq A_{n+1}$ for all $n \in \mathbb{N}$, and $\cup_{n=1}^{\infty} A_n = A$.

That is, for probability measures, one may take limits which for finitely additive probabilities one cannot always take. Finitely additive probabilities are, however, "mathematically convenient" in a different continuity question. Limits of finitely additive probabilities are finitely additive probabilities, while limits of probability measures need not be probability measures (they are still finitely additive probabilities, of course). As discrete masses can become diffuse in the limit, countable additivity may not be preserved in this limit. An example is given by a sequence of probability measures assigning a probability of $\frac{1}{n}$ to each of the first n numbers in $\mathbb{N}$. The limit of these uniform distributions on finite sets is the above discussed uniform distribution on the natural numbers, which was seen not to qualify as a probability measure. Roughly speaking, for probability measures one may take limits at the level of events, or "within" the measure, while for finitely additive probabilities one may take limits at the level of the probability distribution itself.

Further mathematical differences are discussed in the literature, but are beyond the scope of this work. As some references, we mention de Finetti [1972] and Seidenfeld [2001], see also references in the latter.

Finally, countable additivity has one big advantage in applications: it is established simply by what is taught. As an example, Markov chains and processes in the σ-additive setting are standard in any basic probability course and thus textbook. On finitely additive Markov chains there are some works [e.g. Ramakrishnan, 1980, 1981], but these cannot be considered basic probability knowledge. Having extensively quoted de Finetti already, let us now come to his formalization of probability.

4.5 De Finetti's probability

This section presents a simplified version, adapted to the needs of this work, of the mathematical model of probability proposed by de Finetti [1974][26].

De Finetti distinguishes

> theories which [...] axiomatize the formal properties of probability as a real-valued function defined on entities called "events", without reference to any concrete interpretation, [from theories] that attempt to express axiomatically particular concrete interpretations of probability. [de Finetti, 1972, p. 68].

While Kolmogorov's axioms belong to the first group, de Finetti's formalization of probability is of the second kind: he formalizes the subjective interpretation of probability. This considers probability not as a property of some process in the real world such as coin tossing, but as a subjective estimate in a situation of uncertainty. (Subjective) probability is defined as "the *degree of belief* in the occurence of an event attributed by a given person at a given instant and with a given set of information." [de Finetti, 1974, p. 3]. It always refers to a

[26] This book sums up de Finetti's work on probability, published in many works from the late 1920s on.

fixed (real or fictitious) reference person. This person is traditionally referred to as 'You' in subjective probability texts.

De Finetti's formalization provides an answer to the question "where the probabilities come from". In fact, the author gives an operational definition of probability, a rule how to measure the probabilities assigned to events by a reference person. He further stresses at every opportunity that

> To me, mathematics is an instrument which should conform itself strictly to the exigencies of the field in which it is to be applied. One cannot impose, for his own convenience, axioms not required for essential reasons, or actually in conflict with them. [de Finetti, 1974, p. 3].

4.5.1 From possibility to probability, and events

Before considering the concept probability, de Finetti [1974] focusses on the "preliminary logic of certainty", emphasizing that a clear distinction between possibility and probability is fundamental. In this logic,

> there exist only:
> TRUE and FALSE as final answers;
> CERTAIN, IMPOSSIBLE and POSSIBLE as alternatives, with respect to the present knowledge of each individual. [p. 11].

While according to this state of knowledge, several alternatives may be possible, the underlying idea is that exactly one of the possibilities will turn out to be true. The term random is used by de Finetti in this sense as well:

> the meaning that we give to 'random' [...] is simply that of 'not known' (for You), and consequently 'uncertain' (for You), but *well-determined* in itself [...,] it must be specified in such a way that a possible bet (or insurance) based upon it can be decided without question. [de Finetti, 1974, p. 28].

A *random quantity* has "instead of a unique *certain* value, two, or several, or infinitely many *possible* values" [p. 12], depending on the state of ignorance of the individual for whom the quantity is unknown. By equating "true" with 1 and "false" with 0, de Finetti can consider *random events* as particular random quantities with only these two possible values. The notion of 'possible' constitutes "a third attitude" toward an event, but not a third truth value, as de Finetti [1995] emphasizes: "This third attitude does not correspond to a distinct third value of yes or of no, but simply to a doubt between yes or no ...." [p. 182].

For the important concept of an *event*, we will depart from de Finetti's model. We denote events by E and E_1, E_2 etc. Considering a set of events, denoted $\mathcal{E}$, the "state of information" of the reference person may include some knowledge about relations between some events, for example that an event $E_1 \in \mathcal{E}$ implies another event $E_2 \in \mathcal{E}$, which means that if the former is true, also the latter is.

To represent events mathematically,

> we have to assume that a meaning has been assigned to the sure and the impossible events, to the negation of an event [...], to the logical product or intersection of two events [...], and therefore to their union [...], either by equating "events" and "propositions" or, at any rate, by assuming analogous formal properties for the events as abstract entities. The task of explicating the notions of event and of logical operations on events would then be left to theories dealing with the meaning of probability. [de Finetti, 1972, p. 70].

De Finetti (as many others working with subjective probability) prefers equating events and propositions rather than representing events by sets as seen in Kolmogorov's model. The elements of Ω, the elementary events, are artificial concepts in many cases. However, without this representation, the mathematical model is generally not made completely explicit, neither for propositions nor for representing the state of information of the reference person.

While in propositional logic, propositions are defined as well-formed formulae over a fixed set of atomic propositions, here, a proposition is loosely defined as a statement that can be true or false. An attempt of describing propositions formally as in propositional logic within the mathematical model of probability turned out to overly complicate matters.[27] Also, while it is sometimes stated that two propositions may describe the same event [see e.g. Coletti and Scozzafava, 2002, p. 17][28], for example "today is Wednesday" and "tomorrow is Thursday", it is generally not made explicit how these are to be treated when "equating events and propositions", for example, that propositions describing the same event should be assigned the same probability. It is considered intuitively clear what propositions are and how they represent events.

However, even when considering propositions as events, a set representation can be used as a preliminary step to make explicit the state of information of the reference person. This is done by mapping propositions to subsets of any convenient but fixed set Ω. Mapping a proposition to Ω means the reference person asserts this proposition to be certain, mapping it to the empty set means asserting it to be impossible. All propositions that are mapped to non-empty strict subsets of Ω are considered possible. The set Ω represents the information given by the "preliminary logic of certainty", cited above: Ω represents all possibilities in a first step. The idea that "exactly one of the possibilities will turn out to be true" is represented by considering one point $\omega \in \Omega$ as the final state in a second step. Two propositions which describe the same event map to the same subset. The mapping of propositions to subsets also represents the fact that it depends on the state of information whether two propositions describe the same event. A different state of information can be represented by a different map. Another arbitrary example: the propositions "today is the first Wednesday of November" and "today is November fifth" were the same event given the state of information that "we are in 2008"; with the state of information "we are in 2009" they are not. Defining the probability on sets, one has clearly separated the "preliminary" description of the state of information of the reference person from the statement of probability values.

To make a long story short, we represent events and the given preliminary state of information of a reference person as subsets of a fixed set Ω. A given set of events $\mathcal{E}$ will in the following be a set of subsets of this Ω. As for Kolmogorov's formalization of probability, an event hence turns out to be true if the "outcome" ω is contained in the subset representing the event, and false if not, *the certain event* is Ω itself, *the impossible event* is the empty set.

Having a mathematical representation of possibilities, the concept probability can be introduced. Wheras "the possible has no gradation [. . . ,] probability is an additional notion that one applies within the range of possibility, thus giving rise to gradations ('more or less probable')" [p. 27]. These gradations are present in everyday life:

> To any event in which we have an interest, we are accustomed to attributing, perhaps vaguely and unconsciously, a probability: if we are sufficiently interested we may try to evaluate it with some care. [de Finetti, 1974, p. 12].

[27] I thank the participants of the Cartesian Seminar at PIK in autumn 2008 for working with me on this topic. Just a sketch of the attempt here: By defining propositions as in propositional logic, the description of the "state of knowledge" becomes more complicated than when represented by sets. A partial valuation function can be used to represent the state of knowledge, that is some relations between some events. For example, if an event E_1 implies an event E_2, the valuation function v should assigning the value 'true' to the implication $E_1 \Rightarrow E_2$. A representation of "the certain event" and "the impossible event" is less immediate here than when working with sets. Next, probability can be considered a second valuation function, assigning values from the interval $[0, 1]$ to events. However, in this formulation, extra conditions for the probability function have to be stated, such as: if $v(E_1 \Leftrightarrow E_2) = true$, then the probability assigned to E_1 and E_2 needs to be equal.

[28] In this book, for example, it is emphasized that propositions, not subsets are considered as events, but then set theoretical standard notation is used without much further comment, for example $E_1 \subset E_2$ for "E_1 implies E_2".

4.5.2 Betting and coherence

De Finetti proposes an operational definition of probability based on a betting scheme that has the purpose of evaluating probabilities, that is, finding out the numbers assigned by the reference person to the event in a set $\mathcal{E}$. Actually, the original betting scheme is more general than the one presented here: random quantities are considered instead of events. What this scheme elicits is the 'prevision' assigned to the random quantities under consideration by the reference person, corresponding to the expectation under the subjective probability. The mathematical description is more involved, for example, suprema are necessary instead of the maxima used below, because of the possibly infinitely many possible values of a random quantity, whereas an event has only two possible values. In the case of events, prevision reduces to probability and only this case will be treated here.

De Finetti uses a fictitious betting situation as a simple situation of "decision-making under conditions of uncertainty" because

> It is precisely in investigating the connection that must hold between evaluations of probability and decision-making under conditions of uncertainty that one can arrive at criteria for measuring probabilities, for establishing the conditions which they must satisfy, and for understanding the way in which one can, and indeed must, *'reason about them'*. [de Finetti, 1974, p. 72].

Here, "must reason" is to be understood as a rationality criterion: an assignment of numbers to events is not considered "a probability assignment" if it does not satisfy a *coherence* condition. De Finetti describes the nature of this condition as follows:

> Concerning a known evaluation of probability, over any set of events whatsoever, and interpretable as the opinion of an individual, real or hypothetical, we can only judge whether it is *coherent*. If it is not, the evaluator, when made aware of it, should modify it in order to make it coherent. In the same way, if someone claimed to have measured the sides and area of a rectangle and found 3 metres, 5 metres and 12 square metres, we, even without being entitled, or having the inclination, to enter into the merits of the question, or to discuss the individual measurements, would draw his attention to the fact that at least one of them is wrong, since it is not true that $3 \times 5 = 12$. [de Finetti, 1974, p. 9].

The rationality criterion is expressed for the betting situation, and serves to avoid bets with consequences which are "manifestly undesirable (leading to certain loss)" [de Finetti, 1974, p. 72]. From this restriction on the "real-world" decision side, mathematical consequences for probability assignments are derived.

The betting situation is first explained for a single event.

> Given a random event E, You are obliged to choose a value $\alpha(E)$, on the understanding that, after making this choice, You are committed to accepting any bet whatsoever with gain $c(\mathbb{1}_E(\omega) - \alpha(E))$, where $c \in \mathbb{R}$ is arbitrary (positive or negative) and at the choice of an opponent. [de Finetti, 1974, p. 87, adapted].

This means the reference person pays $c \cdot \alpha(E)$ to gain c if the event E is true, and to gain nothing if E is false. The value c is referred to as the *stake*. If c is negative, bookmaker and bettor switch roles: the reference person gets $|c \cdot \alpha(E)|$ from the opponent, having to pay back $|c|$ if E is true, and nothing otherwise.

In real-world decision making where phenomena such as risk aversion play a role, the unit of this gain would be important, however, de Finetti prefers to disregard this aspect for simplicity in the study of probability. He refers the reader to the works by Ramsey [1926] and Savage [1954] and assumes the amounts of money involved in the bets to be sufficiently small to avoid the problem. Thus,

> accepting as practically valid [...] the identity of monetary value and utility within the limits of 'everyday affairs', [i.e.] transactions whose outcome has no relevant effect on the fortune of an individual, [de Finetti, 1974, p. 82]

he chooses for "setting aside [...] the notion of utility, in order to develop in a more manageable way the study of probability" [p. 80].[29]

The coherence condition only becomes relevant when several events are considered simultaneously. Consider a set of events $\mathcal{E}$, that is, a set of subsets of some Ω, representing the state of information of the reference person. The betting situation is extended from one to several bets as follows: the reference person is asked to state probabilities for all events in $\mathcal{E}$, that is, to specify a function[30] $\alpha : \mathcal{E} \to [0, 1]$. Then the opponent may choose a combination of a finite number of bets to be played simultaneously, that is, the same "final value" is observed to decide all bets together. Mathematically, the opponent chooses a number n of events and stakes $c_1, \ldots, c_n$ to be played on these events. The reference person's gain $G : \Omega \to \mathbb{R}$ for such a combination of bets is

$$G = c_1(\mathbb{1}_{E_1} - \alpha(E_1)) + \ldots + c_n(\mathbb{1}_{E_n} - \alpha(E_n)) \tag{4.3}$$

for one 'experiment', one verification of "the true ω".

Coherence is the assumption "that You do not wish to lay down bets which will with *certainty* result in a loss for You", corresponding to the "manifestly undesirable consequences" mentioned above. Given the extended betting scheme, this means a probability assignment α is *coherent* if it does not allow a choice of $n \in \mathbb{N}$ and $c_1, \ldots c_n \in \mathbb{R}$ such that for all $\omega \in \Omega$, the gain $G(\omega)$ defined in formula (4.3) is negative[31] or, in other "words",

$$\max_{\omega\in\Omega}\{c_1(\mathbb{1}_{E_1}(\omega) - \alpha(E_1)) + \ldots + c_n(\mathbb{1}_{E_n}(\omega) - \alpha(E_n))\} = \max_{\omega\in\Omega} G(\omega) < 0.$$

We can sum up into the following (equivalent, only positively formulated)

Definition 4.5.1. Given a set of events $\mathcal{E}$, a function $\alpha : \mathcal{E} \to [0, 1]$ is a *coherent probability assignment*, if for any $n \in \mathbb{N}$, any $c_1, \ldots, c_n \in \mathbb{R}$ and any choice of events $E_1, \ldots, E_n \in \mathcal{E}$,

$$\max_{\omega\in\Omega} \sum_{i=1}^{n} c_i(\mathbb{1}_{E_i}(\omega) - \alpha(E_i)) \geq 0.$$

The operational definition of *the probability (according to You) of the event E* given by de Finetti is the value $\alpha(E)$ chosen (by You) in this setting.

Remark 4.5.2. No structure was assumed for the set of events $\mathcal{E}$. In particular, it need not be an algebra, or even σ-algebra. One can assume an assignment known for an arbitrary set of events. The set $\mathcal{E}$ can be infinite. However, the coherence condition is stated in terms of combinations of finitely many bets. This definition is equivalent to stating the definition for a finite $\mathcal{E}$, with the n in the definition equal to the number of its elements (since all bets with less than n events are included by choosing the corresponding stakes $c = 0$), and then defining coherence for an infinite $\mathcal{E}$ via coherence for any finite subset of $\mathcal{E}$. In the literature this procedure is often found [see Coletti and Scozzafava, 2002; Borkar et al., 2004, for example].

---

[29] In this section, one finds the statement:

> We accept that once we are in the area of mathematical idealization we can leave out of consideration adherence to reality in every tiny detail: on the other hand, it seems rather too unrigorous to act in this way when formulating that very definition which should provide the connection with reality. [p. 80].

While the first part refers to the utility question, where a concession is made for the simplicity of the theory, the second sentence refers to the definition of probability itself, for which he does not accept any axioms chosen for mathematical convenience.

[30] A function with target $\mathbb{R}$ would be sufficient. That the assigned values must lie in $[0, 1]$ follows from the coherence condition as will be shown in the following section.

[31] Here, for example, the formulation in terms of random quantities is more involved: it needs the extra qualification that *uniformly negative* gains are to be avoided and thus the original uses a supremum instead of the maximum. This coherence condition is weaker than others found in the literature, but de Finetti provides explanatory examples argueing why, in certain cases, bets which can only lead to a loss, which is however arbitrarily small, are to be accepted as coherent, see p. 118. Here, since the values of the indicator functions are always 0 or 1, and the sum has finitely many terms, with all other numbers fixed (the c's and the values of α), the supremum is a maximum.

As a consequence of the interpretation of probability as a personal judgement about possibilities, de Finetti does not admit "unknown probabilities". The arbitrary extension of the assignment can be considered as all those events for which probabilities have been evaluated. The idea is that other events have not been taken into consideration. If they are, their probability may not remain unknown. In the betting scheme, when asked about the probability of an event, the answers 'I do not know', 'I am ignorant of what the probability is', 'in my opinion the probability does not exist' do not make sense and are not admitted:

> Probability (or prevision) is not something which in itself can be known or not known: it exists in that it serves to express, in a precise fashion, for each individual, his choice in his given state of ignorance. To imagine a greater degree of ignorance which would justify the refusal to answer would be rather like thinking that in a statistical survey it makes sense to indicate, in addition to those whose sex is unknown, those for whom one does not even know 'whether the sex is unknown or not'. [de Finetti, 1974, p. 84].

4.5.3 Coherent assignments and finitely additive probabilities

Coherent probability assignments are closely related to finitely additive probabilities. Without going into detail for proofs of the general theorems, we rather concentrate on providing the intuition about this relation. To be a *coherent probability assignment*, an assignment α has to satisfy additivity for disjoint events, and $\alpha(E)$ has to be chosen in $[0,1]$ for all $E \in \mathcal{E}$[32]. This can easily be shown by constructing a "certain loss situation" if the conditions are violated. Assume there are in $\mathcal{E}$ two disjoint events E_1 and E_2, and $E_3 = E_1 \cup E_2$, that have been assigned values where $\alpha(E_1) + \alpha(E_2) \neq \alpha(E_3)$. Denoting $\alpha_i = \alpha(E_i)$, suppose $\alpha_1 + \alpha_2 < \alpha_3$, the opposite case can be treated analogously. The opponent can choose a bet of the form

$$c_1(\mathbb{1}_{E_1}(\omega) - \alpha_1) + c_2(\mathbb{1}_{E_2}(\omega) - \alpha_2) + c_3(\mathbb{1}_{E_3}(\omega) - \alpha_3)$$

The possible gains for this bet are:

$$G(\omega) = \begin{cases} c_1(1-\alpha_1) + c_2(0-\alpha_2) + c_3(1-\alpha_3) & \text{if } \omega \in E_1 \\ -c_1\alpha_1 + c_2(1-\alpha_2) + c_3(1-\alpha_3) & \text{if } \omega \in E_2 \\ -c_1\alpha_1 - c_2\alpha_2 - c_3\alpha_3 & \text{if } \omega \in E_3^c \end{cases}$$

If the opponent chooses $c_3 = c$ and $c_1 = c_2 = -c$ for some $c > 0$, this is

$$G(\omega) = \left\{ \begin{array}{ll} -c + c\alpha_1 + c\alpha_2 + c - c\alpha_3 & \text{if } \omega \in E_1 \\ c\alpha_1 - c + c\alpha_2 + c - c\alpha_3 & \text{if } \omega \in E_2 \\ c\alpha_1 + c\alpha_2 - c\alpha_3 & \text{if } \omega \in E_3^c \end{array} \right\} = c(\alpha_1 + \alpha_2 - \alpha_3) < 0$$

that is, the possible gains for this bet are all negative.

Similarly, if some $\alpha(E)$ is chosen outside $[0,1]$, it suffices for the opponent to chose a bet on that event. $c(\mathbb{1}_E(\omega) - \alpha(E))$ has the possible gains $c(1-\alpha(E))$ and $-c\alpha(E)$ depending on whether $\omega \in E$ or not. For negative $\alpha(E)$ these are both negative if c is negative, while if $\alpha(E) > 1$, any choice of $c > 0$ makes all possible gains negative. The condition also implies that the certain event Ω needs to be assigned the probability 1, because any choice of $\alpha(\Omega) < 1$ would allow the opponent to choose any bet $c(\mathbb{1}_\Omega - \alpha(\Omega))$ with $c < 0$ to yield a certainly negative gain $c(1 - \alpha(\Omega))$.

Thus, if $\mathcal{E}$ is an algebra, a coherent probability assignment α automatically is a finitely additive probability on $(\Omega, \mathcal{E})$. The converse can also be shown rather easily: a finitely additive probability on $(\Omega, \mathcal{E})$ (for $\mathcal{E}$ algebra) is a coherent probability assignment. Assume

[32] Here, this condition is anticipated in the definition of α as a function with target $[0,1]$. In the more general case of a random quantity X, the restriction is that $\alpha(X)$ must lie between the infimum and the supremum of the possible values the quantity can take and is hence an actual restriction.

it is not coherent, then there exists a choice of n, $c_1, \ldots, c_n$ and $E_1, \ldots, E_n \in \mathcal{E}$ such that

$$\max_{\omega\in\Omega} \sum_{i=1}^{n} c_i(\mathbb{1}_{E_i}(\omega) - \alpha(E_i)) < 0, \text{ that is,}$$
$$\sum_{i=1}^{n} c_i(\mathbb{1}_{E_i}(\omega) - \alpha(E_i)) < 0 \text{ for all } \omega \in \Omega, \text{ or}$$
$$c_1\mathbb{1}_{E_1}(\omega) + \ldots + c_n\mathbb{1}_{E_n}(\omega) < c_1\alpha(E_1) + \ldots + c_n\alpha(E_n).$$

Consider the constituents of the events $E_1, \ldots, E_n$, say $C_1, \ldots, C_m$. Since any E_i is a disjoint union of some constituents, and the finitely additive probability assigns to each E_i the sum of the probabilities assigned to these constituents, one can rewrite the inequality as

$$c_1^*\mathbb{1}_{C_1}(\omega) + \ldots + c_m^*\mathbb{1}_{C_m}(\omega) < c_1^*\alpha(C_1) + \ldots + c_m^*\alpha(C_m)$$

where the c_i^* are obtained from the c_i by distributivity and sorting according to constituents. This, however, is in contradiction with the assignment being a finitely additive probability: since the constituents partition Ω, the left hand side reduces to one of the values c_i^* for each ω. This cannot be strictly smaller than the right hand side for all ω, because the latter is a weighted average of the values c_i^*. Thus, when defined on an algebra, coherent probability assignments are nothing but finitely additive probabilities.

However, it is an important point in de Finetti's mathematical model of probability that no particular structure for $\mathcal{E}$ is assumed. If $\mathcal{E}$ is not an algebra, the two axioms for finitely additive probabilities alone are not sufficient for coherence. As in the following example[33], the additivity condition may be vacuously satisfied when there are no disjoint events in $\mathcal{E}$ whose union is also in $\mathcal{E}$.

Example 4.5.3. *Let $\{F, G, H\}$ be a partition of Ω. Consider $\mathcal{E} = \{E_1, E_2, E_3, H, \Omega\}$ with $E_1 = F \cup G, E_2 = F \cup H, E_3 = G \cup H$. Consider the assignment $\alpha : \mathcal{E} \cup \Omega \to [0, 1]$ defined by*

$$\alpha(E_1) = \frac{4}{9}, \alpha(E_2) = \frac{4}{9}, \alpha(E_3) = \frac{2}{3}, \alpha(H) = \frac{5}{9}, \alpha(\Omega) = 1.$$

α satisfies the axioms A1 and A2, however, it is not coherent. A bet with certain loss is, for example, $-c(\mathbb{1}_{E_2} - \alpha(E_2)) + c(\mathbb{1}_H - \alpha(H))$ with $c > 0$. Possible gains of this are

$$G(\omega) = \begin{cases} -c(1-\alpha(E_2)) + c(1-\alpha(H)) & = c(\alpha(E_2) - \alpha(H)) < 0 & \text{if } \omega \in H \\ -c(1-\alpha(E_2)) - c\alpha(H) & = c(\alpha(E_2) - \alpha(H) - 1) < 0 & \text{if } \omega \in E_2 \backslash H = F \\ c\alpha(E_2) - c\alpha(H) & = c(\alpha(E_2) - \alpha(H)) < 0 & \text{if } \omega \in E_2^c = G \end{cases}$$

that is, all negative. The assignment would not be intuitively accepted as a probability assignment because while H is a subset of E_2, the latter is assigned a smaller value than the former. As the subset relation represents implication, here, "if H takes place then also E_2 takes place". The degree of belief in E_2 should therefore not be smaller than that in H.

What is however sufficient for coherence, is the condition that the assignment can be extended to a finitely additive probability on an algebra. It can be shown that

Theorem 4.5.4. *For an arbitrary set $\mathcal{E}$ of non-empty subsets of Ω with $\mathfrak{a}(\mathcal{E})$ the algebra generated by $\mathcal{E}$, a function $\alpha : \mathcal{E} \to [0, 1]$ is a coherent probability assignment if and only if there exists a finitely additive probability P on $(\Omega, \mathfrak{a}(\mathcal{E}))$ that agrees with α on $\mathcal{E}$.*

The extension P of α need not be unique. A rather accessible proof can be found in Borkar et al. [2004], who first show the result for a finite $\mathcal{E}$, using the fact[34] that a coherent

[33] Adapted from Coletti and Scozzafava [2002].
[34] They refer the reader to Rockafellar [1970].

probability assignment can be interpreted as a point in the closed convex hull of the possible "points" in terms of the indicators, that is, of the set $\{[\mathbb{1}_{E_1}(\omega), \dots, \mathbb{1}_{E_n}(\omega)] | \omega \in \Omega\}$. Then, this result can be used for the general case, because a coherent assignment for an infinite $\mathcal{E}$ is defined via coherence for any finite subset.

The incoherence of the assignment in example 4.5.3 can easily be illustrated using this coherence condition.

Example 4.5.3 (contd)**.** *Even when deleting H and considering the assignments for the other events only, there is no possibility of extending α to a finitely additive probability P on $\mathfrak{A}(E) = \{\emptyset, F, G, H, E_1, E_2, E_3, \Omega\}$. Such an extension would have to satisfy*

$$P(F) + P(G) = \frac{4}{9}, P(F) + P(H) = \frac{4}{9}, P(G) + P(H) = \frac{2}{3}, P(F) + P(G) + P(H) = 1.$$

But then summing the first three equations and dividing by 2 yields $1 = P(\Omega) = P(F) + P(G) + P(H) = \frac{7}{9} < 1$.

In de Finetti [1974], the result of the theorem follows among other things from the *fundamental theorem of probability* [adapted, p. 112]:

Theorem 4.5.5. *Given the probabilities $\alpha(E_i)$ for $i = 1, \dots, n$ of a finite number of events, the probability $\alpha(E)$, of a further event, E, either,*

(a) turns out to be determined (whatever α is) if E is linearly dependent on the E_i (that is, E can be written as a union of some constituents of the E_i); or

(b) can be assigned, coherently, any value in a closed interval $p' \leq \alpha(E) \leq p''$ (which can often give an illusory restriction, if $p' = 0$ and $p'' = 1$, or even, in limit cases for particular α, give a well-determined result $p' = \alpha(E) = p''$). [...]

Next, de Finetti treats extensions of coherent probability assignments to an infinite (countable or uncountable) number of events. Via transfinite induction[35], he shows that a coherent probability assignment on a given set of events $\mathcal{E}$ can be extended to any set (of subsets of Ω) containing $\mathcal{E}$, maintaining coherence. In particular, it can thus be extended to the generated algebra. Together with the above observation that if defined on an algebra a coherent probability assignment is a finitely additive probability, this implies the extendability of a coherent probability assignment on $\mathcal{E}$ to a finitely additive probability defined on $\mathfrak{a}(\mathcal{E})$, as stated in Theorem 4.5.4.

This possibility of extension is an important difference between coherent probability assignments or finitely additive probabilities on the one hand and countably additive probability measures on the other hand as stated in Section 4.4. Therefore, let us provide a flavour of the idea in the proof via transfinite induction [see the Appendix in de Finetti, 1974, p. 335–338]. The important point is that the conditions of coherence always refer to finite subsets. The assignment is extended event by event. Making an infinite number of steps is not of importance in the setting of coherent probability assignments: if a contradiction to coherence arose for the step of a number that has no antecedent, such as the first ordinal which comes after the natural numbers, "a contradiction would derive from the comparison between two *finite* linear combinations" [de Finetti, 1974, p. 337]. Therefore, the contradiction would already have occured at the finite step involving all events in the linear combinations that caused trouble.

For comparison, de Finetti [1974] gives the example of Lebesgue measure:

> Lebesgue measure, too, can be extended, preserving countable additivity, to an arbitrary non-measurable set, and hence to an arbitrary number of such sets, one at a time. In this case, however, an infinite number of steps can lead to a contradiction without any single step doing so (in the same way as a convergent series [converging to any value $\neq 1$] remains such, if we replace the 1st, 2nd, 3rd,... terms with 1, and so on for any finite number, but not if we replace an infinite number of terms). [p. 337]

[35] "... assuming, of course, the Zermelo Postulate" to well-order the random quantities under consideration. [p. 337].

The possibility of extending any coherent probability assignment to the whole powerset of course entails the possibility of extending any finitely additive probability to the powerset of the set its algebra lives on, mentioned as a mathematical difference between finite and countable additivity in Section 4.4.

An example from the literature that uses this fact is "How to gamble if you must" by Dubins and Savage. They define:

> A *gamble* is, of course, a probability measure γ on subsets of fortunes. In the tradition of recent decades, such a measure would be defined only on a sigma-field if subsets of [the set of fortunes] F and required to be countably additive on that sigma-field.
> If this tradition were followed in this book, tedious technical measurability difficulties would beset the theory from the outset. [...] Experience and reflection have led us to depart from tradition and to assume that each γ is defined for all subsets of F and not to assume that γ is necessarily countably additive. [Dubins and Savage, 1965, p. 8]

To deal with gambles that are defined only for a subclass of $\mathfrak{P}(F)$, rather than to "carry a concept of measurability and integrability throughout the discussion", they decide

> to consider all extensions of each such incompletely defined gamble to the class of all sets of fortunes. According to the Hahn-Banach theorem, such extensions exist in abundance, though in a very nonconstructive sense. If, for example, a gambler starting from \$ 1,000 can reach \$10,000 with probability .07 in every completion of an originally incompletely defined problem, is it not a sensible interpretation to credit him with at least that much in connection with the problem as originally specified? Likewise, if there is something he cannot acchieve (or approach) under any extension, it ought not to be regarded as achievable in the original problem. Finally, if something can be approached for some extensions and not others, then the original problem must be recognized as not sufficiently specified to yield a definite answer. [Dubins and Savage, 1965, p. 9].

Thus, while measure theory is often credited with "mathematical convenience", here measurability is perceived as complicating matters. The first monad of finitely additive probabilities presented in Chapter 6 will assume finitely additive probabilities to be defined on whole powersets in this spirit. In that context it will also be seen that measurability questions complicate matters, especially when a less powerful concept than true measurability is used. Before concluding this section on de Finetti's mathematical model of probability with remarks on formalization, a short description of a computable coherence condition will be included.

4.5.4 Coherence computable

While the previously given definitions of coherence are not computable, Coletti and Scozzafava [2002] provide a computable condition equivalent to coherence. We shall here consider only the finite case, that is, a finite set of events $\mathcal{E}$, for simplicity.

Consider the special case where the sets in $\mathcal{E}$ are a partition of Ω and recall that in this case a finitely additive probability is specified by assigning numbers adding to 1 to the sets in the partition (Remark 4.3.2). Here, a coherent assignment is automatically already extended to the whole powerset. For coherence of the assignment, the values for the E_i have to add up to 1, and this can easily be checked.

For general $\mathcal{E}$, the idea is to find a partition that is related to $\mathcal{E}$, which allows a similar check. It is given by the constituents from Definition 3.1.5. In fact, the coherence condition which can be checked computationally combines Remarks 4.3.2 and 3.1.6: coherence amounts to the existence of an assigment for the constituents which adds up to 1 and is compatible with the given assignment on $\mathcal{E}$.

Coherence of an assignment α is thus equivalent to the solvability of the following system of equations:

$$\begin{cases} \sum_{r:C_r\subseteq E_i} x_r = \alpha(E_i) & i \in \{1,\ldots,n\} \\ \sum_{r=1}^{m} x_r = 1 & \\ x_r \geq 0 & r \in \{1,\ldots,m\}. \end{cases}$$

Any solution can be interpreted as an assignment of probabilities to the constituents: $x_r = P(C_r)$ is the probability assigned to the r-th constituent. Given the last equation together with the inequalities for non-negativity, this determines a finitely additive probability P on $\mathfrak{a}(\mathcal{C})$. The first n equations ensure the compatibility with α: by axiom A2, the probability assigned to $E_i \in \mathcal{E}$ is the sum of the probabilities assigned to the constituents whose union E_i is. Since $\mathfrak{a}(\mathcal{C}) = \mathfrak{a}(\mathcal{E})$ (Remark 3.1.6), the finitely additive probability P determined by the solution of the system is an extension of α.

Example 4.5.6. *In the assignment from Example 4.5.3, replace $\alpha(E_2) = \frac{8}{9}$ for the $\frac{4}{9}$ assigned before. Then α is coherent. The constituents are F, G and H (with probabilities denoted x_F, x_G and x_H), and the system is*

$$\begin{cases} x_F + x_G = P(E_1) = 4/9 & \\ x_F + x_H = P(E_2) = 8/9 & \\ x_G + x_H = P(E_3) = 2/3 & \\ x_F + x_G + x_H = 1 & \\ x_r \geq 0 & r \in \{F,G,H\} \end{cases}$$

The system is solved by $x_F = \frac{3}{9}, x_G = \frac{1}{9}$ and $x_H = \frac{5}{9}$ and the solution is unique.

The solution of the system corresponding to a coherent assignment need not be unique, confirming the non-uniqueness of the extension mentioned above. In general, a coherent assignment on $\mathcal{E}$ is the restriction of many finitely additive probabilities defined on $\mathfrak{a}(\mathcal{E})$.

Remark 4.5.7. In particular, if the number of constituents is maximal, i.e. $m = 2^n$, then any assignment on $\mathcal{E}$ is coherent because there are no restrictions for distributing a probability of 1 to the constituents.

4.5.5 Remarks on formalization

De Finetti defines subjective probability (theoretically) as the degree of belief which a certain individual assigns to some events under a given state of information. Then, he provides an operational definition in terms of a betting scheme, that is, a simple situation of decision making under uncertainty. Degrees of belief that would lead to a "manifestly undesirable outcome", a certain loss, are not accepted as coherent probability assignments. The betting scheme allows to, at least in principle, measure the degrees of belief of a person. There is work addressing the question whether the values elicited by the betting scheme actually are the personal probabilities, and answering negatively, such as Seidenfeld et al. [1990]. Also, de Finetti has in later works used a penalty criterion instead of the betting scheme [see, e.g. de Finetti, 2008]. However, this topic is beyond the scope of this work.

The definition of a coherent probability assignment (Definition 4.5.1) is more general than that of a probability measure given by Kolmogorov: any probability measure is a coherent probability assignment, the converse is not true. The definition of coherent probability assignments is also more general than that of finitely additive probabilities, because no restrictions are made for the domain of definition, $\mathcal{E}$.

The different mathematical models of probability arise from different goals pursued with the formalization. In a chapter "On the Axiomatization of Probability Theory" de Finetti [1972] summarizes different purposes of axiomatizing probability theory. The purpose of "listing explicitly the formal properties which are taken as the starting point for a complete and rigorous construction and development of probability theory considered as a formal theory" is the one Kolmogorov's model can be ascribed to. The formalization seen here serves a different purpose: "to attain greater precision in a conceptual discussion" [de Finetti, 1972, p. 67].

De Finetti considers the subjective interpretation of probability, that is, attaches the meaning of 'degree of belief' to the term probability, and wants to express this mathematically, introducing axioms parsimoniously. This can be seen for example when he states that the coherence condition used should be

> *the weakest one* if we want it to be the strongest one in terms of absolute validity. In fact, *it must only exclude the absolutely inadmissible evaluations;* i.e. those that one cannot help but judge contradictory ... [de Finetti, 1974, p. 10].

From the meaning of the concept probability, he does not arrive at the axiom of countable additivity. On the contrary, he has often argued against this axiom, whose "success owes much to the mathematical convenience of making the calculus of probability merely a translation of modern measure theory", while "no-one has given a real justification of countable additivity (other than just taking it as a 'natural extension' of finite additivity)." [de Finetti, 1974, p. 119]. De Finetti considers countably additive probability measures an important special case, and frequently uses the example of continuous functions, compared to all function, as a comparably important special case. This comparison is warranted by the fact that given the axioms of finitely additive probability, the extra assumption of countable additivity is equivalent to assuming the continuity conditions stated in Section 4.4.

Kolmogorov's model of probability, supposed to build the foundations of a mathematical theory, considers the concept probability a primitive. Kolmogorov does not discuss its meaning, nor how to measure it. He uses an additional axiom that allows to draw more conclusions in the mathematical setting. When making an interpretation of the concept explicit, Kolmogorov gives a definition based on repeatable events, that will be similarly encountered in Section 4.6.1 on the frequentist interpretation.

In short, one of the mathematical models is not supposed to clarify the meaning of the concept and the other one translates just one interpretation into mathematics. Before engaging into the dicussion of formalization based on the example of the concept probability, let us briefly present some other interpretations in the following section.

4.6 Interpretations of probability

This section sketches some interpretations of probability which have been described and discussed in the literature. As stated in Section 2.2, interpretations of a concept are closely linked to operational definitions, in fact, the interpretations of probability can also be distinguished as different assessment approaches. Different interpretations may arise from different aspects of a concept being emphasized. The duality of the concept pointed out by Hacking [1975] will play an important role here.

"Classical probability" has been considered an interpretation of the concept, as seen in the quote by Diaconis on page 39. This interpretation is based on an operational definition that allows to measure numerical probability values in the simple case where a finite set of possibilities is considered and these are (for whatever reason) judged to be equivalent. In this simple setting, the notion 'equally probable' is taken as intuitively clear, it is a primitive. The operational definition makes the intuition behind the theoretical definition, a degree of certainty, precise. What is today called a Fundamental Probability Set arose from assumptions such as the principle of insufficient reason, thus focusing on the belief aspect of probability, and was confirmed by long sequences of observations in dice games, fostering the assumption that the die falls on each of the sides equally easily. Even in cases where this seems much less obvious, Fundamental Probability Sets were applied:

> If we consider only games it is natural to have a theory founded on a set of equally likely alternatives. But as soon as we get real life statistics, we expect the model to be inadequate. However, the statistics actually available chiefly concern the mortality

> curve. Far from displacing the model of equal chances they actually confirm it. Thus by what one is tempted to call an accident, both *a priori* and *a posteriori* considerations made the model of a Fundamental Probability Set of equal chance definitive of probability. [Hacking, 1975, p. 121].

For the belief aspect of the concept we should note that in the early times of probability theory, the question *who* was expressing this belief was not considered. The classical probabilists from Pascal to Laplace had a deterministic world view. Uncertainty was considered to derive only from incomplete knowledge, while nothing would be uncertain for "Laplace's demon". An intellect that at a certain point in time knew the exact positions of all items and all forces in the universe and could analyse these, would not be uncertain about anything, past, present or future. However, incomplete knowledge here is to be understood as the knowledge of humanity as a whole, what single people knew or did not know in specific situations was not an issue. The belief aspect in the classical interpretation is also discussed by Hacking [1975] in a chapter on "probability and the law" about the work of Leibniz. Hacking points out that the study of conditional rights leads Leibniz "to a theory of what we might now call 'partial implication'." [p. 87]. Leibniz's work *De conditionibus* shows similarities to (conditional) probability theory, but we will not go into detail here.

As soon as one leaves the simple setting of finitely many equivalent possibilities, the question how to assign numbers to probabilities is not answered by the mathematical model itself. Here, different interpretations of probability arise, together with different assessment approaches and operational definitions. A basic distinction is made between "objective" and 'subjective' probability approaches. While, as was seen in section 4.5, subjective probability is considered the degree of belief of a person, 'objective' interpretations "locate probability in the world" [Hàjek, 2007, p. 20]. De Finetti [1972] describes these as "theories which hold that the value of probability is uniquely determined by objective conditions of a logical, statistical, physical or any other nature" [p. 84]. This attitude was seen in the quote from Gnedenko's work on page 39, where the notion of probability is said to have an "objective role as a numerical characteristic of definite real phenomena".

Objective interpretations can be further subdivided into logical interpretations, propensity interpretations and frequentist interpretations. The first two shall be mentioned without going into any detail here. The logical interpretation views probability as a generalization of deductive logic. It has been studied systematically for example by Carnap. The propensity interpretation considers probability as a physical propensity that certain outcomes are produced in experiments. A name connected to this interpretation is Popper. The interpretation has been accused of "giving empty accounts of probability" and being "metaphysical rather than scientific" [Hàjek, 2007]. Since the frequentist interpretation is most commonly used, it will be sketched in an extra section.

4.6.1 Frequentist probability

The frequency aspect of the concept probability is the focus of frequentist interpretations of probability, which can again be further subdivided into an *empirical* (or finite frequentist) and an *asymptotic* (or infinite frequentist) probability interpretation. The underlying idea, seen in the long quote by Bernoulli on page 41, is to identify the probability of an event with its frequency in a sequence of independent trials.

In the empirical approach, long but finite sequences are considered, and the probability of an event is operationally defined as the number of occurences divided by the number of trials. For example, Venn, when discussing the proportion of births of males and females, states "probability *is* nothing but that proportion"[36]. It is not made explicit how many observations make a sequence a "long sequence", but the importance of this qualification is pointed out already by Bernoulli: "some natural instinct" makes one "feel sure" that the numbers obtained improve with the length of the sequence used (again, see the quote

[36] See Hàjek [2007].

on page 41). The empirical operational definition of probability shows a similarity to the classical one in that cases are counted, however, here the counting is done in a sequence of actual outcomes instead of an abstract space of possibilities.

Hajek lists some problems of this definition, for example, the "problem of the single case" that in a sequence of length one, an event can only have probability 0 or 1. Generally, in a sequence of length n, relative frequencies occur only in steps of $1/n$, which for example "rules out irrational probabilities; yet our best physical theories say otherwise." [Hàjek, 2007]. Also, a sequence could be considered as one trial in an experiment producing sequences. Then, the sequence that has occured would always have probability 1, contrary to the intention with which this operational definition is used.

The empirical approach is used not only for sequences of repeated trials, but also for sets. For example, sets of people who have or do not have a given disease are considered to determine the probability of a person having this disease. Here, the set of people used to evaluate the probability is referred to as the "reference class". It implies the so called "reference class problem": the probability depends on the reference class chosen. In the disease example anything form 'the population of the whole world' to 'people from a certain area, of a certain age and sex, etc.' could be chosen. This illustrates that the assumption of objective probabilities existing "in the world" is problematic: which reference class gives this one objective probability, or, at least, gets close to it? Kolmogorov's hypothesis that a definite probability exists seems easily refutable by this reference class problem in many cases.

For measuring probability in the empirical sense there further is the following

> ... problem: if one stops at a finite number of experiments, it is not entirely correct to say that the frequency one obtains is an approximation of the "true probability" in the same sense as in a measurement of a length with a certain degree of precision. In fact, [...] the frequency is close to the probability only *with a certain probability* which again depends on the "true probability". [Boldrighini, 2009, p. 256],

as can be deduced within the mathematical model of standard probability theory (considering the "true probability" given) from the laws of large numbers. Moreover, Hàjek [2007] notes that, similarly to "ill-calibrated thermometers", one wants to allow for the case where observed frequencies produce misleading results [see p. 16], and that, contradicting an intuition about measurement,

> according to the finite frequentist, a coin that is never tossed, and thus yields no actual outcomes whatsoever, lacks a probability for heads altogether; yet a coin that is never measured does not thereby lack a diameter. [p. 16]

Distinguishing 'counting favourable outcomes in a given sequence' and 'counting favourable cases in a given set' as two operational definitions, together these can be considered one assessment approach, due to the similarity of methods used.

In the asymptotic approach, probability is defined as the limiting frequency in an infinite sequence, thus requiring that the trials may be repeated an infinite number of times. Here, as for the empirical case, the reference class problem occurs. There further is a "reference sequence problem": even sticking to the same sequence, reordering may change the limit. Probabilities operationally defined as limits cannot actually be measured, because a limit cannot be observed in any (however long) finite sequence. In fact, finite sequences do not provide any constraint for the limit. A formalization of this interpretation was given by von Mises, however, it is basically out of use.

4.6.2 Subjective probability

Section 4.5 on de Finetti's mathematical model of probability has shown one example of a subjective interpretation, together with an operational definition, the betting scheme. The subjective probability interpretation gives up the idea of a 'true probability', as for example Diaconis [2007c] expressivley states: "Coins don't have probabilities, people have probabilities."

Other operational definitions are betting schemes with different detailed assumptions. For example, some authors use a regularity condition which entails that only impossible events (mathematically the empty set) may be assigned the probability zero. De Finetti argues against this choice[37] but we will not go into the details here. One also finds cases in the literature where a mathematical model using countable additivity is proposed together with a subjective interpretation. Jeffrey [2004] provides a betting scheme argument for countable additivity. The coherence condition expressed by bets is referred to as a 'Dutch Book argument' in the literature. Hàjek [2007] discusses limitations and problems of measuring probabilities by betting schemes.

> For example, you may have reason to misrepresent your true opinion, or to feign having opinions that in fact you lack, by making the relevant bets (perhaps to exploit an incoherence in someone else's betting prices). Moreover, as Ramsey points out, placing the very bet may alter your state of opinion. Trivially, it does so regarding matters involving the bet itself (e.g., you suddenly increase your probability that you have just placed a bet). Less trivially, placing the bet may change the world, and hence your opinions, in other ways (betting at high stakes on the proposition 'I will sleep well tonight' may suddenly turn you into an insomniac). And then the bet may concern an event such that, were it to occur, you would no longer value the pay-off the same way. (During the August 11, 1999 solar eclipse in the UK, a man placed a bet that would have paid a million pounds if the world came to an end.) [p. 25].

However, the problems "stem largely from taking literally the notion of entering into a bet" and may be avoided by considering the probability assigned as "the betting price you regard as fair, whether or not you enter into such a bet." [Hàjek, 2007, p. 25].

To avoid these problems, there are operational definitions that use a penalty criterion. However, the general methodology in the operational definitions is to ask a person to state degrees of belief with the understanding that the stated numbers will enter into a game played afterwards, where the reference person may lose (and sometimes also win) something, and a condition of coherence decides whether an assignment is accepted as a probability. Thus, these can be considered the same assessment approach, corresponding to the subjective probability interpretation.

Obviously, this subjective interpretation focuses on the belief aspect of the concept probability. As was seen in Section 4.5, any probability assignment which satisfies certain rules (coherence) is accepted as a probability assignment some person could give. No judgement is made about whether an assignment is better than another[38] by the mathematical model, because no 'true' probability is looked for. In fact, de Finetti [1974] states right up front in the preface, and in all capital letters, "PROBABILITY DOES NOT EXIST" [p. x].

Important works in this field are, for example, those by Ramsey [1926] and Savage [1954]. Axioms of subjective probability are discussed for example by Fishburn [1986], references therein provide further literature. A seminal work is also the collection of articles by Kyburg and Smokler [1980]. The closely related topics of preferences and utility (not discussed here) are found in these works. When the focus is on updating a given subjective probability with data observed, one also finds the term "Bayesian" probability. We will not go into this either.

The operational definitions given for subjective probability are more generally applicable than for example the frequentits ones, in that probabilities can be assigned without supposing repeatable events. This makes subjective probability an interesting concept for vulnerability to climate change (see Section 4.8). Having thus sketched interpretations, it will be discussed in the following section how the mathematical models help provide clarity.

[37] See for example de Finetti [1974, p. 120].

[38] Of course, this has been investigated. See, for example, Chapter 3 "Does it Make Sense to Speak of 'Good Probability Appraisers'?" in de Finetti [1972]. An experiment on probability assignments concerning the results of football matches was carried out at the University of Rome, considering weekly probability assignments made by a group of mostly students for the matches of the Italian soccer championship.

4.7 Discussion: interpretations and mathematizations

Hàjek [2007] concludes his discussion of interpretations of probability saying that

> ...there is still much work to be done regarding the interpretation of probability. Each interpretation that we have canvassed seems to capture some crucial insight into it, yet falls short of doing complete justice to it. Perhaps the full story about probability is something of a patchwork, with partially overlapping pieces. [p. 31]

Similarly, Diaconis [2007a] states that "the discovery processes of uncertain reasoning is far from over."

For probability, at least a frequentist and a sujective interpretation still coexists. People maintaining one interpretation often voice disapproval of the other and not always in a neutral fashion. Kolmogorov [1991], for instance, dismisses subjective probability as "erroneous" in his encyclopedia entry. He discusses a weather example: considering the statement that on the 23rd March 1930 "snow probably still covered the fields" around Moscow, he stumbles about a quasi-subjective interpretation:

> As regards such estimates, it must be borne in mind that if they are applied to a statement which, by its very nature, can only be true or false, the estimate of its probability can only have a temporary or a subjective meaning, i.e. it merely reflects our attitude in the matter. [Kolmogorov, 1991, p. 302].

However, he considers the statement justified only by a "valid general rule" reflecting "the objective characteristics of the climate around Moscow" and hastens to add that the statement must be modified as soon as one finds out the actual climatic conditions from that year: "the snow on the fields near Moscow had vanished already on March 22" [Kolmogorov, 1991, p. 302]. That at this point the concept probability becomes superfluous, is not taken into consideration. Rather, Kolmogorov insists that rules (as about climate conditions) expressed in probabilistic terms "do have an objective meaning", and concludes that

> ...the calculation of mathematical probability in order to arrive at a degree of reliability of certain statements concerning individual events is no longer a mere expression of the subjective belief that the event will or will not take place. Such an idealistic, subjective understanding of the sense of mathematical probability is erroneous. If pursued to its logical conclusion, it would result in the absurd claim that valid conclusions about the world around us can be arrived at in complete ignorance, by merely analyzing subjective, more or less reliable opinions. [p. 303]

A caricature of the subjective interpretation found in the book by Maĭstrov [1974] shows that not all participants in the debate are well-informed about all interpretations. First, Borel is quoted as follows:

> Imagine a thousand Parisians passing by a seven story immovable property; they all agree to call it a house; however, they refuse to call a stone structure serving as a shelter to two rabbits and three hens a house. Let us consider an average structure (of these two); here opinion may be divided; if 748 out of 1000 voters call this structure a house, it would therefore be correct to assert that the probability that this structure is a house is 0.748 and the opposite probability is 0.252. [as cited by Maĭstrov, 1974, p. 242].

Maĭstrov comments: "Such an arbitrary interpretation of probability could have arisen only as a result of the ambiguity and vagueness of this notion." [p. 243] The translator S. Kotz adds in a footnote the remark: "We note, however, that this type of argument is basic for a subjective definition of probability." The "probability" as sketched by Borel, an average in the 'yes' or 'no' answers from 1000 people, is far from the same "type of argument" of probability considered the degree of belief of a single fixed person. In the example, the single persons do not express any uncertainty, each person states a certain 'yes' or a certain 'no'. The only idea in common with subjective probability seems to be that a question may be answered (subjectively) differently by different people.

Also the frequentist, or more generally, objective interpretations are being described in rather unfriendly manners. De Finetti [1974], for example, considers definitions of 'objective' probability "useless *per se*" [p. xii] and "irredeemably illusory" [p. 4]. Galavotti [2008] adds that objective probability is "an obstacle to a correct understanding of probability." [p. xviii][39].

While these two interpretations each put the emphasis on one of the two aspects in Hacking's "duality" of the concept probability, belief and frequency, both cannot leave out of consideration the other aspect. Kolmogorov was just quoted referring to estimates which have a subjective meaning. He seems to struggle with the (subjective) belief aspect coming to the fore unwantedly. "De Finetti's Theorem"[40] establishes (in a mathematically rigorous fashion) a connection of subjective probability to the frequency aspect and can be used for updating subjective probabilities of a "repeatable" event, when some trials of the experiment under consideration are observed.

Meaning and measurement of the concept probability are closely linked. For classical probability, it was seen that the theoretical definition of probability as a degree of certainty is made precise by the operational definition, which then illustrates the intuitive meaning of probability.

The objective interpretations propose operational definitions, and in the obvious examples, such as tossing a (biased) coin, again, the operational definition illustrates the meaning of the concept probability. The focus in the meaning here is on the frequency aspect: the measurement simply is a frequency. However, since repeatable events are required for the existence of a definite probability, the interpretation cannot deal with many questions involving uncertainty, such as single events. The question when exactly a (frequentist) probability exists is left open.

For the subjective approach, the existence of probabilities is more clearly defined. A degree of belief of a person exists for anything the person is uncertain about, it was seen that de Finetti does not admit unknown probabilities. Concerning the measurement of these probabilities of a person, reservations are stated in the literature as to whether the betting scheme actually measures subjective probability. However, again, the measurement can be considered an illustration of the meaning of the concept: it is intuitive that a person would pay less for a bet that she thinks to "probably" lose than on one where the chances of winning seem better. The fact that operational definitions provide the meaning of a concept in a rather concrete manner, and thus provide information about the interpretation used, will be important in the vulnerability context in the next chapter.

What does mathematics have to add to an analysis of a concept and its interpretations? First of all, mathematics has been used fruitfully in working with the concept itself, which turns out counterintuitive at times. The classical calculus of probability was a first mathematization that helped answer questions too complicated to treat using natural language only. Even with the mathematical model of probability, some results remain unintuitive, such as a simple game with four dice in which to each one there exists another one having a probability of 2/3 of scoring the higher number (that is, the intuitively expected property of transitivity is found lacking).[41] It has been argued that the basic ideas in probability

[39]Galavotti here comments on de Finetti's statement that "probability does not exist", concluding that de Finetti's "radical empiricism" deems objective probability thus.

[40]This theorem is not further discussed in this work, see e.g. Heath and Sudderth [1976], or Chapter 11 in de Finetti [1974].

[41]The game can be found for example at the museum of mathematics "Mathematikum" in Giessen, which, as an attempt at making mathematics interesting to a wider audience, connects back to Chapter 2. The game works as follows. There are four dice, labelled

A 3 on all faces

B four faces 4 and two faces 0

C three faces 5 and three faces 1

D four faces 2 and two faces 6

theory are not natural to classical mathematics [Kol'man, 1936, as cited by Maĭstrov, 1974] and there are famous examples of "pitfalls and delusions encountered by far from second-rate mathematicians" [Maĭstrov, 1974, p. 133] in probability theory[42]. Nevertheless, about 120 years after the Pascal-Fermat correspondence, Euler writes[43] that he intends to solve some "difficult problems on the basis of the long well-known principles of the calculus of probabilities".

Classical probability remained the foundation of the mathematical theory of probability which was much further developed for several centuries[44]. On the mathematical side, many results were established, for example limit theorems, and probability was being applied in the natural sciences, requiring "additional development of the mathematical techniques and tools of probability theory" [Maĭstrov, 1974]. All this occured without making progress in the foundations of the theory [see, e.g. Rényi, 1969].[45]

Also, the area of applicability of probability theory had never been clearly outlined. Many applications were unfounded or were thus considered. Maĭstrov [1974, see Chapter V] describes a case where probability theory was (mis)used for political purposes in Russia. As a result, probability was not considered mathematics, or not mathematics only, and, by the time that many areas of mathematics were being set on an axiomatic basis, probability theory acquired a rather dubious reputation.[46] Many people expressed the need for an axiomatization of probability theory. While the classical calculus of probability had provided the basic mathematics to solve some problems, the further development of probability theory as a mathematical discipline left the field where classical probability could be applied, the finite sets of possibilities. Neither the foundations of the mathematical theory nor the interpretation of concepts were clear at that point.

Enter Kolmogorov's axioms. These have been widely accepted as the solution to the problem of confusion: "The intuitive correctness of Kolmogorov's formalization was recognized almost at once, and is now universally adopted" [Suppes, 1968, p. 654]. These axioms, taking the concept probability as a primitive, do not explain its meaning and even less provide one universally accepted interpretation of the concept. The mathematical theory has clear foundations in these axioms, but they exclude certain examples, as seen for the uniform distribution on the natural numbers.

It is therefore not surprising that, in addition to the interpretation, also the question 'what is the best mathematical model of probability?' is up to debate. This debate produces just as fervent remarks as the one on interpretations. A textbook example is Williams [1991], whose introduction to probability theory is interspersed with (partly derogatory) comments on any attempt at considering probability without measure theory.[47]

Two players play, the first one chooses a die, the second one chooses a die knowing which one the first player has chosen. With the chosen dice, both roll simultaneously 10 times in a row. At each roll, the player scoring the higher number gets a point. The player who has more points at the end wins. For a single toss, A wins over D, D wins over C, C wins over B, and B wins over A with probability 2/3 each. The game is unfair in that the second player, seeing the die chosen by the first player, always has the possibility to choose a die which, with probability 2/3 scores better on the single toss.

[42] A frequently mentioned example is D'Alembert with his calculation of the probability for obtaining heads at least once in two consecutive throws with a fair coin as 2/3. He considered the alternatives 'heads; tails, heads; and tails, tails' as the Fundamental Probability Set.

[43] "Solutio quarundam quaestionum difficiliorum in calculo probabilium", 1785, here in the translation found in Maĭstrov [1974, p. 104].

[44] See, for example, Todhunter [1949] and chapters IV and V in Maĭstrov [1974].

[45] "Obwohl zahlreiche neue Ergebnisse erzielt und die Anwendungen in Naturwissenschaften und Wirtschaftsleben zu grosser Bedeutung gelangten, wurde in Richtung der Grundlagen der mathematischen Theorie der Wahrscheinlichkeitsrechnung im Grunde kein wesentlicher Fortschritt erreicht." [p. 83]

[46] "...haben zu Anfang des 20. Jahrunderts die meisten Mathematiker die Wahrscheinlichkeitsrechnung nicht als organischen Bestandteil und noch weniger als gleichrangiges Kapitel der Mathematik angesehen, sondern eher als einen zwischen Mathematik und Physik bzw. Philosophie stehenden ziemlich fragwürdigen Zweig betrachtet."[Rényi, 1969, p. 83]

[47] Interestingly, he seems to think that this is done only by philosophers, whom he does not seem to take seriously anyhow. An example:

> *Something to think about.* Some distinguished philosophers have tried to develop probability theory without measure theory. [...] The moral is that the concept of 'almost surely' gives us (i) absolute precision, but also (ii) enough flexibility to avoid the self-contradictions into which

For the axiom of countable additivity, some of de Finetti's objections, not always presented in a neutral fashion, have been given in the previous sections. The articles de Finetti [1930c], Fréchet [1930a], de Finetti [1930a], Fréchet [1930b] and de Finetti [1930b] portray the debate between the two authors about this axiom in the year 1930. An interesting point about the confused situation and the adoption of one or the other mathematization of the concept probability, that is, requiring finite or countable additivity, is made by de Finetti [1972]:

> For a long time the opposing viewpoints have been applied indiscriminately [. . .]. If no contradictions seemed to arise, or at least, to be felt, this was largely due to the disconnected character of the applications. However, when the time came to place probability theory on unambiguous foundations, the following dilemma became inescapable: *either to exclude problems of the type of an integer chosed at random (and then give reasons why such problems should not be considered), or else, to forego conclusions based on [countable] additivity (and then be ready to accept some technical complications).* [p. 87].

This "dilemma" is one that obviously cannot be solved by mathematics: the choice of axioms is extramathematical. In this case, the advantages are general applicability for the finitely additive versus mathematical conclusions from measure theory for the countably additive choice. As measure theory has been established at the beginning of the twentieth century and has become rather common knowledge in mathematics since, some time later finite additivity could be considered mathematically more interesting: "some of the new problems that the finitely additive approach does introduce are of mathematical interest in themselves." [Dubins and Savage, 1965, p. 8]. Gangopadhyay and Rao [1997] describe:

> Whether there is a need to develop probability in a finitely additive setting, is a debatable question. One must go back to the writings of de Finetti to see the spirit. At a purely mathematical level, this exercise seems to have some interesting points. [p. 644].

One other point to be considered here is the representation of events as subsets of a given set and random variables as functions on this set. In fact, it could be considered cheating to present de Finetti's work on probability using this set representation which he has vehemently objected to. However, not only does de Finetti's approach take quite some getting used to, at this point it simply appears less clear.[48] De Finetti provides the intuition about a random quantity that first is unknown but later is known via several possible values and one true value, which must be one of the possible values. The relation between the true values of several random quantities does not appear obvious.[49] When several random quantities (or events) are considered at the same time, as in the betting scheme, it is not immediately clear how the true values of these relate to each other.

The set representation fixes the relation between the true values of the random quantities in a clear way, as they are all functions of the same outcome of an underlying experiment. In the case of events, de Finetti also speaks of their indicator functions. The question what the domain of definition of these indicators is seems to remain open. The set of constituents of a given set of events (which can equally be defined when events are propositions, not subsets) can be used: one constituent will be the true constituent in the end and determine the values of the indicators of all events. This description, provided also by de Finetti, is similarly clear, probably because it is not very far away from the set representation. In fact,

> those innocent of measure theory too easily fall. (Of course, since philosophers are pompous where we are precise, they are thought to think deeply . . . .) [p. 25].

[48]This may or may not be due to the habit of standard probability theory for the intuitive point of view. That without set representation more axioms for probability would be needed in a fully formal description was pointed out in Footnote 27.

[49]Again, this may be a question of habit. The discussions of possible points and the "linear ambit" of several random quantities at least seems more complex than the standard representation. Examples provided are clear, and geometric illustrations add to the intuition, however, these concern simple special cases, such as events instead of random quantities and three dimensions only.

it is now hard to imagine probability theory without such a representation. Here, Suppes' statement about Kolmogorov's axiomatization being intuitively accepted has been found warranted.

Let us return to different interpretations of probability. Given one or more mathematical models, these can be discussed "at another level" as Suppes pointed out. Interpretations of probability may be compared on the background of a given mathematical model. The criterion of 'admissibility' of an interpretation proposed by Salmon [1966] contains the statement that a "fundamental requirement for probability concepts is to satify the mathematical relations specified by the calculus of probability".[50] For example, Hàjek [2007] states for each interpretation which of the axioms proposed by Kolmogorov it satisfies. That coherent probability assignments are finitely additive (and when defined on algebras are finitely additive probabilities) and only finitely so, has become clear above. However, Kolmogorov's axioms and the interpretation proposed by Kolmogorov in the encyclopedia are not as closely related as one might expect. Relative frequencies in finite sequences are finitely and (trivially) countably additive, but limits of relative frequencies are not necessarily countably additive [see, e.g. de Finetti, 1972, Section 5.19].

Vice versa, in the debate about the different mathematical models, interpretations may be used to argue for or against a model. This has been seen in the section on de Finetti's probability, where he refutes countable additivity on the basis of the subjective interpretation. He even uses interpretations he does not adhere to for this purpose. In a discussion of both different interpretations and different mathemtical models [de Finetti, 1972], he argues against an algebra structure for the domain of definition of probability functions and against countable additivity from a frequentist point of view. The fact that limiting frequencies are not necessarily countably additive and the fact that given two events for which limiting frequencies exist, this need not be the case for the intersection, according to de Finetti, are reasons for giving up these axioms. Jumping to the conclusion that Kolmogorov's position in the debate is weakened, would however be rash. Kolmogorov was explicitly not concerned with the interpretation in setting up his axioms and considered probability a primitive. While objecting to subjective probability, he considers both classical and frequentist measurements as warranted in some cases. The formal framework of vulnerability presented in the following chapter will also leave open the question "where concrete numbers come from". However, vulnerability will not be a primitive as the framework aims at clarifying the meaning of the concept.

Before concluding this chapter on probability, we take a look at how this concept is used (or avoided) in the climate change context, as an example that further has implications for vulnerability assessments.

4.8 Probability in climate change research

Climate change decision making mostly is decision making under uncertainty, in that it concerns future climate. Probability suggests itself as a tool to be used, however, its use is debated. The mathematical model of probability used is generally that from standard probability theory. Both the frequentist and the subjective interpretation of probability can be found.

The frequentist interpretation seems to dominate, however, it is sometimes problematic. Single events, for example, are of importance in the climate change context. Questions like 'What does a frequentist probability mean for the tipping of the thermohaline circulation? Which are the regularities warranting the existence of a definite probability of such an event, and how can one obtain a number for this probability?' reveal that the frequentist interpretation does not always make sense in the climate change context. Even for events where

[50] What is meant by the calculus of probability of course has to be specified in this case. Salmon's axioms differ from Kolmogorov's, see Salmon [1966, p. 59].

long data records are available, such as paleoclimatic temperature records over hundreds of thousands of years obtained from ice-cores, the fact that one is considering anthropogenic climate change implies that these records may not be useful to deduce probabilities of future events from past records of climate, where most data come from a time before the industrial revolution. Also, when the object of study is the global climate system, one cannot use a common technique to produce frequentist probabilities: making experiments.

Schneider [2002], who advocates the use of subjective probability, explains that

> the probability of some level of future climate change is not determinable directly by any set of frequency experiments – so-called frequentist probabilities – and instead will rely to some degree on scientific judgments based upon as much empirical observation as is possible. [p. 444]

Schneider describes these subjective judgments made as "essentially Bayesian judgments about the plausibility of the assumptions and structure of the systems model" [p. 444]. The fact that "any such estimates will be highly subjective and often carry a fairly low confidence" [p. 443] is perceived as problematic in climate change research.

Some problems seen with the frequentist interpretation do not apply to the subjective interpretation: single events, as well as events for which no frequencies are available can, at least theoretically, be ascribed subjective probabilities. The actual measurement, for example by expert elicitation, is another question that shall not be discussed here, see for example Morgan and Henrion [1990]. Considering probabilities given by experts, de Finetti's coherent probability assignments could be an advantageous mathematical model because the extension of the probability function is arbitrary. Coherent assignments can be defined for any set of events, one can

> assume or suppose [an assignment to be] defined or known for all (and only) the random quantities (or, in particular, events) belonging to some completely arbitrary set [. . .]: for instance those for which we know the evaluation explicitly expressed by the individual under consideration. [de Finetti, 1974, p. 84].

This individual under consideration could be the expert, the model seems convenient for the situation of an elicitation.

However, subjective probabilities are equally controversial as frequentist ones. In fact, probability is sometimes rejected in the context of uncertainty about future climate, future emissions etc., without specifying an interpretation.

A central issue in the debate about the use of (subjective) probability are the IPCC SRES scenarios of future emissions [Nakićenović and Swart, 2000]. 40 scenarios have been developed, based on four qualitative storylines differing in demographic, social, economic, technological and environmental developments. There are six groups of scenarios (as one storyline is further subdivided), and for each group a so called illustrative scenario is identified. The scenarios are described as "equally sound" and all of them "are equally valid with no assigned probabilities of occurrence." [Nakićenović et al., 2000, p. 4].

Schneider [2001] tells that in the scenario development, participants of a preliminary meeting he attended could not agree on likelihoods of the storylines. Therefore, "in an attempt to avoid endless disputes" [p. 18], no probabilities were assigned to scenarios. He criticizes the "probability vacuum" resulting from this decision:

> In the risk-management dilemma that constitutes climate-change policymaking, I would definitely put more trust in the probability estimates of the SRES team - however subjective - than those of the myriad special interests that have been encouraged to make their own selection. [p. 19]

That decision makers will use (their own) probability estimates is inevitable. According to Pittock [2001], even adaptation and mitigation strategies such as increasing robustness or resilience "tacitly assume some estimates of likelihood . . . " [p. 249]. In a certain sense, implicit probabilities are assigned by any consideration of (some of) the 40 scenarios. One may consider them all in the same way or make some differences, but to avoid an implicit relative weighting of scenarios is not possible. For a finite number of scenarios, a weighting immediately translates to a probability measure, as also suggested by Schneider [2001] referring to "the relative likelihood of each scenario" [p. 19]. From the subjective probability point of view, where probability is a degree of belief, one might argue that probabilities cannot not exist. Similar to de Finetti's statement that there are no unknown probabilities (see page 55), there are no non-existent probabilities for a given individual facing a decision under uncertainty.[51] Experts' probabilities have the advantage that established techniques to control biases in subjective judgements can be used in the elicitation process, apart from the fact that experts are knowledgeable on the respective subject.

For non-experts, the assignment of equal probabilities strongly suggests itself, given the description of the scenarios as "equally sound". The formulation is very similar to the expression 'equally possible' that, in classical probability, was reduced to 'equally probable'. Immediately the reference class problem arises: does one choose the 4 storylines, all 40 scenarios or the 6 illustrative scenarios as the equally probable alternatives? That the answer is unclear may have the benefit of discouraging this simple use of classical probability, but the problem how to obtain probabilities for risk management questions is not solved.

Equal probabilities are not what is intended by the makers of the SRES scenarios. Grübler and Nakićenović [2001], in a one-page reply to Schneider [2001], explain that

> 'dangerous' levels of climate change will need to be identified by research into the adverse impacts on natural and human systems, independent of the question of how likely they are to occur, and covering the full range of scientific uncertainty.

While they consider "assigning subjective probabilities of occurrence" neither appropriate nor feasible, they later interpret probability in a frequentist sense and confine its use to the natural sciences. The subjective interpretation of probability is not further discussed in this reply. A main concern of the authors is "a danger that Schneider's position might lead to a dismissal of uncertainty in favour of spuriously constructed 'expert' opinion."

There is a mismatch between the actual concern and the question whether probabilities should be assigned to scenarios or not. Scenarios are described as projections, and explicitly as not being predictions [see, e.g., Carter et al., 2007], and one often finds the terms 'deep uncertainty' associated to future descriptions in climate change research. At the same time probability is unduly associated with prediction, for example by phrases like "predicting the future in terms of probabilities." [Dessai and Hulme, 2004, p. 113][52].

The idea that probability reduces uncertainty is sharpened in the subtitle of Grübler and Nakićenović [2001]: "We need to research all the potential outcomes, not try to guess which is the likeliest to occur." This much stronger description, in that assigning probabilities does not necessarily mean concentrating on the likeliest possibility only, seems to capture the authors' concern about the full range of uncertainty better. They want to avoid that decision-makers, when given probabilities, dismiss all but the most probable scenarios as irrelevant. A similar concern is expressed by Lempert et al. [2004], who consider two frameworks for decision making. The "Predict-Then-Act Approach" ranks policy alternatives "on the basis of their expected utility, contingent on the probabilities" [p. 2], while the "Assess-Risk-of-Policy" framework starts with available policy options and assesses uncertainties associated with these. The authors support the latter as the framework which "may provide more policy relevant results" [p. 7]. According to the authors, the Predict-Then-Act framework "can suggest implicit criteria that scientist and analysts do the best job when they reduce uncertainty as much as possible" [p. 6]. That probability reduces uncertainty is suggested also when they describe probability as providing "factual statements about the world" [p. 4], and as suggesting "a single, correct answer" [p. 5].

Rather than a problem of the concept probability, this is a problem of how probabilities are used. If one assumes that when given probabilities, decision-makers will automatically

[51] The question could be discussed more thoroughly in Decision Theory, which however we do not go into here.

[52] According to these authors, analysis based on large data sets and aggregated numbers, and the increase in computational power have "arguably led to the emergence of predicting the future in terms of probabilities."

use only aggregate information like an expected value and consider it as certain, it may indeed be useful to refrain from assigning probabilities. However, the close link between probabilities and prediction established by some works means misunderstanding the concept probability, and especially subjective probability according to de Finetti. The problem is not new:

> There will be many and frequent occasions to warn against errors, misunderstanding, distortions, obscurities, contradictions and the other endless troubles which are so difficult to avoid when dealing with probability, and which are always essentially the result of ignoring the same warning: *prevision is not prediction*! [de Finetti, 1974, p. 98].

Here, prevision corresponds to expectation, which is even a more aggregate information than probability itself. De Finetti explains the differences at length, describing prediction as trying to guess among the possibilities the one that will occur, while prevision, and thus probability, "acknowledges that what is uncertain is uncertain" [p. 71] and merely expresses a person's attitude of expecting something rather than something else to happen or a quantity to be small rather than large[53]. The fact that "the future is unknown" is by no means contradicted by the concept probability. The description of the second role of probability, given by Lempert et al. [2004], "a coherent mathematical framework for summarizing information" is much closer to these ideas. It seems unfortunately not as widespread in the discussion about attaching probabilities to scenarios.

The non-assignment of probabilities to the SRES scenarios has consequences for vulnerability assessments as seen in the following chapter. Many assessments deduce possible future evolutions from the IPCC SRES scenarios and thus do not use a probabilistic description of the uncertain future. This is different for risk assessments in the context of natural hazards, where the "classical definition of risk: probability times consequence" [Schneider, 2002, p. 443] is applied.[54]

Let us conclude this section with a summary and recommendation by Dessai and Hulme [2004]:

> Probabilities of climate change will remain subjective – there is no such thing as 'true' probabilities – so it is extremely important for researchers to be as explicit as possible about their assumptions.

Of course, this recommendation should hold for science in general. As subjective probabilities are admittedly subjective, they are generally regarded more suspiciously than other scientific descriptions of uncertainty.

4.9 Conclusions

This chapter has introduced the concept probability and several mathematizations of it. These are classical probability, Kolmogorov's (now standard) probability model, finitely additive probabilities, and coherent probability assignments. As presented in this work each model is a generalization of the previous one. Classical probability can be seen as a special case of the standard model, where the sample space is finite and all elementary events are assigned equal probability. The standard model, where a countably additive probability measure is defined on a σ-algebra over an underlying set, is generalized by finitely additive probabilities, defined on an algebra over a set, and dropping the axiom of countable additivity. Lastly, coherent probability assignments were presented as functions defined on an arbitrary set of subsets[55] which can be extended to finitely additive probabilities on the algebra generated by the domain of definition.

The different mathematizations were seen to serve different purposes. Kolmogorov's axioms build mathematical foundations for a mathematical theory that was already devel-

[53] See also the quote on page 35 about this attitude which goes beyond the mere assertion of possibility.
[54] See Section 5.5 for the use of "uncertainty" in that context.
[55] The original work by de Finetti does not use the set representation of events.

oped to some extent. Taking the concept probability as a primitive, the mathematization does not explain what probability means nor how it can be measured. De Finetti gives an operational definition that allows to measure probabilities. He wants to use mathematics explicitly to clarify concepts and is thus concerned with the connection between the mathematical theory and the meaning of the concept. The example of probability has thus shown that according to the different purposes that formalization is used for, it produces rather different results. While all results are mathematical models, their relation to the concept, in terms of clarification of the meaning and measurement of it, can be very different.

As discussed in Section 4.7, the debates about what probability means, how it can be measured and which mathematical model best represents it are not conclusively answered as yet, and probably never will be. Frequentist and subjective interpretation co-exist, considering probability a characteristic of uncertainty deriving from statistical regularities or a degree of belief of a person, respectively. These interpretations each emphasize one aspect of the duality of the concept made out earlier, a frequency aspect and a belief aspect inherent to probability. Formalizing concepts was shown to have some of the benefits proposed in Chapter 2. For example, the set representation makes explicit in a clear fashion how different random quantities relate. Given a mathematization, different interpretations can be discussed more clearly, but also vice versa, interpretations are used to justify some mathematical axioms or reject them.

Finally, an application of probability was looked at: probability in climate change research, in particular the example of the IPCC SRES scenarios. The debate about whether the uncertain future should be described in probabilitistic terms or not again illustrated that the concept probability is far from clear. A misunderstanding encountered here is that probability serves to reduce uncertainty, while it merely serves to describe it. This problem may derive from how probabilities are being used, there seems to be a fear that decision-makers usually throw away all information and consider the most likely alternative as certain. It was also seen that the subjective interpretation of probability is not as well-established as the frequentist one. The meaning of probability in this interpretation seems less clear and its subjectivity is heavily opposed while subjectivity of other scientific descriptions of uncertainty, such as scenarios, is much less discussed.

A misunderstanding about formalization often encountered in the vulnerability context should have been dispelled in this chapter: formalization does not mean providing the one and only "correct" definition in a conceptual debate. On the contrary, for probability it has led to different mathematical definitions in addition to the different interpretations. These however could be used to discuss interpretations more clearly, confirming Suppes' statement that formalization may raise the discussion to another level.

Chapter 5

Vulnerability

You have to wait until tomorrow to find out what tomorrow will bring.
[Murakami, 2009, p. 104]

The topic of this chapter is the formalization of vulnerability, which was motivated by the confusion in the terminology discussed in Chapter 2. Proposing mathematics as a language to clarify concepts, a formal framework of vulnerability to climate change has been developed by the FAVAIA research group[1] at the Potsdam Institute for Climate Impact Research (PIK), as part of the ADAM project[2]. This framework has evolved over several years and has been published in different stages of its development [Ionescu et al., 2005, 2009; Wolf et al., 2007, 2008; Hinkel, 2008; Ionescu, 2009; Wolf, 2009]. The formal framework consists of two mathematical models of vulnerability, one of the ordinary language concept and one of the scientific concept.

The mathematical model of the ordinary language concept is presented in detail in Section 5.1. In a nutshell, vulnerability is defined mathematically as a measuring function of the possible future harm that may occur to an entity. This definition serves as a basis for the analysis of theoretical scientific definitions of vulnerability in Section 5.2. The mathematical model of the scientific concept vulnerability developed by Ionescu [2009] is presented and adapted to our purposes in Section 5.3. Given the general and precise mathematical definitions, in Section 5.4, different vulnerability assessment approaches are identified and explain part of the confusion: the same term is used for different kinds of assessment results. The assessment approaches are also related to interpretations of vulnerability identified in previous conceptual literature. Another source of confusion is explained by using the formal framework to analyze risk assessments in the context of natural hazards in Section 5.5.

Throughout this chapter, the formal framework of vulnerability is displayed in three forms, from a mathematical description via a semi-mathematical diagram to a box-and-arrow-diagram closer to what is commonly used in conceptual work on vulnerability. The different modes of presentation, which owe to the fact that the formal framework's primary audience are non-mathematicians, are discussed in Section 5.6. Finally, Section 5.7 bridges back to the previous chapter before Section 5.8 concludes.

5.1 A mathematical model of vulnerability in ordinary language

This section serves an introductory purpose in two respects: first, it demonstrates the formalization process to some detail, and second, the result of this process – a general mathematical model of vulnerability – is the basis for an analysis of scientific definitions of vulnerability later. Ordinary language was chosen as a starting point for the framework development

[1] www.pik-potsdam.de/favaia
[2] http://adamproject.info/

because a common language in this interdisciplinary field must be very general to point out the commonalities and differences of approaches to vulnerability from different disciplines.

The formalization process outlined in Section 2.5 is followed here: the ordinary language meaning of vulnerability is analysed to identify the primitives in Section 5.1.1. An example mathematical model in a probabilistic setting illustrates some ideas in Section 5.1.2. The primitives are translated into mathematics in a more general setting, and vulnerability is mathematically defined in Section 5.1.3. While this formalization is very general and introductory, of course the focus is directed towards vulnerability to climate change. Hence, reference to the scientific concept will be made; most examples refer to scientific vulnerability assessment. This context provides natural examples, because the aspect of measurement will be prominent in the mathematical model. When referring to the meaning of the concept vulnerability, terms like 'assessing' vulnerability are used without reference to a real-world vulnerability assessment, but rather to an abstract description how vulnerability could be evaluated. In fact, the formal framework of vulnerability is not to be confused with a framework for assessing vulnerability: it does not provide guidelines how to conduct an assessment. Nevertheless, taking it into consideration may help start an assessment in a systematic manner, for example, by making explicit what the primitives of vulnerability, identified in the following section, are in the situation under consideration.

5.1.1 Analysis

The first step in the formalization process is an analysis of vulnerability as used in ordinary language. In ordinary language people have a working understanding of words and precise definitions are rarely needed. The analysis in this section identifies primitives of the concept vulnerability in its working understanding, supported by a glance at the etymology of the word and by a dictionary definition.

'Vulnerable' has Latin origins: 'vulnerare' literally translates as "to inflict a wound on, wound; to damage (things)" and further "To wound (mentally or emotionally), hurt, distress; to damage the interests of" [Glare, 1982]. The suffix -able (Latin: -abilis) conveys the notion of possibility: 'being wounded' may (but need not) happen at a later point in time.

These elements are explicit in the Oxford Dictionary of English[3] entry "vulnerable" [Soanes and Stevenson, 2005]:

> exposed to the possibility of being attacked or harmed, either physically or emotionally: [...] *small fish are* ***vulnerable to*** *predators.*

The adjective "vulnerable" is ascribed to somebody or something, that can be very generally described as an *entity*. An entity is vulnerable (now) if there exists a possibility that it is attacked or harmed (later). To decompose the concept vulnerability into primitives, we consider the notions of *possibility* and *harm* separately.

The property "vulnerability" is ascribed to an entity at a certain point t in time on the basis of information about the future relative to t. In the following, t will also be referred to as the *present*, and the *future* is to be understood as relative to t. Since information about the future is required at time t, it involves uncertainty; quotes along the lines of "The future is, of course, unknown." [Adger, 2006, p. 276] are ubiquitous in the vulnerability literature. Viewed from time t, there is not one certain future evolution of the entity but several evolutions are considered possible. In de Finetti's terms, we consider a range of possibilities or, similar to a random quantity that has several possible values, a random evolution[4].

Uncertainty is essential in the concept vulnerability. An entity that is going to be harmed for certain is not the prime example of a vulnerable entity. Reconsidering the present vulnerability of an entity at a later point in time t_1, when time has passed and the evolution

[3]This definition is used as one representative choice. The everyday usage of the word "vulnerability " is sufficiently clear for the present purpose without a detailed analysis of different dictionary definitions.

[4]See Section 4.5.

of the entity after the present time t has been observed, the vulnerability at time t is not necessarily determined by this observed evolution. When no harm has happened, one cannot conclude that the entity was not vulnerable, it might just have been lucky. One observed evolution generally does not carry enough information about the *possibility* of harm that was given at t. In the literature this is summed up for example by Blaikie et al. [1994], "vulnerability is a hypothetical and predictive term" [p. 58], or by Calvo and Dercon [2005] "vulnerability is used as the magnitude of the *threat* [...], measured ex-ante, before the veil of uncertainty has been lifted." [p. 2].

The ordinary language use of statements like "someone is more vulnerable than someone else" reveals that the adjective vulnerable has the grammatical property of comparison, that means the degree to which the modifier modifies its complement can be distinguished.[5] Consequently, the abstract noun vulnerability, constructed with the noun-forming suffix '-ity' describes a property that admits degrees. This property is asserted at the present but involves a view of the uncertain future. We will refer to present and future as the *time aspects* of the concept, which become important in the discussion of the scientific concept.

We have identified

- the entity,
- its uncertain future evolution, and
- the notion of harm

as the primitives of vulnerability. The relation found between these primitives can be summarized as follows: imagine viewing the uncertain future of the entity by seeing several possible future evolutions, and assuming that these are all possibilities. The entity is vulnerable if in some possible future evolution harm occurs to it. And, roughly speaking, it is more vulnerable if more harm is possible. This basic idea is made precise by the formalization in the following sections: vulnerability is a monotonic aggregation of possible future harm.

An important grammatical construction for discussing "vulnerability *to* climate change" is "vulnerability to something" as seen in the dictionary's example sentence "small fish are vulnerable to predators". This something (the predators in the example sentence) is considered a cause of harm to the entity. It is a further primitive for this case and will be referred to as *the stimulus*. The uncertain future evolution considered in this case comprises the entity and the stimulus in interaction. In order to describe this interaction, a description of the entity's environment is needed, where environment is understood as anything distinct from the entity taken into consideration in the description. According to Gallopín [2006], "vulnerability becomes a property of the relationship between the system and its environment" [p. 296] where the system is our entity while, when no stimulus is considered, vulnerability is "a property of the target system" [p. 297].

The construction 'vulnerability to something' can also be found with a different use, as is the case for 'vulnerability to poverty', where the complement refers to the kind of harm considered instead of the stimulus causing harm. While not as important in the climate change context, this construction will be considered briefly in Section 5.3.

The example probability suggests some similar questions as discussed in the previous chapter to be asked for the concept vulnerability: when, if at all, does a "true vulnerability" exist? Can Kolmogorov's hypothesis of the existence of a definite probability[6] also be considered for vulnerability, in other words, does the hypothesis that "a definite vulnerability exists" have to be "verified or justified in each individual case"? Or is vulnerability a subjective estimate that can be measured only with respect to a given reference person? While some definitions of vulnerability ("vulnerability is a perception", see the definition

[5] An example of a non-comparative adjective, a special case from the English literature excepted, is 'equal'.
[6] See Section 4.3.4.

by Kuban and MacKenzie-Carey [2001] in the table on page 85) can be understood in this direction, this discussion is not as prominent for vulnerability as for probability (see, however, Section 5.4.4).

A further conceptual question for vulnerability could be whether one also finds the duality observed for the concept probability. This seems likely because Hacking [1975] finds a similar duality contained already in the concept possibility itself. "It is possible that" refers to something that might or might not be the case, but one does not know; thus, possibility has a *knowledge* aspect. The second aspect occurs in the formulation "it is possible for someone to do something". This *physical* aspect of possibility describes "actual abilities independent of our knowledge of them" [Hacking, 1975, p. 123]. Comparing to probability, this aspect is similar to the "facility" with which a die was said to fall with a certain face up, see also the quote on page 37. The interpretation of possibility as ability will be found in vulnerability in the concept capacity, see Section 5.2.2.

As an aside, looking up "expose" in the same edition of the Oxford English Dictionary, one finds the construction "expose someone to" which is used in the above definition of vulnerability ("exposed to the possibility ..."). This construction is defined as "cause someone to be vulnerable or at risk" [Soanes and Stevenson, 2005]. Such (partial) circularities in definitions are often hidden in ordinary language and may add to the confusion.

5.1.2 Probability suggests a simple mathematical model

Having to translate an entity and its uncertain future evolution into mathematical concepts after a chapter on probability theory, a probabilistic model of vulnerability suggests itself. We will sketch the vulnerability basics in a simple probabilistic model before going into the more general mathematization of vulnerability in the following section.

Some preliminary remarks are necessary: to avoid repeating the 'the entity, or the entity and its environment' we will in the following refer to 'the system' as containing the entity and its environment as far as under consideration in order to assess the entity's vulnerability (to the stimulus). Moreover, in all of the following, we consider a fixed time point t as a reference time point at which the entity's vulnerability is assessed and a finite time horizon $t_h > t$ as the scope of the assessment. The further evolution after that time is not taken into consideration for assessing the vulnerability of the entity at time t. We assume that $[t, t_h] \subset \mathbb{R}$, achieved for example by an appropriate mapping of real-world dates to real numbers. The time difference between t and t_h considered in vulnerability assessments varies, from for example one year, typical in studies of vulnerability to poverty, to a hundred years in climate change studies.

A natural simple probabilistic model for describing a system and its uncertain future evolution is a homogeneous Markov chain on a finite state space[7] (see Definition 4.3.4). The situation of the system at a fixed time point is described by a *state*, denoted s, an element of the state space of the Markov chain, S. A state collects all the relevant information about the entity and its environment at that point in time. The uncertain future for one time step is described by a probability distribution over states, that is, possible next situations of the system. In the homogeneous Markov chain model this probability distribution depends only on the given state.

Consider a starting state $s_0 = s \in S$ fixed, where the lower index denotes the time step, from which the system makes a number of say h transitions up to the time horizon t_h. The possible future evolutions of the system considered to assess the entity's vulnerability in state s are the trajectories of h steps, $s, s_1, \ldots, s_h$ with $s_j \in S$ for all $j = 1, \ldots, h$. Their probabilities can be computed as described in Section 4.3.3. Denote the resulting probability distribution over the trajectories by $P^{(s)}$.

[7]We choose the simplest model since the aim is to illustrate some features of vulnerability, not to give a maximally general probabilistic model. Also, the finite state space exempts us from the decision for or against countable additivity.

An entity is considered vulnerable if harm is possible in its uncertain future evolution. In this model, this suggests an evaluation of harm for each of the trajectories describing a possible future evolution. Let us look at the mathematical representation of harm in more detail.

Excursus on the primitive harm

The notion of negative effects or harm presupposes a notion of comparability, interpreted as 'worse than'. The mathematical representation chosen in the formal framework is a partial strict order $\prec$, that is, an anti-reflexive and transitive binary relation:

- nothing is worse than itself, or $\forall x \neg(x \prec x)$, and
- if x is worse than y and y is worse than z, then x is worse than z, or $(x \prec y) \wedge (y \prec z) \Rightarrow (x \prec z)$.

The possible partiality of the order accounts for the fact that problems of non-comparability easily arise in vulnerability assessments. The mathematical model should thus allow for the result that some situations are not comparable and require (further) value judgement that lies outside of the mathematical framework.

Denote the set of trajectories representing evolutions by Ev and its elements by ev. Then, the mathematical representation of the harm evaluation is a function $h : Ev \longrightarrow H$, where H is a partially ordered a set of harm values. The prime example of a partially odered H in vulnerability assessments is a set of tuples with a (total) order on each component, where the components record different phenomena such as 'people affected', 'monetary damage', etc. Often, total orders on the tuples are derived by weighting the components, where choosing the weights of course involves value judgement. For example, indices are created from several indicators by weighting these. In the following we will consider $H = \mathbb{R}^n$ for simplicity, with the standard partial order given by the Euclidean norm.

While the function h is applied to trajectories of states of the whole system, the interpretation is that it measures harm which occurs to the entity under consideration. In the case of vulnerability to a stimulus, the interpretation is even more specific: the function h measures "harm to the entity that is due to the stimulus". This could be made explicit in the notation: e.g., $h_{stimulus}$ could be replaced for h everywhere in the formulae to follow. However, the formalization does not change for this case.

For further reference, and in order to not confuse with measures in the sense of measure theory, we define:

Definition 5.1.1. Functions that take values in a partially ordered set will be referred to as *measuring functions*.

Back to vulnerability

From the Markov chain representation of the system's future evolution, we have a probability measure over trajectories. The evaluation of harm is represented by a measuring function on these trajectories. Vulnerability tells something about possible future harm as viewed from the starting state. This suggests to aggregate the harm values of the possible evolutions. The first aggregation that comes to mind in a probabilistic context is the expected value.

The harm function $h : Ev \longrightarrow \mathbb{R}^n$ can be considered as a random variable (n-dimensional) on Ev. The probability measure $P^{(s)}$ is defined on Ev, wherefore the expected value can be computed.

A possible definition of vulnerability is expected harm:

Definition 5.1.2 (vulnerability as expected harm)**.** $v(s) = E_{P^{(s)}}(h) = \int h dP^{(s)}$

Varying the starting state s, vulnerability mathematically is a measuring function on states, $v : S \rightarrow \mathbb{R}^n$. Note that the order on vulnerability values for different states can again be partial, because, while each component is totally ordered, there is no default order on $\mathbb{R}^n$. One could again choose the Euclidean norm. The value $v(s)$ associated to a state is of course to be interpreted as *the vulnerability of the entity in state s*, or, if harm due to a stimulus was considered, *the vulnerability of the entity to the stimulus in state s.*

The expected value as the vulnerability aggregation is just one possibility out of many. However, not all aggregation functions that can be applied to a probability distribution over harm values and produce values in a partially ordered set are necessarily sensible aggregations in this vulnerability model. Ionescu [2009] gives as a counterexample the most likely harm value, which, when used to aggregate, may produce unintuitive results. Roughly speaking, an entity is considered more vulnerable when more harm is possible. The most likely harm value as aggregation allows examples where this is not satisfied. Keeping the same probability distribution and increasing some harm values in it may lead to a lower vulnerability value as in Ionescu's example consisting of the harm values $10, 0$ an 1, with the probabilities $0.4, 0.3$ and 0.3, respectively, and the non-decreasing function

$$g(n) = \begin{cases} 1 & \text{if } n = 0 \\ n & \text{otherwise.} \end{cases}$$

If g is applied to the harm values above, these become $10, 1$ and 1. Therefore, the most likely harm value changes to 1, which has a probability of 0.6. An aggregation like this, for which the vulnerability of a given situation is higher than that of the modified situation in which no harm value has decreased and some have increased, seems contrary to the working understanding of vulnerability in ordinary language. Conditions that an aggregation function has to satisfy in order to be admissible in the vulnerability definition will be discussed in the following section, where the general mathematical definition of vulnerability is introduced.

5.1.3 A general mathematical definition

The previous section has shown a possible mathematical definition of vulnerability in a probabilistic setting, illustrating its basic elements. Borrowing from the theory of dynamical systems, the system under consideration at a given point in time is described by a *state*. This should contain all relevant information for assessing vulnerability about the situation of the entity and its environment at that time. States are collected in the set of states S. The system's uncertain future evolution as viewed from that state at that time requires a more general mathematical translation than in the probabilistic case. This will be given below. Then, the harm measuring function, which remains as above, is again applied to all possible future evolutions provided by the description of the uncertain future, and vulnerability will result by aggregating. Again, mathematically, vulnerability will be a property of a state, $v(s)$, meaning *the vulnerability of the entity (to the stimulus) in state s*. Let us consider the more general mathematical description of the uncertain future evolution of the system.

Uncertain future evolution

There are various ways of describing the uncertain future as seen from the perspective of a state at a fixed point in time. Which one is most useful depends on the available information in each concrete case. The Fourth Assessment report by the IPCC (Chapter 2 by Working Group II) lists artificial experiments, analogues, projections, sensitivity analysis, scenarios and storylines, and probabilistic futures as characterisations of future conditions [Carter et al., 2007]. While probability is a well-established mathematical method of capturing some information about uncertainty, its use in the climate change context is debated, as seen in Section 4.8.

The use of scenarios, which are "not predictions or forecasts [...] but alternative images without ascribed likelihoods of how the future might unfold" [Carter et al., 2007], is quite common in studies of vulnerability to climate change because many make use of the SRES scenarios, which come without probabilities. Mathematically, such a future description is given by associating a set of possible future evolutions to a given state. It is referred to as *non-deterministic* description. A further mathematical description of uncertainty is the *fuzzy* one, used for example to represent qualitative expert judgement. There is a common structure in these examples, given by the fact that sets of possible future evolutions are considered. While in the non-deterministic description this is all the information, in both the probabilistic and the fuzzy description further information about the elements of the set is given. In the case of discrete sets, numbers are attached to the possible future evolutions, probabilities in the probabilistic case, degrees of membership in a given set in the fuzzy case. The general mathematical concept that captures all three examples is a functor, defined in Section 3.3. Generalizing from the probabilistic case, where one has probability distributions over a set, we will refer to the result of applying a functor to a set as the set of *uncertainty distributions*.

First, however, we need to define future evolutions of the system mathematically. Since the situation of the system at a given time is described by a state, an evolution of the system mathematically is a trajectory of states that the system "passes through in time". The time interval under consideration for assessing an entity's vulnerability is $[t, t_h] \subseteq \mathbb{R}$, as introduced in the previous section. Let T be the set of time points considered, that is, a sequence of say n discrete time steps, $t = t_0 < t_1 < \ldots < t_n = t_h$, or, for a continuous evolution of time, $T = [t, t_h]$, the interval itself. An evolution of the system is a function from time to states, $ev : T \to S$ associating with each point in T the state that the system finds itself in at that particular point in time. In the probabilistic example, Ev denoted trajectories starting in the fixed starting state. To define a vulnerability measuring function v on S, let Ev be the set of all evolutions ev, regardless of their starting point. Not all evolutions in this set need to be considered possible evolutions of the system, in particular, a possible future evolution viewed from state s at time t has to start in s. However, only possible evolutions will occur in the descriptions of the uncertain future as viewed from the starting state. In in this abstract model, we will consider, for example, the harm measuring function h defined on the whole set Ev. That this may be too strong a requirement for practical purposes is not of importance here. In fact, most vulnerability assessments will find out only $v(s)$ for one state, namely the given situation. Yet, to clarify its properties, the mathematical definition will again be a measuring function.

To translate into mathematics the first two primitives, an entity and its uncertain future evolution, we again look back to the probabilistic case in the previous section. There, a probability distribution over evolutions, $P^{(s)}$, was associated with a starting state s to describe the uncertain future of the system as viewed from that state. Here, instead of the probability distribution, an *uncertainty distribution over possible future evolutions* of type $F\,Ev$ gives the corresponding description, where F is the functor chosen to represent uncertainty. Since this description depends on the starting state, we consider given a "future" function $f : S \longrightarrow F\,Ev$ that to a state s associates $f(s)$ of type $F\,Ev$, the uncertain future evolution of the system as viewed from state s at time t.

In the computational model of vulnerability given by Ionescu [2009], this function f, which supplies the uncertainty distribution over possible future evolutions when given a starting state, is computed as the structure of micro-trajectories of a monadic dynamical system. Here monadic dynamical systems are of advantage:

> The computation of possible evolutions is complicated by the need to model the interactions between different types of systems: non-deterministic systems representing scenarios, deterministic models of physical processes or stochastic systems resulting from data analysis. [Ionescu, 2009, p. 46].

Ionescu provides combinations of monadic dynamical systems, but we will not go into any details here.

Back to vulnerability

Having generalized the probabilistic future description to a functorial one, the following steps in defining vulnerability are very close to those described in the previous section.

Taking advantage of the functor, the application of the harm measuring function h can be concisely written as $F\,h : F\,Ev \longrightarrow F\,H$. This means the harm function is applied to each possible future evolution within the uncertainty distribution, while the latter does not change, providing an uncertainty distribution over harm values. This generalizes the probability distribution induced by a random variable.

Note that the functorial description of the uncertain future evolution of a system allows the application of *any* measuring function for harm defined on Ev. The separation of the primitives means that in a vulnerability assessment one of these two components could be exchanged without having to make any changes to the other one. There are no restrictions imposed on either component by the other one, except for the type Ev. In a computational assessment using this structure, the two components could be constructed independently. An especially useful case may be that of exchanging a harm measuring function by a new one, that may result for example from stakeholder involvement. Also, vulnerability could be computed for several measuring functions to show the consequences of different evaluations of harm.

As in the probabilistic example, vulnerability is defined as an aggregation over the uncertainty distribution over harm values. Having seen that aggregation functions can produce examples which do not reflect the ordinary language intuition about vulnerability, some aggregations will be considered appropriate, while others will not. What makes an aggregation function appropriate will however be discussed further below.

Definition 5.1.3. Consider given a function $f : S \longrightarrow F\,Ev$ representing the uncertain future of a system as viewed in each state, two partially ordered sets H and V, a measuring function for harm, $h : Ev \longrightarrow H$, and an appropriate aggregation function $g : F\,H \longrightarrow V$. The vulnerability of the entity under consideration in state s is

$$v(s) = g(F\,h(f(s))).$$

Thus, $v : S \longrightarrow V$ is a measuring function defined on states.

A simple but illustrative example is that of a non-deterministic uncertainty description, so that $f(s) \in \mathfrak{P}(Ev)$, and a harm function that only takes two values, indicating that 'harm has occured' or 'no harm has occured' for every element of the set $f(s)$. That is, h is a predicate, which is the simplest case of a measuring function. Denoting 'harm has occured' by the value 1 and 'no harm has occured' by the value 0, one can define vulnerability as the predicate 'harm has occured in a possible future evolution'.

Definition 5.1.4 (vulnerability as a predicate)**.**

$$v(s) = \begin{cases} 1 & \text{if there is an evolution } ev \in f(s) \text{ with } h(ev) = 1 \\ 0 & \text{otherwise} \end{cases}$$

Ionescu [2009] uses the notation of the functional programming language Haskell, where the aggregation function just used is predefined under the name *any*. Further, instead of f, Ionescu uses *possible* as name for the function supplying the future description to a state. He writes the above definition (point-free and without brackets for function application) as

$$vulnerable = any\ harm \circ possible$$

which reads just like the ordinary language description: an entity is vulnerable if harm is possible in the future. In the general case, the yes/no distinction typical of a predicate is too rough to capture vulnerability, which admits degrees. Let us turn to the question which functions qualify as aggregation functions.

Aggregation

So far, only the type of aggregation function was fixed, $g : F\,H \longrightarrow V$. The aggregation function takes an uncertainty distribution over harm values and aggregates into a vulnerability value, so this is a crucial point where assumptions about the concept vulnerability can be made explicit.

A first condition for making an aggregation appropriate involves the idea of monotonicity. The general intuition is very clear: more harm should mean more, or at least not less vulnerability. It is however not easy to make this idea precise in natural language. What exactly does "more harm is possible" mean, when several future evolutions, possibly with probabilities or other further information attached, are considered at the same time? In a statement like "more harm is likely", one could additionally change the word order to get "harm is more likely". What should vulnerability do in this case?

In the mathematical setting, a minimal monotonicity condition can easily be stated taking advantage of functoriality once again: if the uncertainty distribution stays the same and "inside" it the single harm values increase or stay constant, but do not decrease, the vulnerability value that the aggregation function outputs should certainly not decrease.

Condition 5.1.5 (Monotonicity)**.** *For any non-decreasing function $\alpha : H \to H$, and any y of type $F\,H$ (i.e. any uncertainty description over harm values),*

$$g(y) \leq g(F\alpha(y)).$$

Only functions g with this property qualify as aggregation functions. This monotonicity condition provides a first sensibility check for candidate functions in vulnerability assessments. The aggregation which chooses the most likely harm value, discarded as unintuitive in the previous section, in fact violates this condition, as shown above. Without this check a seemingly natural aggregation such as the most likely harm value could lead to unintuitive vulnerability evaluations.

The above posed questions involving the word 'likely' cannot be discussed in this very general setting, because further uncertainty information than the mere possibilities is necessary. Thus, other properties for the aggregation function may hold only for some type of uncertainty distribution. Calvo and Dercon [2005], for example, discuss vulnerability axioms in the probabilistic setting, where they can pose requirements such as "probability-dependent effect of outcomes" [p. 10]. Adding conditions for certain types of uncertainty distribution may imply that the same aggregation function qualifies with some types of uncertainty distributions but not with others. Maximum possible harm is an example that could be used to aggregate in the non-deterministic case, representing an application of the precautionary principle, but not in the probabilistic case if dependence on the probability measure is required.

Of course, properties of the aggregation may be argued about. An interesting question might be a dependence of the vulnerability measuring function on the time horizon chosen. If the harm measurement is such that it can also be applied to shorter evolutions, theoretically, vulnerability can be measured for any time horizon before t_h. At the least, this reveals that a clearer notation could be useful: with the definitions made so far, all measurement results would simply be 'the vulnerability' of the entity in the same starting state. However, the results by no means have to be the same, when a different time horizon is considered. To enhance clarity, the time horizon could be explicitly attached to the term vulnerability, as for example '10-year vulnerability'. Further, one could ask whether the vulnerability measurements for the same starting state should behave in a certain way in relation to the time horizon considered, for example whether harm should be somehow "discounted" in the further future.

A discussion of mathematical properties that an aggregation function for vulnerability should satisfy can also help uncover assumptions concerning the uncertainty distribution, not only the aggregation itself. An example is the idea that "having more possibilities should not make an entity more vulnerable". Consider an entity with several possible futures and add one more possibility for which the harm measuring function yields very bad values. The question whether the aggregation function should satisfy the above statement may be answered in two ways, following different lines of argumentation. Considering an entity which can make an informed choice, the entity could simply disregard the new possibility, nothing should change. When the possibilities are, however, considered as alternatives that can happen, it would make sense to let the vulnerability measurement increase when the new possibility is added. The different assumptions about how possibilities are considered here correspond to the physical and the knowledge aspects, respectively. These can become explicit in a discussion about formal properties of elements of the formalization.

5.1.4 Remarks on formalization

In the probability chapter, formalization was seen to help raise the discussion in a situation of confusion to another level. The same can be said already at this point for vulnerability.

The concept as used in ordinary language was analyzed into primitives: an entity in a given state, a description of the uncertain future evolution, a notion of harm, and for 'vulnerability to something', a stimulus causing this harm. Vulnerability was mathematically defined as a measuring function, reflecting that the property vulnerability admits degrees. This measuring function is obtained by aggregating an uncertainty distribution over harm values in possible future evolutions of the system. It is rather difficult to state properties that all vulnerability aggregation functions must necessarily satisfy. Still, the (rather weak but not vacuous) monotonicity condition could be stated at this level of generality.

The debate about properties that vulnerability aggregation functions should satisfy can be carried out more clearly than discussing properties of vulnerability using natural language. Also, the precise description of the structure of vulnerability encourages to investigate the working understanding one has of vulnerability: which aggregation function would represent a certain view on vulnerability best? By trying to come up with a mathematical description (especially by choosing the aggregation function), a more precise understanding may be reached. The debate is hence raised to another level.

The precision of the mathematical definition may seem exaggerated for vulnerability as used in ordinary language: here, people will rarely think about explicit possible future situations or specify the evaluation of harm they have in mind. Also, the everyday usage of vulnerability lacks the measurement aspect that is present in the formal definition. When using the word vulnerability in ordinary language, people are generally not concerned with *measuring* vulnerability, although the property admits degrees. One uses statements like "someone is more vulnerable than someone else", when this is obvious. However, the problem of finding out "who of these two is more vulnerable?" is rather unfamiliar in everyday situations. This mismatch is not a problem, since the formalization of the ordinary language term was not meant as an end in itself.

Both the high level of precision and the accentuation of the measurement aspect, which led to the fact that examples from the scientific context came in very naturally, show that the basic mathematical model presented lends itself for investigation of the scientific concept vulnerability. It will be the common background against which the structure of scientific definitions of vulnerability are analysed in the following section. There, the 'mathematical model of vulnerability' will always refer to the model presented here.

5.2 Scientific theoretical definitions of vulnerability

This section presents an analysis of a list of scientific definitions[8] of vulnerability from the literature, reported in the table on pages 83 to 86. In the table, the definitions can still be read line by line when one skips the blank spaces, therefore it was not deemed necessary

[8] We include "descriptions" like "vulnerability refers to . . . " rather than only definitions stating "vulnerability is . . . ".

source		(graded) property		dimen-sions	entity		uncertain future	risk	action		harm		stimulus		other
Timmer-man,	V. is	the degree	to which		a system				acts adversely						
1981			to				the occurence					of	a hazardous event.		
UNDRO, 1982	V. is	the degree								of	loss				
			to		a given ele-ment or set of elements			at risk							
						resulting from	the occurence					of	a natural phenomenon	of	a given magni-tude.
Kates, 1985	V. is						the capacity		to suffer		harm				
						and			react adversely.						
Cham-bers, 1989	V. refers to						exposure to				contin-gencies and stress,				
						and	difficulty in		coping	with	them.				
Dow, 1992	V. is					the differential	capacity								
			of		groups and individuals	to			deal with				hazards	based on	their positions within physical and social worlds.
Cutter, 1993	V. is						the likelihood								
			that		an individual or group	will be	exposed to			and	adversely affected	by	a hazard.		
	It is	the interaction										of	the hazards of place (risk and mitigation)		with the social profile of communities.

source		(graded) property		dimen-sions	entity		uncertain future	risk	action		harm		stimulus		other
Blaikie et al, 1994	The	character-istics	of		a person or a group	in terms of their	capacity to		anticipate, cope with, resist and recover	from the	impact	of a	natural hazard.	It in-volves	a combination of factors
	that deter-mine the	degree	to which		someone's life and livelihood			is put at risk				by a	discrete or identifyable event	in nature or society.	
	V. is	a measure	of		a person or group	's	exposure			to	the effects	of	a natural hazard,		
	inclu-ding	the degree	to which		they		can		recover	from	the impact	of	that event.		
Bohle et al., 1994	V. is best descri-bed as	an aggregate measure	of human welfare that in-tegrates	social, economic and political			exposure to			a range of poten-tial	harmful		perturbations.		
	V. is a	multi-layered and multi-dimension-al social space	defined by the deter-minate	political, economic and institution-al			capabilities								
			of		people										in specific places at specific times.
Cannon, 1994	V. is	a measure of the degree and type				of	exposure to	risk	generated by	differ-ent socie-ties		in rela-tion to	hazards.		
Cutter et al., 2000	V. is						the potential			for	loss of property or life	from	environmental hazards.		

source		(graded) property		dimen-sions	entity		uncertain future	risk	action		harm		stimulus		other
Kuban & MacKenzie-Carey,2001	V. is	a perception				of	the lack of capacity to		defend	against	injury, harm or damage	from	a threat or a hazard.		
Turner II et al., 2003	V. is	the degree	to which		a system, subsystem, or system component		is likely to experience				harm				
						due to	exposure to						a hazard, either a perturbation or stress/stressor.		
UN/ISDR, 2007	V. refers to	the conditions	deter-mined by	physical, social, economic, and environ-mental		factors or processes which increase	the susceptibility								
			of a		community					to	the impact	of	hazards.		
UNDP, 2004		A human condition or process	result-ing from	physical, social, … (as above)		factors which determine	the likelihood and scale			of	the impact	from	a given hazard.		
Twigg, 2004	V.:	The extent	to which		a person, group or socio-economic structure		is likely to				be affected	by	a hazard		
			(related to		their		capacity to		anticipate it, cope with it, resist it and recover	from	its impact.				
Calvo and Dercon, 2005	V. is used as	the magnitude				of		the threat		of	poverty,				measured ex-ante, before
						the veil of	uncertainty								has been lifted.

source		(graded) property		dimen-sions	entity		uncertain future	risk	action		harm		stimulus		other
Adger, 2006	V. is	the state				of	susceptibility			to	harm				
						from	exposure			to	stresses associated	with	environmental and social change		
						and from	the absence of capacity to		adapt.						
IPCC, 2007	V. is	the degree	to which		a system	is	susceptible to, and unable to		cope with,		adverse effects	of	climate change,		including climate variability and extremes.
	V. is	a function				of the character, magnitude, and rate of							climate change and variation	to which	
					a system	is	exposed,								
					its	sensitivity,		and							
						adaptive	capacity.								
Moench and Dixit, 2007	Capa-cities are	the characte-ristics	of		communities and people	which	can be used	to	respond to and cope with		disasters.				

to report the definitions again as simple text. The table can be considered the result of a first step in the analysis: the definitions have been sorted into a common scheme guided by the mathematical model of the ordinary language concept. In the formalization process, this section belongs to step 1, concept analysis. The selected definitions are taken from the field of global change, that is, together with definitions of 'vulnerability to climate change', also definitions of the more general 'vulnerability to environmental change' are included in the list. Scientific use of the concept vulnerability in the fields of, for example, computing science or medicine is however not considered. The choice of definitions was supported by review articles as for example Adger [2006] and O'Brien et al. [2007] as well as by a list provided under www.vulnerabilitynet.org/definitions. The present list is not supposed to be comprehensive. Many more pages could be filled with vulnerability definitions from the literature, starting for example with the 35 definitions collected by Thywissen [2006].

5.2.1 Mapping expressions to the formalization

The scientific concept is defined before the background of the working understanding of vulnerability that people have. If it were not based on the ordinary language use of the word, there would have been no reason to use the term vulnerability. According to the hypothesis that the scientific concept refines the ordinary language concept vulnerability, we use the mathematical model presented in the previous section as a basis for the analysis here. The scientific definitions listed are mostly longer than the ordinary language definition and also seem more complex in that they contain other concepts such as impact, hazard, capacity and so on. The following analysis will show that the scientific definitions generally fit the structure of the basic mathematical definition and refine it. The table groups expressions used in the definitions according to the primitives identified for vulnerability in ordinary language and adds some further columns. This immediately shows that there is a common structure shared by the definitions in the table, and shared with the mathematical definition. The analysis proceeds column-wise. For the most important columns, the expressions used in the definitions are listed here.

column	**expression**
(graded) property	the state; the conditions; the characteristics; a human condition or process; a perception
	a measure; an aggregate measure; a measure of the degree and type
	the degree; the extent; the magnitude
	a function; a multi-layered and multi-dimensional social space
entity	a system, subsystem, or system component; a given element or set of elements
	groups and individuals;a person, group or socio-economic structure; a community; people
	someone's life and livelihood
uncertain future	the occurence; the potential; the likelihood and scale; uncertainty
	the capacity; capabilities; adaptive capacity; can; the lack of capacity; difficulty; the absence of capacity; unable to
	exposure to
	susceptibility
harm	harm; impact; loss; injury, harm or damage; contingencies; stress; adverse effects; loss of property or life; poverty
stimulus	a perturbation or a stress/stressor; the threat; a hazard; a hazardous event; a natural hazard; environmental hazards; the hazards of place
	a natural phenomenon; a discrete or identifyable event;
	environmental and social change; climate change and variation
action	act adversely; defend; resist; cope with; deal with; adapt; anticipate
	suffer; recover

The primitive *entity* is present in most definitions.[9] The expressions used in the definitions can be divided into two groups: those referring to an abstract description of the entity as a 'system' etc, and those referring to a real-world entity such as 'people'. All expressions in the column qualify as entities in the sense of the above formalization, the more abstract ones even suggest a model in terms of a dynamical system. Wanting to study the vulnerability of, say, 'a community' presupposes a description of this community (no matter whether quantitative or qualitative)[10]. Mathematically, such a description of the situation is a *state*.

The primitive *harm* occurs as the expression 'harm' itself or is paraphrased by expressions such as '(adverse) effects' and 'contingencies'. Other expressions specify a certain kind of harm, such as 'poverty', 'loss', or even more specifically 'loss of property or life'. Different aspects are emphasized by the expressions used, but they can all be represented by the harm measuring function in the mathematical model of vulnerability. Whether one considers 'loss of property or life from environmental hazards' or 'adverse effects of climate change', one provides an *evaluation* (again, quantitative or qualitative) of 'harm to the entity triggered by the stimulus'. The mathematical representation should thus take values in a partially ordered set to allow a comparison of some values.

Since the definitions considered are special cases of the construction 'vulnerability to something', also a *stimulus* is given. Two groups of expressions can be made out: a neutral description, such as a '(discrete or identifiable) event', 'a natural phenomenon, or 'climate change and variation', and a description with negative connotation, such as 'a perturbation or stress/stressor'. Some expressions in the second group also convey a possibility of something negative, such as 'a hazardous event', 'a threat', 'a natural hazard'. The ISDR defines

> Hazard: a potentially damaging physical event, phenomenon or human activity that may cause the loss of life or injury, property damage, social and economic disruption or environmental degradation. [UN/ISDR, 2007].

Two of the neutral expressions from the stimulus column recur in this definition: the hazard may be an 'event' or a 'phenomenon'. In the mathematical model, harm (occuring to the entity) is measured in the possible future evolutions of the whole system, confirming the often stated fact that a hazard is not determined by the (natural) event per se, but by the interaction of the event and the entity [e.g., Burton et al., 1993]. The mathematical model also reveals that within a definition of vulnerability it is redundant to call the event a hazard. A substitution of the hazard definition into the vulnerability definition renders a statement of the form: "Vulnerability is the possibility of harm from a *potentially damaging event that may cause harm*."

Expressions found in these three columns suggest to translate the scientific concept vulnerability into mathematics in a similar way as done for ordinary language vulnerability. For the primitive *uncertain future*, the situation is different. Here, the decomposition of the concept vulnerability into other concepts will be seen. Some expressions are already captured by the mathematical model of vulnerability. Certainly, 'uncertainty' is, and 'the occurence', which refers to the possibility of something happening, could be represented by a non-deterministic future description with the two possibilities of occurence or non-occurence of the event under consideration. Further, 'the potential' is a general description of possibility, hence captured by the mathematical model, whereas 'the likelihood' and similar entries ('is likely to experience', 'the likelihood and scale of') refer to a probabilistic description of uncertainty that is also already present in it. However, other expressions in this column deserve a closer look because other scientific concepts from the vulnerability terminology such as exposure and capacity appear. We postpone the analysis in order to deal with some more straightforward remaining columns first.

[9]The definitions which do not specify an entity describe the property abstractly, as in the dictionary definition. This property is, however, attributed to an entity when the concept vulnerability is used.
[10]The same is true even for 'someone's life and livelihood'.

The columns in the table which do not refer to primitives of vulnerability group expressions which play similar roles in the definitions. The column titled *risk* collects expressions which combine two of the primitives into a single word that could thus not be taken apart and sorted where the parts would have belonged: 'risk' ('is put at risk') and 'threat' contain both the uncertain future and a notion of harm. In the case of risk, probabilistic information about the uncertainty is usually assumed. The relationship between the concepts vulnerability and risk is treated in Chapter 5.5. The column *other* was added to collect parts of definitions that do not fit into any group. It will not be discussed.

The column *(graded) property* can be directly interpreted in terms of the mathematical model of vulnerability: the scientific concept is defined as a property by the expressions 'the state', 'the conditions', the 'characteristics', 'a human condition or process' and 'a perception'. This suggests a mathematical representation as a predicate like in Definition 5.1.4. The following expressions additionally state that the property can be measured: 'a measure', 'an aggregate measure', 'a measure of the degree and type', 'the degree', 'the extent'and 'the magnitude'. Therefore, defining vulnerability as a measuring function as in Definition 5.1.3 seems more useful. Theoretical definitions can be read as defining $v(s)$, the result of applying this measuring function to a given state.

The expression 'a perception', found in the definition by Kuban and MacKenzie-Carey [2001], recalls the discussion in Section 5.1.1: the information that a property is perceived adds a subjective element, in that a reference person who perceives it is needed. Subjectivity in relation to vulnerability is not generally discussed as a characteristic of the concept, but for its measurement, see Section 5.4.4.

Thus far, the mapping of the table to the mathematical model of vulnerability has been quite straightforward, and the differences between expressions did not yield essentially different definitions. For example, the concept vulnerability does not seem to change much depending on whether harm is described as an 'impact' or 'loss'. Even when vulnerability is defined diversely as a 'property' or a 'degree', which could be formalized by the different vulnerability definitions 5.1.4 and 5.1.3 respectively, one would rather formalize both cases using the more general Definition 5.1.3. In fact, no definition using the term 'property' states that the property is not comparative, so that the definition as a predicate seems too restrictive. Anyhow, the predicate is a special case of a measuring function.

5.2.2 Uncertainty refined

Apart from the already discussed expressions that are captured by the mathematical model of vulnerability, in the column *uncertain future evolution* one finds specific concepts from the vulnerability terminology such as exposure and capacity, and the expression susceptibility which is related to the scientific concept sensitivity.

A glance at the ordinary language meaning of these terms reveals that an *uncertain future evolution* is a common primitive of exposure, susceptibility and capacity: an entity is 'exposed' to something if this might (but need not) happen to it and 'susceptible' if it may be affected by something happening in the future. 'Capacity' specifies a positively connoted possibility that refers to the entity's actions. Each of the concepts further specifies the uncertain future evolution of entity and stimulus in interaction by considering a certain aspect of it. Here, the refinement of the scientific concept (compared to the ordinary language one) occurs.

The scientific concept is decomposed to explain "how vulnerability comes about". The most basic decomposition of vulnerability is given by Chambers [1989]:

> Vulnerability has thus two sides: an external side of risks, shocks, and stress to which an individual or household is subject; and an internal side which is defencelessness, meaning a lack of means to cope without damaging loss. [p. 1].

Both sides explain why an entity might be vulnerable. Considering that vulnerability arises from the interaction of an entity and a stimulus in the entity's environment, the external side focusses on the stimulus, while for the internal side, the focus is on the entity. Having made such a distinction of two sides of vulnerability, the concept is then defined in terms of concepts that correspond to these sides. In the definition by Chambers (see table, page 83), the two sides occur as 'exposure' and 'difficulty in coping'. Other definitions in the table use other expressions which correspond to an external and an internal side of vulnerability in this sense or, in other words, to a focus on the stimulus and a focus on the entity. The respective concepts 'exposure' and 'capacity' (and also 'susceptibility' and 'sensitivity') are discussed in brief extra sections in the following. First, however, a preliminary remark is due.

In studies of vulnerability to environmental change, "a consistent focus on the social-ecological system" [Adger, 2006], is observed. This common object of study is generally abbreviated SES and also referred to as coupled human-environment system [e.g., Turner II et al., 2003]. It specifies the form of the states under consideration. In most studies, the system of interest contains a social and an ecological component that interact. Further, in many cases, the entity is part of the social system while the stimulus manifests in the ecological system, for example in the case of the vulnerability of a community to climate change. This need not always be the case but a general association of this kind prevails. This has consequences for the two sides of vulnerability distinguished by Chambers: each side shows a focus on the component of the SES predominantly considered when studying stimulus or entity respectively.

Exposure

The concept *exposure* shows structural similarities to the concept vulnerability. Apart from the circularity between the definitions of 'vulnerable' and 'exposed to' found in the Oxford Dictionary of English (see Section 5.1.1), the same grammatical construction is used and the concepts share a primitive.

Definitions in the table use the grammatical construction 'exposure to' either as 'exposure to a stimulus', for example Turner II et al. [2003]: "exposure to a hazard", or as 'exposure to harm due to a stimulus', for example Adger [2006] "exposure to stresses associated with environmental and social change". Within the definition of vulnerability, exposure thus describes the external side, being closely related to the stimulus.

Just as the concept vulnerability, exposure has the uncertain future as a primitive: "exposure to the stimulus" at a time t describes a future possibility of the stimulus manifesting. Recalling[11] the duality of the concept possibility made out by Hacking [1975], possibility here contains both the knowlegde and the physical aspect. It is physically possible that the stimulus manifests, but one does not know whether or when this will be the case. In the mathematical model of vulnerability, exposure can be described as the information contained in an uncertainty distribution of stimulus values or an aggregation thereof.

Only when considering the status in the vulnerability literature is there a big difference between vulnerability and exposure. While most authors define vulnerability, exposure is rarely defined as a scientific concept. For example, neither the ISDR glossary nor the IPCC glossary of Working Group II for the Fourth Assessment Report have an entry 'exposure'. Thywissen's [2006] glossary contains 7 definitions of exposure, one fifth of the number of vulnerability definitions. The IPCC definition from the Glossary of Terms of the Third Assessment Report, "Exposure: The nature and degree to which a system is exposed to significant climatic variations." [McCarthy et al., 2001], is not actually a definition of exposure, which it paraphrases by using the associated adjective. One thus has to assume the meaning of the term 'exposed' given by its working understanding in ordinary language, and this is not precisely determined.

Capacity

The concept *capacity* occurs in the definitions in the table as 'capacity' and 'capability', in the negatively formulated 'lack' and 'absence of capacity' (with 'difficulty' as a weaker

[11] See Section 5.1.1.

negative formulation), as well as in the verb 'can' and the negative counterpart 'is unable to'. The ordinary language meaning of capacity given by the Oxford Universal Dictionary [Onions, 1959] contains the following points:[12]

- the power, ability or faculty for anything in particular
- capability, possibility
- position, condition, character, relation.

In Hacking's duality of the concept possibility, capacity corresponds to the physical aspect of possibility, more precisely to actions an entity is able to undertake. Capacity generally has a positive connotation. If negatively formulated, as in a lack of capacity for example, the concept corresponds to the internal side of vulnerability because it focuses on the entity. The present capacity of an entity refers to future actions which the entity is (physically) able to do. Since one does know what it will do, capacity has an uncertain future as a primitive similar to exposure and vulnerability. One could thus try to describe capacity via an uncertainty distribution over an entity's actions.

The third point in the dictionary definition uses an expression that was found in vulnerability definitions: a 'condition' describes capacity as a property. The conceptual literature emphasizes this present property aspect of the concept capacity, for example Gallopín [2006]: "Capacity of response is clearly an attribute of the system that exists prior to the perturbation" [p. 296]. The uncertain future inherent to the concept, however, often remains implicit. Mathematically, capacity will be a measuring function, but without using an explicit uncertainty distribution.

Often, capacity is further specified in vulnerability definitions. One finds the concepts adaptive capacity, coping capacity, capacity of response and resilience (the latter has been formalized by Hofmann [2007] but is not treated in this work). Definitions of these concepts often state what is meant by the qualifying adjective, while the term "capacity" itself is rarely defined explicitly. An example is the Parry et al. [2007] definition of adaptive capacity:

> The ability of a system to adjust to climate change (including climate variability and extremes) to moderate potential damages, to take advantage of opportunities, or to cope with the consequences. [p. 869].

Here "capacity" is paraphrased using "ability", the meaning of both terms has to be inferred from ordinary language.

The present property aspect of capacity is emphasized also in definitions as "Capacities are the characteristics of communities and people which can be used to respond to and cope with disasters." [Moench and Dixit, 2007]. This definition has the same structure as the vulnerability definitions, in fact, it can be (and was, see page 86) sorted into the table. It is proposed, however, not for a component of vulnerability, but for a concept complementary to it. In the table some vulnerability definitions include capacity while others do not. In some cases, keeping the concepts separate is an explicit aim. For example, the ISDR definition, after the part reported in the table, refers the reader to the definition of capacity: *"For positive factors, which increase the ability of people to cope with hazards, see definition of capacity."* [UN/ISDR, 2007]. And the ProVention Consortium states in this regard

> vulnerability and capacity cannot always be seen as two ends of a spectrum. [...] For the sake of clarity, we therefore propose to confine all the resources and capabilities of communities under the term "capacity" and to restrict the word vulnerability to factors that contribute to putting people at risk. [Davis et al., 2004].

This means the concept capacity, while used as a component of vulnerability in some definitions, is used as a concept on equal footing with the concept vulnerability here. Being

[12] Another part of the dictionary entry refers to a measure of volume or content, which is not of interest here.

considered a concept like vulnerability but "under a change of sign" in the mathematical sense, it is not surprising that this definition fits into the table. Gallopín's [2006] conclusion that concepts "are related in non-trivial ways" [p. 301] is confirmed here. One finds the same concepts declared complementary in one case and one declared a component of the other in another case.[13] Moreover, from a given definition of the concept capacity alone, one cannot deduce whether it is used as a component of vulnerability or a complementary concept to it in that instance.

Having made distinctions between vulnerability and capacity as negative and positive sides of a coin, or between exposure and capacity as external and internal sides of vulnerability, a question that arises is how these interact. There are further concepts in the vulnerability definitions which address some kind of interaction between entity and stimulus: susceptibility and sensitivity.

Susceptibility and sensitivity

The concepts susceptibility and sensitivity are used to describe an interaction between entity and stimulus in some definitions of vulnerability. While for sensitivity as a scientific concept, often definitions are provided, this is generally not the case for susceptibility. Again one is left with the ordinary language meaning, and this proves insufficient for clear conclusions. Sensitivity and susceptibility can be found as synonyms, for example, in online dictionaries. However, one also finds an important difference between the two concepts: 'susceptibility' defined as the 'likelihood of being affected' comprises a notion of uncertainty. 'Sensitivity' on the other hand, does not refer to uncertainty[14] when it is defined as an if-then-relationship or a degree of being affected, as in the IPCC definition: "Sensitivity is the degree to which a system is affected, either adversely or beneficially, by climate variability or change." [Parry et al., 2007, p. 881]. 'Being affected' happens at the time when the stimulus occurs (or later in case of inertia), and sensitivity is plainly a measure of this, it does not have a predictive quality.

Within the vulnerability of an entity to a stimulus at time t, sensitivity throughout each possible future evolution, that is from t to t_h, is of interest. Sensitivity thus plays a role similar to the harm measuring function, as will be seen in the formalization of the scientific concept. It does not show the same structure as vulnerability, capacity and exposure, which contain an uncertain future evolution and hence (could) contain an uncertainty distribution.

The concept susceptibility remains rather unclear. It is not considered a scientific concept, but supposed to be intuitively clear from ordinary language. It is defined neither theoretically nor operationally and seems to occur in some theoretical definitions without further thought as to what it is precisely meant to mean. In the definitions in the table, susceptibility plays different roles. Consider the definition by Adger [2006] and by the IPCC on page 86. In Adger's definition, susceptibility appears at the same level as vulnerability, and seems to derive from exposure and capacity. The IPCC definition decomposes vulnerability into susceptibility and the 'inability to cope'. The second sentence of this definition describes vulnerability as a function of three components. It should be uncontroversial to associate the inability to cope with the (in this case, adaptive) capacity of the entity. This suggests that susceptibility is determined by "the character, magnitude, and rate of climate change and variation to which a system is exposed" and sensitivity. Here, the concepts susceptibility and capacity seem to be on equal footing, while only together they combine into vulnerability. The lacking definition of the concept susceptibility thus cannot be inferred from the role played by the concept in definitions of vulnerability.

5.2.3 Vulnerability composed

Two further columns in the table have to be discussed. Expressions in the column *action* mostly specify capacity, similar to the way in which the stimulus specifies harm. Again,

[13]For vulnerability and risk, each can be found as a component of the other, see Section 5.5.

[14]This is why it was sorted in an empty column in the table.

there are general terms such as 'deal with', 'react adversely' and 'defend'. Other terms express a timing relation with an event: 'cope with' refers to a short-term reaction, 'adapt' implies a longer time scale, while 'recover' relates to the time after harm has taken place. In the list 'anticipate, cope with, resist and recover' all these time scales, plus one that lies before the event under consideration, are gathered. One exception is the definition by Kates [1985], where the 'capacity to suffer harm' shows an unfamiliar use of the usually positively connoted term capacity. A more established formulation would be the "possibility to suffer" harm. Grammatically a verb in the active form describing what the entity does (and therefore found in this column), the meaning of 'suffer' describes the entity as passively experiencing harm, and it could also be sorted into the column harm.

Finally, the column *dimensions* provides more practically oriented information. It specifies which 'factors or processes' to take into account in the system description of a vulnerability assessment, with 'physical' and 'environmental' referring to the ecological subsystem and 'social, economic, political' and 'institutional' to the social part of the system.

The preceding analysis has identified common elements of the various vulnerability definitions and related these to the mathematical model of vulnerability. In fact, these elements, found in the column headings, are the primitives of vulnerability (entity, stimulus, uncertain future and harm), a combination of two primitives (risk), further specifications for some primitives (dimensions, action), and the formalization result of putting the primitives together (graded property). While not all definitions use all elements, all definitions follow the basic structure "vulnerability is a property of an entity consisting in the fact that future harm caused by a stimulus is possible", or, wanting to emphasize the comparability, "vulnerability is an aggregate measure of possible future harm due to a stimulus." The hypothesis that the ordinary language concept vulnerability is a common denominator of scientific definitions of vulnerability is thus confirmed.

Relation between expressions in definitions

How the different elements are linked to each other within each definition also plays a role. Again, the only essential difference seems to be found in the column 'uncertain future'. There are definitions which are at the level of generality of the mathematical model for this primitive: the oldest definitions reported, by Timmerman [1981] and UNDRO [1982], as well as Cutter et al. [2000], UNDP [2004] and Calvo and Dercon [2005] do not specify the uncertainty with an emphasis on either of the entity and the stimulus.

When several concepts are mentioned for this primitive in one definition, these may be considered as being at the same level, as in Chambers [1989]: exposure *and* difficulty, Parry et al. [2007]: susceptible *and* unable to[15], and Bohle et al. [1994] where 'exposure' and 'capabilities' are used in two separate sentences[16].

In other cases, capacity is subordinated to exposure, which "includes" the "degree to which they can recover" in the definition[17] by Blaikie et al. [1994], or subordinated to a general uncertain future description by statement in brackets: "(related to their capacity to ...)" [Twigg, 2004]. Adger [2006] subordinates both exposure and capacity to susceptibility to harm.

In some definitions, only one of the concepts exposure and capacity occurs under the primitive uncertain future. It is clear that capacity does not occur in the definitions of vulnerability by the authors who explicitly want to separate the two concepts, as for example UN/ISDR [2007]. However, other definitions seem to focus on the "external side" or mention the general possibility of harm, without explicitly naming capacity for no obvious reason.

[15] A previous version of this definition, given by McCarthy et al. [2001], differs from this one in one word: where the 2007 definition reads "susceptible to, and unable to cope with", the 2001 version was "susceptible to, or unable to cope with". In the old version, susceptibility and lack of capacity to cope could be understood as alternative definitions, the change of 'or' into 'and' was made to emphasize their interplay.

[16] Note that while under www.vulnerabilitynet.org/definitions the two sentences appear as one definition, in the source these sentences are separated by a about a page.

[17] The three sentences actually constitute two definitions. The first two sentences can be found on page 9, the third sentence on page 57 of Blaikie et al. [1994].

An example in the table is the definition by Turner II et al. [2003]. Vulnerability may also be defined in terms of capacity only, as here present with the definitions by Dow [2003] and Kuban and MacKenzie-Carey [2001]. These differences without a clear motivation enhance the confusion in the terminology.

5.2.4 Conclusions

We draw the following conclusions from the analysis of theoretical definitions of vulnerability: the hypothesis that the everyday language definition represents a common denominator for scientific definitions of vulnerability is confirmed. The same primitives occur and vulnerability is considered a measurable property consisting in the fact that future harm is possible. Different definitions further specify this common basis in various ways. Many of the more specific expressions still allow for a representation by the mathematical model of ordinary language vulnerability.

One particularly important specification is given by splitting the uncertain future into other scientific concepts such as exposure and capacity. In a formalization of the scientific concept vulnerability these should be taken into account. A first step would be to analyze these concepts. Unfortunately, scientific definitions that one could base an analysis on are not necessarily available. Also, the role they play in the vulnerability definition does not provide the necessary information for a formalization. The concept capacity even occurs as a component or as a complementary concept to vulnerability, meaning that it plays very different roles, although similar definitions are proposed for both cases.

That theoretical definitions were found to be similar does not allow any conclusion about how vulnerability is assessed, since "in many assessments the theoretical definition put forward is far away from the methodology applied" [Hinkel, 2008, p. 41].

Considering

- the residual ambiguities in the definitions,
- the fact that theoretical definitions are taken out of context in such an analysis (here as in glossaries)
- and the fact that, as was seen for probability, the interpretation of a concept used can be more easily, and sometimes only, deduced from operational definitions

one can agree with Buckle et al. [2001] who want to "avoid pedantic debate" [p. 7] about definitions. A myriad of detailed theoretical definitions of concepts is around and has not helped to clarify the confusion. One might even argue that by adding another tree to the proverbial forest, each new definition contributes to the confusion. Theoretical definitions leave one with a sense of arbitrariness and seem to be quite irrelevant to the policy context. In this sense, there is no need for *one single definition* of vulnerability.

The mathematical model of the scientific concept vulnerability presented in the following section is based on operational definitions instead.

5.3 A mathematical model of vulnerability to climate change

This section sketches the formalization of vulnerability based on computational assessments, presented by Ionescu [2009]. It then proposes small modifications to better connect the mathematical model to the conceptual literature on vulnerability and clearly discuss of assessment approaches, respectively interpretations of vulnerability. Also, a diagram used for presenting the mathematical model to a non-mathematical audience is given. In this section, words are used to name the newly introduced functions. This convention, more common in computing science than in mathematics, here helps to recall what the many functions stand for.

5.3.1 Ionescu's formalization

Ionescu [2009] introduces the formalization of vulnerability from a computational point of view.

> [...] there is a tantalizing similarity underlying the definitions and the usage of "vulnerability", at least in those studies which are attempting to "project future conditions" by using computational tools, and which can thus be termed computational vulnerability assessment. [p. 2]

Comparing to the activity of writing computer programs, where "one finds oneself using the same computational patterns over and over again, with small yet important differences" [p. 2], he transfers the task of "inventing the proper concepts and tools for describing these patterns as instances of the same structure" [p. 2] to the problem of capturing the commonalities found in computational vulnerability assessments. He hopes for computational benefits such as "more robust design, better code, and code reuse", but also - and this is the more interesting benefit from the point of view of this work - "to gain a better understanding of the concept of vulnerability itself" [p. 3].

Apart from definitions, Ionescu analyzes three computational vulnerability assessments, that operationalize the IPCC definition of vulnerability: Metzger and Schröter [2006], O'Brien et al. [2004] and Luers et al. [2003][18]. While differences in interpretation are pointed out, uncovering the common structure of the various assessments is the main goal of the analysis. At the case study level, where vulnerability measurement methods are described, one encounters operational definitions of vulnerability. These being in mathematical form already, the problem of residual ambiguity, observed for theoretical definitions in the previous section, is essentially eliminated.

Ionescu presents a refined mathematical model of vulnerability for the scientific concept which captures the commonalities found in the analysis. It incorporates mathematical concepts for sensitivity and adaptive capacity. Sticking to the notation used in this work, these are added to the model of ordinary language vulnerability as follows.

To represent that "the estimate of the harm suffered by the entity considered is no longer just a function of the impacts, [...] but also of the sensitivity to the stressor associated with those impacts" [p. 39], separate measuring functions are considered and then combined into the measuring function for harm, $h : Ev \longrightarrow H$.

$$\begin{aligned} &\mathit{impacts} : Ev \longrightarrow I \text{ and } \mathit{sensitivity} : Ev \longrightarrow Q \text{ are combined, that is} \\ &h(ev) = \mathit{impacts}(ev) \odot \mathit{sensitivity}(ev) \text{ with } \odot : Q \times I \longrightarrow H \end{aligned} \tag{5.1}$$

A translation of the dictionary example sentence "small fish are vulnerable to predators" illustrates. Two predicates on the set of possible evolutions are given

$$\mathit{wounded} : Ev \longrightarrow Bool \quad \text{and} \quad \mathit{predators} : Ev \longrightarrow Bool$$

where the first checks whether the small fish have been wounded and the second whether the wounds have been caused by the predators. Then, harm is defined as the conjunction of the two predicates:

$$\begin{aligned} &h(ev) = \mathit{impacts}(ev) \odot \mathit{sensitivity}(ev), \\ &\qquad \text{where } \mathit{impacts} = \mathit{wounded}, \mathit{sensitivity} = \mathit{predators}, \odot = \mathit{and} \text{ so that} \\ &h(ev) = (\mathit{wounded}(ev)) \; \mathit{and} \; (\mathit{predators}(ev)) \end{aligned}$$

Further examples show how the measurements made in the analyzed case studies can be expressed as instances of (5.1). In two studies, the measuring function harm is computed directly out of that for sensitivity, while in the third study, just as in the dictionary example, harm is a commutative combination of the two measurements *impacts* and *sensitivity*.

[18] This last study does not actually put forward the IPCC definition of vulnerability but a definition that "accords with the IPCC definition" [Ionescu, 2009, p. 30]. See, however, Section 5.3.2.

Ionescu [2009] points out that the "symmetry between impact measurements and sensitivity measurements is very similar to the symmetrical way in which the expression 'vulnerability to' is used" [p. 42] and concludes that "in a certain sense, the model of sensitivity outlined [...] justifies, or at least serves to explain, this usage." The formalization hence has explained one of the "conceptual ambiguities" listed by Füssel and Klein [2006]: whether vulnerability "should be defined in relation to an external stressor such as climate change, or in relation to an undesirable outcome such as famine" [p. 305].

The second new element added to the formalization is a mathematical description of adaptive capacity. It is introduced from a computational perspective by Ionescu [2009]. Seeing the difficulty that "the set of possible evolutions is usually not computable" [p. 43], he proposes "thinning the set of results produced by $f(s)$, by considering the evolutions from the point of view of the entity under consideration." [p. 43, adapted].

The analyzed assessments "make use of the assumption that there are a number of 'standard' scenarios which describe the possible evolutions of the climate" [p. 43], derived from the SRES scenarios and/or from statistics. These describe possible evolutions at a macro scale. Ionescu considers the case where each possible evolution can be represented as a modification of a standard evolution depending on the entity's action. He represents the possibilities of action of the entity by a function

$$doable : S \to G\ Actions$$

where G is a functor and *Actions* denotes the type of actions the entity might take. Then, he constructs the function associating the possible evolutions to a state by

$$f(s) = (standard(s)) \otimes (doable(s))$$

where *standard* is a function giving the standard evolutions for a state, possibly returning a different type of uncertainty distribution, but the details are not important here. The operation $\otimes$ combines the uncertainty distributions over *standard* evolutions and *doable* actions.

Then, vulnerability is computed according to the formula from the previous section with this new structure of the function f, starting:

$$\begin{aligned} v(s) &= g(Fh(f(s))) \\ &= g\Big(Fh\big((standard(s)) \otimes (doable(s))\big)\Big) \end{aligned}$$

It is then assumed that the harm measuring function and the aggregation function distribute over the operation $\otimes$, and yield

$$\begin{aligned} &g\Big(Fh\big((standard(s)) \otimes (doable(s))\big)\Big) \\ &= g_1\big(Fh(standard(s))\big) \oplus g_2\big(doable(s)\big) \end{aligned}$$

for aggregation functions g_1, g_2 and an operation $\oplus$ resulting from distributivity. Thus, if distributivity is given, one finds that vulnerability is "a measure of the potential harm registered along standard evolutions, combined with the measure of the local effects due to the actions of the entity" [Ionescu, 2009, p. 44]. This description of the second term warrants to denominate it "adaptive capacity". Thus, with

$$\begin{aligned} stdVulnerability &= g_1 \circ Fh \circ standard \text{ and} \\ adaptiveCapacity &= g_2 \circ doable \end{aligned}$$

one has

$$v(s) = \big(stdVulnerability(s)\big) \oplus \big(adaptiveCapacity(s)\big).$$

In the following, we refer to the $stdVulnerability$ measurement as 'macro-vulnerability'. This formula will be the basis of the diagram presented in Section 5.3.4. The derivation of the adaptive capacity term here will be discussed in relation to the vulnerability literature in Section 5.3.3. Ionescu [2009] further asserts consistency of the definition of "adaptive capacity as a measure of the influence of the actions or courses of action of the entity considered on vulnerability" with the IPCC definitions of vulnerability and adaptive capacity. He remarks that in the climate change community, "adaptive capacity is used to account for the lack of predictive models for the complex systems (especially the social systems) involved" [p. 45] and justifies its use also from the computational point of view: "adaptive capacity is what allows us to actually compute the vulnerability by reducing the set of possible evolutions to manageable proportions." [p. 45].

The formalization clearly presents the refinement of the ordinary language concept vulnerability into the scientific concept that includes the concepts sensitivity and adaptive capacity. The definition is in fact a refinement because it was derived from the definition of the ordinary language concept by considering within this definition a more specific structure for the function f. The decomposition of the vulnerability measurement into the two measurements of macro-vulnerability and adaptive capacity emerged from the mathematical model of the ordinary language concept, and is justified under few additional assumptions, the new structure of the evolutions and a distributivity condition. This gives a possible explanation, which is clear and tangible, for decompositions of vulnerability including these concepts seen in the theoretical definitions in the previous section. In order to include also the concept exposure found there, we will in the following section present what is mostly a different interpretation of the definitions given above.

5.3.2 Exposure and a stimulus measuring function

As a small change to the formalization seen above, we will reinterpret the *impacts* function. In Ionescu's model, the stimulus only enters the definition of vulnerability via the sensitivity function. While this is sufficient to represent the computational assessments considered, the stimulus needs to enter more prominently when we want the mathematical model to capture the decomposition of vulnerability seen in the theoretical definitions and especially in risk assessments (in Section 5.5).

In fact, the stimulus climate (change) appears already in the future evolutions: scenarios describe the possible evolutions of the climate, before harm is taken into consideration. In some assessments, the evolutions are even given in terms of the stimulus only.[19] Otherwise, one can consider the presence of the stimulus for each single possible future evolution within the uncertainty distribution, for example by a function $stimulus : Ev \to D$, that, when applied to an evolution extracts the information about the stimulus. Here, D is a set of stimulus values, and can be assumed partially ordered. For example, descriptions of the climate system are often given in terms of variables such as global mean temperature. In an assessment of vulnerability to climate change an example might be taking the SRES emissions scenarios as possible future evolutions and then using a climate model which computes the state of the climate system on the basis of given emissions. Then the function *stimulus* is constructed by comparing for each scenario the climate model output with some reference climate system state (to obtain climate *change*). The set of the climate change values for all scenarios represents "exposure to climate change".

In the general case, one can describe *exposure* mathematically as the uncertainty distribution over the stimulus values, obtained by applying $Fstimulus$ to the given uncertainty distribution over evolutions (or an aggregation of it). In the case where possible future evolutions are given completely in terms of the stimulus, the identity can be used as *stimulus* function.

[19]See, for example, Section 5.5.2.

Ionescu's example of impacts and sensitivity for the small fish and the predators from the dictionary suggests to measure harm by measuring impacts and then discarding those impact values which were not influenced by the stimulus. This procedure cannot be found, for example, in the description of assessments of vulnerability to climate change by Kelly and Adger [2000]:

> The assessment of vulnerability is the end point of a sequence of analyses beginning with projections of future emissions trends, moving on to the development of climate scenarios, thence to biophysical impact studies and the identification of adaptive options. At the final stage, any residual consequences define levels of vulnerability . . . [p. 327].

Here, the procedure is to first compute levels of the stimulus and – from these – compute impacts. This may occur via sensitivity, as will be seen in Section 5.5.

The formalization by Ionescu is based partly on studies of vulnerability to poverty via the consideration of the work by Calvo and Dercon [2005]. It is there that the consideration of *impacts* and *sensitivity* seems to originate. In fact, what Ionescu classifies as *impacts* in the study by Luers et al. [2003] comes from a measure of well-being, a concept typically enountered in the literature on vulnerability to poverty, and much less frequently so in climate change vulnerability studies. Closeness of the study by Luers et al. [2003] to assessments of vulnerability to poverty is also suggested by the use of the construction "vulnerability to":

> Luers et al., who have both impacts and sensitivity, use "vulnerability to stressors" whose effects are measured by sensitivity most of the time, but slip without comment into "vulnerability to poverty" or "vulnerability to food insecurity" when relating their work to existing literature. [Ionescu, 2009, p. 42].

For describing some studies of vulnerability to climate change and later also risk assessments, we suggest to replace the *impacts* function by a *stimulus* function as described above. Since both are (not further specified) measuring functions, the mathematical definition does not actually change, but the interpretation does. The small fish and predators example can be translated to this setting by replacing the *impacts* predicate *wounded* with a *stimulus* predicate *predators*, and use for *sensitivity* a predicate which establishes whether the predators have *wounded* the small fish, that is, the degree to which the small fish are affected by the predators. The two predicates thus switch roles.

$$
\begin{aligned}
&\text{Still, } \mathit{predators} : \mathit{Ev} \longrightarrow \mathit{Bool} \text{ and } \mathit{wounded} : \mathit{Ev} \rightarrow \mathit{Bool}, \text{ but} \\
&\quad h(ev) = \mathit{stimulus}(ev) \odot \mathit{sensitivity}(ev) \text{ where} \\
&\quad \mathit{stimulus} = \mathit{predators}, \mathit{sensitivity} = \mathit{wounded}, \text{ and } \odot = \mathit{and} \text{ which however yields} \\
&h(ev) = (\mathit{wounded}(ev)) \ \mathit{and} \ (\mathit{predators}(ev)).
\end{aligned}
$$

This is the same harm definition as before, showing that merely a reinterpretation has taken place. Harm is still defined as the conjunction of the two predicates. Now, however, the emphasis is different: sensitivity "filters" the cases where the small fish have been wounded by the predators out of those where predators where present, and not out of all cases where the small fish were wounded.

A general model should contain a general measuring function, *joker*, which can be used for the stimulus or the impacts depending on necessity. This makes the focus on stimulus or impacts in the concepts *vulnerability to climate change* and *vulnerability to poverty*, respectively, more prominent in the mathematical model.

In both cases, the measuring function for harm is combined out of the measuring function for sensitivity and the joker function. The image of symmetry between sensitivity and impact measurements given by Ionescu [2009] can be adapted. There now is a symmetry between stimulus and impacts. One of them is focused upon and, so to say, comes first: in the uncertainty distribution over *joker* function values. Then, sensitivity is applied to determine 'harm due to the stimulus' by computing impacts out of given stimulus values, or by computing the influence of the stimulus in the given impacts. That is, sensitivity plays the role of what was not the joker function. Then, the symmetry is given between the joker and the non-joker, stimulus and impacts, in this order or the other way around.

5.3.3 Capacity

In the formalization of vulnerability by Ionescu [2009] the concept adaptive capacity is used, as in the IPCC definition of vulnerability and many vulnerability assessments. Closely related concepts in the literature are for example coping capacity, and capacity of reponse. We will here refer more generally to capacity. To describe a given assessment, this can be further specified as necessary.

Considering macro-vulnerability and adaptive capacity in relation to the conceptual literature, a difference appears. While both concepts are themselves defined by further concepts in the mathematical model, these are of interest for macro-vulnerability but irrelevant for adaptive capacity. The components of macro-vulnerability are similar to the primitives of ordinary language vulnerability, mathematically, an uncertainty distribution and a harm measuring function. These primitives also occur in vulnerability assessments.

The components of adaptive capacity do not have a direct interpretation in assessments. Doable actions and a function that aggregates into a measurement on the present state can be motivated from ordinary language definition of capacity given in Section 5.2.2, where "possibility" and "ability" as well as the reference to the present state, "condition", occured. In fact, ordinary language might have suggested a formalization similar to that of vulnerability, using an uncertainty distribution over possible future actions and aggregating this into the measurement on the present state. However, a formal definition of capacity which emphasizes these components would miss the point of most if not all vulnerability assessments that treat capacity: this concept is virtually always measured only on the present state.[20] The measurement occurs, for example, via indicators such as GDP per capita, literacy rate and labour participation rate of women (which are examples from the ATEAM project). They measure the state a system is in, "to capture society's ability to implement planned adaptation measures." [Metzger and Schröter, 2006]. Of course, these indicators are chosen because they provide some information about the possible future evolutions and possible actions. A range of the possibilities of action is delimited by measuring criteria that make actions possible. One of Ionescu's examples is capital as an indicator of possible future investments: the level of capital that an entity has at its disposal provides an upper bound to (and, taking zero to be the lower bound, thus a range of) possible investments. The indicator measurement provides the boundaries to the range of possibilities an entity has. However, doable actions and an aggregation do not occur explicitly in assessments.

This is obvious already in Ionescu's [2009] text: the functions *doable* and g_2 are not used when considering the concept adaptive capacity in the case studies.

> In the study of O'Brien et al., "adaptive capacity" was a measure of the current state of the farming units under consideration from the point of view of their ability to adapt to future impacts, thus, a measure of the actions or courses of action available to them.
> [...]
> In the study conducted by ATEAM, "adaptive capacity" was measured in a similar manner to the O'Brien study, therefore a measure of the "doable" actions. [p. 45].

Also, the descriptions of adaptive capacity as a "characterization of the initial state from the point of view of the actions available to the entity and their influence on the potential harm" shows that actions and their influence on the future evolution are implicit in the concept, which explicitly is a characterization of the present state.

The mathematical concepts used to derive adaptive capacity as a component of vulnerability in the mathematical model, the functions *doable* and g_2 are, so to say, for internal use during the formalization only. Once adaptive capacity is established, they become obsolete, for example, when using the mathematical model for literature analysis. While within the

[20] There are measurements of capacity for modelled future states, but these then are measurements of the future capacity as for example in the ATEAM study, see page 103.

formalization, “This usage of “adaptive capacity” is justified if the distributivity condition given above is fulfilled.” [Ionescu, 2009, p. 46], such a condition cannot be checked in actual vulnerability assessments, in which the elements for which the distributivity could hold do not even exist.

Also, a statement like “adaptive capacity is introduced in order to make possible evolutions computable” might not be agreed with by people from the vulnerability community. The concept does not originate from computational assessments, but rather is adopted into them when groups conducting assessments become more interdisciplinary. It serves to shift the emphasis from stimulus focussed assessments towards the entity as described for example by Füssel and Klein [2006] and is motivated by work on vulnerability in the social sciences.

When the concept capacity is used not as a component of vulnerability but as a complementary concept, still the primitives do not become explicit. We return to measurements of this type in Section 5.4.1.

5.3.4 Representation by a diagram

Over the course of the development of the formal framework of vulnerability to climate change, different representations where used to present the framework to different audiences. A diagram has been developed for non-mathematical audiences. The basic idea is to represent the uncertainty distribution by the following sketch of a *non-deterministic* description, yielding the simplest possible diagram:

$$\begin{array}{ccc} & \nearrow & ev_1 \\ s & \rightarrow & ev_2 \\ & \searrow & ev_3 \end{array} \tag{5.2}$$

The interpretation is that from the state s, the system can take one of the three evolutions ev_1, ev_2, ev_3. This simplistic example is useful to illustrate functoriality. The application of a function to all evolutions inside the uncertainty distribution can easily be sketched by denoting a function application in all nodes on the right-hand side.

The diagram for the scientific concept vulnerability follows in a very general version. The use of Greek letters as placeholders avoids introducing yet new terms in the already overloaded terminology. It is used later to compare different assessment approaches. For illustrating the structure of one approach only, the Greek letters can be replaced by function names such as *impacts* and *sensitivity* as used above.

$$\left.\begin{array}{ccccc} & & & \nearrow & \left.\begin{array}{c}\alpha(ev_1) \odot \beta(ev_1) \\ \alpha(ev_2) \odot \beta(ev_2) \\ \alpha(ev_3) \odot \beta(ev_3)\end{array}\right\} \gamma(s_{ECO}) \\ & \nearrow & s_{ECO} & \rightarrow & \\ & & & \searrow & \\ s & & & \cdots\nearrow & \\ & \searrow & \delta(s_{SOC}) & \cdots\rightarrow & \\ & & & \cdots\searrow & \end{array}\right\} \varepsilon(s) \tag{5.3}$$

The first step, producing two states s_{ECO} and s_{SOC} illustrates the foci on one component in the social-ecological system in each part of the assessment (see Section 5.4.1). This step is not essential for interpreting assessments in terms of the framework, as one can set $s = s_{ECO} = s_{SOC}$ for assessments where all measurements are made starting with the same state, and the other way around, in assessments where one finds two parts using different starting states, here interpretable as s_{ECO} and s_{SOC}, an interpretation in terms of one single s could be to consider them together as one state. In that case, each part of the assessment uses only a part of that s. The idea that parts of an assessment can be carried out using different descriptions of the same system, that is descriptions from different points of view, is emphasized by this choice.

α, β and δ are measuring functions, with the respective domains of definition. $\odot$ combines values of the types which α and β output. γ is an aggregation function, that is, a measuring function defined on uncertainty distributions, over the output type of this combination. ε is a measuring function defined on a tuple, with types of the results of γ and δ. Curly braces represent aggregations. These as well as the symbols $\oplus, \otimes$ and $\odot$ are used for generality.[21] Interpretations for the Greek letters are:

α: the *joker* function, that is, a characterization of the stimulus, respectively of impacts, measured in every possible future evolution. If a *stimulus* function is used, the uncertainty distribution over the values of α corresponds to the *exposure* of the entity to the stimulus.

β: the *sensitivity* measuring function
The combination $\odot$ of α and β constitutes the *harm* evaluation in each possible future evolution.

γ: an aggregation of *harm* values in possible future evolutions.

δ: a measurement on the present state, that implicitly concerns an uncertain future (hinted at by the dotted arrows). Above, this is *capacity* and thus implicitly further contains a notion of action of the entity and is positively connoted. In the following section, this measurement occurs also with a negative connotation.

ε: a (purposefully very general) combination of γ and δ.

In the following, we use the abbreviation "ε-assessment" to describe "a vulnerability assessment, with a methodology which can be described by the steps leading to a result of the kind ε in Diagram 5.3". In the next section, different approaches to assessing vulnerability will be distinguished by description in terms of different parts of this mathematical model/diagram and interpretations will be mapped to these. That is, the formal framework will be applied to analyze vulnerability literature.

5.4 Concept clarification

Many vulnerability assessments can be described as special instances of the mathematical model of vulnerability from the previous section. Ionescu [2009] shows this for some assessments which operationalize the IPCC definition. Other than the assessment approach found in the analysed case studies, there are two approaches which do not contain all the above measurements but can be described in terms of γ or δ only.

5.4.1 Three assessment approaches

The first assessment approach, producing a result of type γ, is the traditional one in studies of vulnerability to climate change, also referred to as *impact assessment* [see e.g. Füssel and Klein, 2006]. A description of assessments of this type by Kelly and Adger [2000] has been quoted on page 98, the main points were that vulnerability is assessed by projecting future emissions trends, developing climate scenarios, and studying biophysical impacts and adaptive options. The residual consequences that, according to this description, define levels of vulnerability in the mathematical model correspond to the harm measurements for the

[21] Also, we want to avoid implicit assumptions that symbols familiar in everyday mathematics may lead to, see Sections 5.5 and 5.6.

different evolutions. These may be aggregated into one level of present vulnerability, but this last step is not always made in assessments. In the case of scenarios, that is, a non-deterministic uncertainty description, where the uncertainty distribution simply is a set, formally, the identity function can serve as an aggregation function, when one reinterprets the set as a tuple. Since the harm values for each scenario are (partially) ordered, this yields partially ordered vulnerability values.

The reference to the present state may not be explicit without the aggregation step, but via the identity as aggregation, the formal definition of γ also contains impact assessments as a special instance. A focus on the stimulus is observed in these assessments, where models are constructed to project climate change, while the description of the entity receives less attention. For example, the entity might enter the assessment via a roughly specified assumption about adaptation, such as the metaphorical ‘dumb’, ‘typical’, ‘smart’ or ‘clairvoyant farmer’ described by Füssel and Klein [2006], ranging from no reaction to climate conditions via adjusted management practices to perfect foresight and unrestricted implementation of adaptation.

Because the primitives figure explicitly in the assessment steps when vulnerability is determined from levels of harm in projected possible future evolutions of the system, the mathematical model describes this assessment approach to some detail.

This is different for vulnerability assessments producing a result of type δ. The assessment is conducted by taking measurements on the present state of the system. Often, these are indicators or indices, that is, weighted combinations of indicators. This approach can be seen as a pendant to the measurement of adaptive capacity as described in the previous section, where now the connotation of what is measured is negative. As discussed before, a separation of the primitives within the measurement is not feasible. In fact, vulnerability indicators are supposed to provide information about *vulnerability*, that is, ‘possible future harm’ and are not separable into ‘possible futures’ and ‘harm’.

The same indicator may occur in measurements with positive or negative connotation when the order is reversed, for example, a high level of GDP may be used to indicate a high level of adaptive capacity or a low level of vulnerability.

While for γ a focus on the stimulus was observed, measurements of type δ, both with positive and with negative connotation, focus on the entity. Vulnerability is here sometimes stated to exist “independently of external hazards” [Brooks, 2003, p. 4], or is at least denominated “generic”, see for example the “generic vulnerability” in Brooks et al. [2005] or the ATEAM study’s “generic index of macro-scale adaptive capacity” [Metzger and Schröter, 2006, p. 209].

At the same time, however, it is often pointed out that also the measurement δ depends on the stimulus, for example by Luers et al. [2003]:

> Many factors may determine a system’s ability to modify its vulnerable conditions, including the rate of change of the disturbing forces, and the social and natural capital of the system. For example, if the average temperature increases gradually, farmers may be able to adapt by changing crops or management practices that result in a shift in the well-being function, which would result in lower vulnerability. However, if the temperature change was rapid the same farmers might be limited in their abilities to adapt the changing conditions. [p. 259]

A difference in the motivation of γ- and δ-assessments will be discussed in Section 5.4.4. Before, however, let us sketch the third approach and consider distinctions of interpretations made in the conceptual literature, because these provide further information.

Assessments which produce a result of type ε integrate the two above approaches. Examples (with a positively connoted δ in the measurement of adaptive capacity) are the studies analyzed by Ionescu [2009]. The combination of results of type γ and δ may occur by first aggregating the harm results from future scenarios and then combining this with the (present time) measurement of adaptive capacity, as done in the study of O’Brien et al. [2004].

A different type of combination occurs in the ATEAM study by Metzger and Schröter [2006], where also for the adaptive capacity index the future evolution is estimated using the SRES scenarios. Maps are created to provide a visual combination of potential impacts (represented by hue) and adaptive capacity (represented by saturation). Different maps are built for different time slices and for the different SRES scenarios, that is, no true aggregation γ over the different possible futures, the scenarios, takes place before the combination ε with the adaptive capacity measurement δ is carried out. The result can be considered as several vulnerability assessments for the different time slices, each providing a result ε in which the information for the different scenarios is not aggregated but presented separately by the maps for the different scenarios.

The fact that in all three assessment approaches the term vulnerability is used for the assessment result, immediately makes one source of the confusion asserted in the terminology obvious.

5.4.2 Vulnerability interpretations

In this section, we briefly sketch four examples from the conceptual literature which distinguish different interpretations of vulnerability. The examples are the (conceptual part of the works by) Kelly and Adger [2000], Brooks [2003], O'Brien et al. [2007] and parts from Füssel and Klein [2006]. For details, the reader is referred to the respective original works.[22] Then, the three pairs of different interpretations found will be related to the assessment approaches identified previously.

Starting point and end point vulnerability Kelly and Adger [2000] distinguish interpretations (their term: definitions or conceptualisations) of vulnerability to environmental stress according to whether vulnerability assessment is considered the end point, focal point or starting point of a study. They describe the distinction in terms of assessment methods.

The description of the first case was quoted on page 98: emissions are projected, impacts analyzed and adaptive options identified. In this setting "the level of vulnerability is determined by the adverse consequences that remain *after* the process of adaptation has taken place" [p. 327].

Vulnerability as the focal point of an assessment is found in the food insecurity and natural hazards literature as well as in some climate change studies. It is considered an overarching concept that comprises "the exposure to stress and crises, the capacity to cope with stress, and the consequences of stress and the related risk of slow recovery" [Watts and Bohle (1993) as cited on p. 327].

The vulnerability interpretation that "represents a potential starting point for any impact analysis" [p. 327] focuses on the social dimensions: Kelly and Adger quote the definition by Blaikie et al. [1994] (reported in Table, page 84 as the first sentence), which defines vulnerability in terms of the human dimension alone. They further refer to the usage of the Late Latin *vulnerabilis*, a "term used by the Romans to describe the state of a soldier lying wounded on the battlefield, i.e., already injured, therefore at risk from further attack" [p. 328]. Arguing that vulnerability is defined primarily by prior damage, not by the future stress, they emphasize that

> the vulnerability of any individual or social grouping to some particular form of natural hazard is determined primarily by their existent state, that is, by their capacity to respond to that hazard, rather than by what may or may not happen in the future. [p. 328].

With this interpretation vulnerability assessment "is not dependent on predictions of adaptive behaviour" [p. 327].

[22] Within each paragraph all pagenumbers refer to the work under consideration if not specified otherwise.

This article provides one of the rare examples where the interpretation of vulnerability can be identified reading the short theoretical definition alone. After having sketched the three interpretations, and stated their adherence to the third one, the authors define vulnerability as "the ability or inability of individuals and social groupings to respond to, in the sense of cope with, recover from or adapt to, any external stress placed on their livelihoods and well-being." [p. 328]. The definition, as well as the further description, shows a focus on capacity and the social side, and, in fact, the concept is termed 'social vulnerability'.

Biophysical and social vulnerability Brooks [2003] presents two interpretations (his term: categories of vulnerability definitions). First, biophysical vulnerability, determined by "hazard, exposure and sensitivity" [p. 4],

> has arisen from an approach based on assessments of hazards and their impacts, in which the role of human systems in mediating the outcomes of hazard events is downplayed or neglected. [p. 4].

Second, social or inherent vulnerability, described as the internal state of a system, is "determined by factors such as poverty and inequality, marginalisation, food entitlements, access to insurance, and housing quality" and "is not a function of hazard severity or probability of occurence" [p. 4]. The relation between the two interpretations, "social vulnerability may be viewed as one of the determinants of biophysical vulnerability" [p. 4], is inconsistent with the distinction made. If social vulnerability is considered in assessments of biophysical vulnerability as a determinant, the social system should not seem neglected.

Outcome and context vulnerability O'Brien et al. [2007] base their distinction of outcome and contextual interpretations of vulnerability (their term: interpretation, conceptualization, definition) on the distinction by Kelly and Adger [2000]. They identify a difference in assessments: "Outcome vulnerability is considered a linear result of the projected impacts of climate change on a particular exposure unit (which can be either biophysical or social), offset by adaptation measures." [p. 75]. On the other hand, "contextual vulnerability [...] is based on a processual and multidimensional view of climate-society interactions." [p. 76]. Different framings of the climate change problem, "products of different discourses on climate change - discourses that represent distinct world views and approaches to science" [p. 78] are considered the sources of the different interpretations. The problem in conceptual work that clear cuts in the distinctions made are often lacking is illustrated here: the authors point out the difficulty in assigning institutional analysis to either side of their distinction because it can be found in assessments with both interpretations of vulnerability.

An Evolution of Conceptual Thinking Füssel and Klein [2006] distinguish three schools of thought that each have their own interpretation of vulnerability (their term: model for conceptualizing and assessing vulnerability). First, in the *risk - hazard framework*, vulnerability is the dose-response relationship between an exogenous hazard to a system and its adverse effects. This will be treated in Section 5.5. Second, the *social constructivist framework* regards (social) vulnerability as an a priori condition of a household or a community that is determined by socio-economic and political factors. Finally, a third school of thought, prominent in global change and climate change research, considers vulnerability an integrated measure, which includes an external and an internal dimension. Examples of the third approach are the IPCC definition of vulnerability as well as Cutter's [1996] 'hazards of place' model, which aims to integrate biophysical and social determinants of vulnerability.

Concerning the evolution of assessments, Füssel and Klein [2006] distinguish prototypical assessment stages and illustrate these by diagrams. *Impact assessments* "superimpose future climate scenarios on an otherwise constant world to estimate the potential impacts of anthropogenic climate change on a climate-sensitive system" [p. 324]. *(Climate) vulnerability assessments*, further classified into *first-* and *second-generation* vulnerability assessments, still have a descriptive purpose. While first-generation assessments include some non-climatic factors and consider adaptation potential, in second-generation vulnerability assessments the focus is shifted to the ability of a system to adapt to climate change. These assessments consider climate change and the potential response options in a wider context. Finally, *adaptation policy assessments*

> are characterized by the intensive involvement of stakeholders, by a strong emphasis on the vulnerability of a population to current climate variability, by the formulation and evaluation of response strategies that are robust against uncertain future developments, and by the integration of adaptation measures with existing policies. [p. 324].

5.4.3 Correspondence: interpretations and assessment approaches

This section relates interpretations of vulnerability to assessment approaches. Three distinctions just sketched, starting-point and end-point vulnerability, biophysical and social vulnerability, and outcome and contextual vulnerability, can be mapped to the distinction made here between the two assessment approaches: assessments of end-point, biophysical and outcome vulnerability produce a result of type γ while starting-point, social and contextual vulnerability assessments can be described in terms of δ.

Assessment methods and time aspect

With the mathematical model of the scientific concept vulnerability, assessments can be distinguished by their structure, that is, the steps undertaken and methods applied. A result of type γ is obtained by projecting possible future evolutions, measuring harm for each of these and (possibly) aggregating. In such an assessment, the focus is on the future time aspect of vulnerability. Obviously, for the measurement on the present state, δ, the focus is on the present time aspect.

The three distinctions of vulnerability interpretations under consideration all mention the difference between assessment methods as well as the focus on one time aspect for each interpretation. The qualifyers "outcome and contextual" chosen by O'Brien et al. for the two interpretations hint at both criteria: future outcomes are projected, or a present context is analyzed in the assessments of the respective types. Füssel and Klein [2006], in their list of "conceptual and semantic ambiguitites", pose the question whether vulnerability "is an inherent property of a system or contingent upon a specific scenario of external stresses and internal responses" [p. 305]. An "inherent property" would suppose a measurement of type δ, while the measurement result γ is "contingent" upon the scenarios used in modelling the uncertain future.

The future time aspect dominates in assessments of end-point, biophysical and outcome vulnerability, as clearly expressed for example by O'Brien et al. [2007] for outcome vulnerability: it "directs attention towards future impacts of climate change, rather than towards present vulnerability" [p. 80]. For starting-point, social and contextual vulnerability, on the other hand, it is emphasized that vulnerability is a property of the entity at the present time. For example, Kelly and Adger [2000] highlight the fact that vulnerability of people to a hazard is "determined primarily by their existent state, that is, by their capacity to respond to that hazard, rather than by what may or may not happen in the future" [p. 328]. The present state is the object of analysis, which entails a "shift of focus away from the speculative future." [Kelly and Adger, 2000, p. 329]. However, that the future is implicitly considered is obvious for example when Eriksen and Kelly [2007] suggest "a process-based approach to indicator studies" of vulnerability instead of "aggregating static indicators of local conditions" [p. 496]. A *process* cannot be defined on one single point in time, and clearly, the time of interest in a process-based indicator of vulnerability is the future, because indicators of vulnerability should be indicators of possible future harm.

Entity, stimulus, and components of the SES

The two assessment approaches producing results of type γ and δ have as focal points in assessments the stimulus and the entity, respectively. They can thus be mapped also to the two sides of vulnerability in Chambers' 1989 distinction, γ to the external and δ to the internal side. Since, in studies of vulnerability to climate change, the stimulus manifests in the ecological component of the SES, while in most cases the entity under consideration is part of the social component, these focal points coincide with a focus on the ecological system for γ-assessments and on the social system for δ-assessments. The qualifying adjectives chosen by Brooks, "biophysical" and "social", highlight these focal points in the two interpretations of vulnerability.

End-point, biophysical and outcome vulnerability are said to emphasize the ecological component of the system and the stimulus, here the hazard:

> The hazards and impacts approach typically views the vulnerability of a human system as determined by the nature of the physical hazard(s) to which it is exposed, the likelihood or frequency of occurrence of the hazard(s), the extent of human exposure to hazard, and the system's *sensitivity* to the impacts of the hazard(s). [Brooks, 2003, p. 4]

This occurs at the expense of the social aspects and the the human dimension, "rather neglected in past studies..." [Kelly and Adger, 2000, p. 328], or in the formulation of O'Brien et al. [2007], "the focus is disproportionately on nature [. . . , while] society is typically represented as one box . . . " [p. 76].

The social system is the focus of starting-point, social and contextual vulnerability. "In this formulation, vulnerability is something that exists within systems independently of external hazards." [Brooks, 2003], or, in less explicit terms, "[v]ulnerability is considered to be influenced not only by changing biophysical conditions, but by dynamic social, economic, political, institutional and technological structures and processes [. . .]" [O'Brien et al., 2007, p. 76]. Here, the stimulus merely "sets the context for the study" and it is not considered necessary "to define precisely the nature of the potential impact" [Kelly and Adger, 2000, p. 328].

Interpretations are here distinguished in terms of focal points of assessments. Rather than by an exact criterion, the difference is made by the degree of detail to which stimulus and entity, or the respective parts of the social-ecological system, are represented, and by their (relative) importance in the assessment. The separation cannot be strict because the interaction between stimulus and entity is essential for determining the entity's vulnerability to the stimulus. Hence, it is frequently acknowledged in the conceptual literature that the point which is not focussed upon cannot be completely disregarded. Even for assessment results of type δ, "it is intrinsic to the definition that vulnerability must always be linked to a specific hazard or set of hazards" [Kelly and Adger, 2000, p. 327]. Similarly, results of type γ are "contingent on assumptions about concurrent socio-economic developments" [Füssel and Klein, 2006, p. 306], that is, the social system.

Two combinations

Having seen two different levels at which the distinctions between interpretations of vulnerability are made in the conceptual literature, theoretically, four combinations of focal points would be possible: stimulus-future, entity-future, stimulus-present and entity-present. However, only the first and last one exist: a focus on the stimulus and the future aspect of vulnerability corresponding to results of type γ, and a focus on the entity and the present time aspect, corresponding to results of type δ.

The two interpretations of vulnerability related to these have been ascribed to different "framings of the climate change problem" by O'Brien et al. [2007]. These authors embed outcome vulnerability in a scientific and contextual vulnerability in a human-security framing. Differences between the framings show in "prioritized questions, focal points, methods, results and proposed responses." [O'Brien et al., 2007, p. 78]. Different disciplinary backgrounds and problem contexts such as natural hazards or food security have been mentioned as sources of the different interpretations of vulnerability.

Without going into detail on the origins of these interpretations here, the mathematical model provides one simple explanation for the given association of focal points: the difference between assessments producing a result of type γ or δ is in the methods delineated. Possible future evolutions are made explicit only in the assessment approach with a focus on the ecological component of the SES. For this component, accepted theories of the system evolution, which can be implemented in models, are much more readily available than in the case of the social system, and it may not even be desirable to project people's actions in assessments. In fact, social vulnerability, that is, the interpretation corresponding to measurements of type δ, emerged in the context of disaster risk[23] as the main concept to explain the social causation of disasters [see Cannon, 1994]. Recognizing vulnerability when considering the present state of a system is essential in this approach: the fact that "it is vulnerable people who are the victims of disasters" [p. 17] is pointed out not to be a tautology by Cannon [1994] who further explains:

> The purpose is to demonstrate that there are particular characteristics of different groups of people (derived from economic, social and political processes) which mean that with the impact of a particular type of hazard of a given intensity, some avoid disaster and others do not." [p. 17]

In conclusion, the time aspect focused upon appears clearly in the assessment methods used. The methods again are closely linked to the focal system component of the SES: for the ecological component and the stimulus, models are available, while for the social component and the entity, measurements on the present state are preferred. That is, the assessment approach describing the kinds of methods used can make the link between the focal points observed together.

5.4.4 Measurement

Especially with the interpretation of vulnerability corresponding to the δ-assessment approach, measuring vulnerability has been described as problematic in the literature. Cannon [1994] notes a "great difficulty in accurately determining levels of vulnerability" [p. 29] and Adger [2006] states that researchers "struggle to find suitable metrics for vulnerability" [p. 274], for example. Luers et al. [2003] list some problems for vulnerability indicators, which are measurements of kind δ:

> While the indicator approach is valuable for monitoring trends and exploring conceptual frameworks, indices are limited in their application by considerable subjectivity in the selection of variables and their relative weights, by the availability of data at various scales, and by the difficulty of testing or validating the different metrics. [p. 257].

Subjectivity connects back to the questions discussed in Section 5.1.1: here, the existence of an objective vulnerability seems to be assumed, but it is clear that measurements are subjective. Therefore these measurements are not considered adequate. For the case of intelligence, Henshaw [2006] notes that measurement "provokes strong emotional reactions" which are not mainly caused by the attempts at measurement themselves, "it's how those measurements are *used*." [p. 89], namely in decisions involving high stakes, for example the admission of a student to a certain school. Similarly, vulnerability measurements may be used in decisions with high stakes, such as adaptation funding decisions. Measurements which are perceived as more subjective than others are controversial in such contexts.

The "difficulty of testing or validating the different metrics" mentioned by Luers et al. [2003] can, from a theoretical point of view, be attributed to uncertainty: vulnerability measurements cannot be validated even in principle by observing the evolution of a system

[23] See also Chapter 5.5.

due to the primitive possibility, as discussed in Section 5.1.1. The observation does not fully describe the prior possibility. This would not be a practical problem here, because the actual interest in vulnerability arises from the interest in future harm as seen in the quote by Cannon [1994] on page 107. If by reducing whatever is measured under the label vulnerability, disaster could be generally avoided, the interest in what the *possibility* of disaster was at the time of the measurement can be expected to be minimal. The 'a posteriori' measurements of harm, however, do not satisfactorily measure vulnerability 'a priori' in practical situations. Luers et al. [2003], for example, state that

> In certain case studies, depending on the type of stressor and outcome variables of concern, the relative impacts of stressors in a region could be used as objective ex-post measures of vulnerability. [...] This simple approach, although useful, is not easily applied to a greater variety of stressors and outcome variables. [p. 256]

Further, even considering that one is not interested in previous possibilities, measurements of vulnerability to climate change are hard to test or validate. One cannot run laboratory experiments, and the time horizons in climate change vulnerability assessments do not allow to just wait for observations that would validate measurements.

Returning to vulnerability measurements in general, Eriksen and Kelly [2007] name two criteria for indicator selection: theory and statistics. A generally agreed theory of vulnerability is not available, but steps towards a theory of vulnerability have been made. These are mostly found in connection with the δ-assessment approach. Vulnerability research with focus on the present and the social system investigates "underlying causes" [O'Brien et al., 2007] or "root causes" [Blaikie et al., 1994] of the aggregate property vulnerability.[24] It aims at developing a *theory of vulnerability.*

Using an indicator to measure vulnerability, the relation between the indicator and vulnerability is presupposed. For example, low values of wealth and education, but high values of income inequality are generally associated to high vulnerability. Adger [2006] remarks that "unless the variable and causal links are well established, the relationship may not hold." [p. 275]. This may be problematic: the fact that no generally accepted theory exists means precisely that causal links are not well established. While there seems to be agreement that "certain factors [...] are likely to influence vulnerability to a wide variety of hazards in different geographical and socio-political contexts" [Brooks et al., 2005], and these "generic" determinants of social vulnerability [Brooks, 2003] provide a step towards a general theory of vulnerability, the differences between different cases is often emphasized in the literature.

The mathematical model can be used to illustrate the problem: in indicator measurements (or more generally, in δ-assessments), the primitives are implicit in the measurement. Using theory or statistics to select an indicator, one adapts measurements from one case to another or from the past to the future. However, the primitives are known to differ between cases. For example, O'Brien et al. [2007] point out that "what constitutes 'damage' or 'negative effects' varies across contexts and cultures." [p. 77]. In the aggregate measure, exchanging "gardening in England" for "skiing in Norway", to use their example, may be difficult.

These difficulties become apparent in many case studies: indicators or determinants are proposed as measures of vulnerability but accompanied by counterexamples where the causal links do not hold, or are even reversed. One of many instances is inequality, used as an indicator of collective social vulnerability by Adger [1999] who at the same time points to the fact that "it is argued that under certain circumstances inequality facilitates provision of services for the good of communities by those with cumulated assets." [p. 255]. Phrases like "vulnerability changes from place to place" or "vulnerability can change over time" [Cutter, 1996] summarily acknowledge the problem: the point is not that levels of vulnerability differ but that the causal structure of vulnerability does.

[24] See Section 5.5.2.

In γ-assessments that use modelled futures, a theory about how the system is going to develop is present. Together with a given evaluation of harm and an aggregation, vulnerability can be deduced from the model results. However, since a model of the system evolution is specific to each case, coming up with a general theory of vulnerability would exceed the possibilities of a single assessment. Knowledge about vulnerability "lying scattered" in different case studies is an accurate description of the situation by Hinkel [2008].

We conclude this section with a remark on the formalization of γ- and δ-assessments: the mathematical model describes the measurements made in γ-assessments to some degree of detail, the primitives of vulnerability can be found in assessments. The refined mathematical model accounts for the decomposition of harm into sensitivity and another component, the characterisation of the stimulus or impacts. While the aggregation step may be missing in assessments, in this case the identity can play the role of the aggregation function. For δ-assessments, the description given in the mathematical model is very rough. The primitives of vulnerability occur in the measurements undertaken only in an implicit manner. The derivation of adaptive capacity by Ionescu [2009] explains the implicit primitives which helps in concept clarification. However, the model does not provide information about how measurements are undertaken, except that these are applied to the present state.

5.4.5 Confusion explained

The formalization of vulnerability has been useful in reducing six vulnerability interpretations to two, by linking them to assessment approaches. There are further smaller points where it can help clarify issues, some examples are given in the following.

The distinctions made in the literature have not always been clear and consistent. Criteria for distinguishing the interpretations concern focal points of assessments, thus the distinctions often are not crisp. For example, Füssel and Klein [2006] refer to the four assessment stages they distinguish as "prototypes" and note that "actual climate change assessments may well combine features from more than one stage." [p. 309]. Of course, also a mathematical model cannot provide crisp cuts where there are none.

However, in some cases, the distinction can be made more precise. For example, O'Brien et al. [2007] mention that in their distinction between outcome and contextual vulnerability

> indicator approaches to vulnerability [...] may be associated with either type of study. Indicator studies can be used both to enhance understanding of the causes of vulnerability and to quantify the extent of the problem. [p. 80].

Having made explicit the two time aspects of vulnerability, the mathematical model can be used to distinguish where within a study an indicator is used. On the one hand, there are indicators of harm, applied to measure harm in the possible future evolutions. An example is provided by the DIVA tool [see Hinkel and Klein, 2007], a model which for different choices of SRES scenarios and adaptation policies outputs values for many indicators such as "people flooded", "wetlands lost" etc, measured for the modelled possible evolution of the system. These are the indicators used to "quantify the extent of the problem" in O'Brien et al.'s formulation. On the other hand, indicators of vulnerability are applied to the present state, as described previously. By considering what an indicator measures, namely in one case harm (in a possible future evolution) and in another case a possibility of future harm (in the present state), it can be associated to an assessment approach.

The borders between the interpretations are blurred by assessments from the integrated approach, which use measurements of both type γ and δ. The attempt to classify an ε-assessment, given the distinction of only two interpretations corresponding to γ- and δ-assessments, poses a problem. The IPCC definition of vulnerability illustrates: O'Brien et al. [2007] classify it as of type γ: "An example of an end-point definition can be found in the IPCC Third Assessment Report ..."[25]. At the same time, for Füssel and Klein [2006],

[25] See footnote 15 about the connector 'and' in this definition.

> Vulnerability, according to the IPCC definition, is an integrated measure of the expected magnitude of adverse effects to a system caused by a given level of certain external stressors. [...] Vulnerability, [...] includes an external dimension, which is represented here by the 'exposure' of a system to climate variations, as well as an internal dimension, which comprises its 'sensitivity' and its 'adaptive capacity' to these stressors. [p. 306].

This describes a measurement of type ε. Interestingly, sensitivity is here classified into the "internal dimension" of vulnerability, while in the long quote from Brooks [2003] on page 106, as well as in the mathematical model, it appears within the γ-part of the intergated (that is, ε-) assessment, which corresponds to the external side.

The inconsistency identified in the conceptual framework by Brooks [2003] in Section 5.4.2 also seems to owe to the non-consideration of the integrated approach. The statement "social vulnerability may be viewed as one of the determinants of biophysical vulnerability" comes as a surprise after having mapped social vulnerability to our δ and biophysical vulnerability to our γ. The inconsistency can be avoided by considering social vulnerability (and just as well biophysical vulnerability) as a determinant of vulnerability in ε-asssessments. This is probably intended by Brooks [2003] who also states that "it is the interaction of [the] hazard with social vulnerability that produces an outcome ..." [p. 4].

The formalization has made explicit the two time aspects of the concept vulnerability in a present state s and possible future evolutions ev. Moving from left to right in Diagram 5.3, a step from the present to the projected future is made, but the aggregation yields a result about the present state. That in some assessments the present vulnerability, $v(s)$, is measured based on projected future harm seems to cause some confusion with the intuition that vulnerability causes harm. In the measurement, the order is reversed: projected harm determines present vulnerability. Füssel and Klein [2006] explain their diagram representing "first-generation vulnerability assessment":

> The arrow that points from *impacts* to *vulnerability* is different from most of the other arrows. This thin arrow indicates that the potential impacts of climate change on a particular system (in concert with its adaptive capacity [...]) determine the vulnerability of that system to climate change. However, it does not suggest that impacts cause vulnerability. We note explicitly that the direction of causation shown in [the diagram] would be different if vulnerability to climate change were defined according to the first or second school of thought ... [p. 317].

The "schools of thought" that are referred to are the ones described in Section 5.4.2: the risk-hazard framework with vulnerability as a dose-response relationship and the social constructivist framework with (social) vulnerability as of type δ. The fact that the authors deem it necessary to explain this difference between how something is measured and the direction of causation suggests that the diagram alone might be misunderstood by people from the first two schools of thought.[26] The multitude of diagrams found in the conceptual literature on vulnerability has not been discussed in this work. It can be seen as source of confusion similar to theoretical definitions, see for example Füssel and Klein [2006] for the different (implicit) interpretations that can be given to arrows.

All these examples discussed interpretations and assessment approaches of vulnerability. Adding theoretical definitions and the results from the analysis in Section 5.2 to the picture, a basic result is a gap between the two, described by Hinkel [2008] as follows:

> In most case studies, operational definitions are not derived systematically from the theoretical ones and the relation between the two often remains obscure. [p. 46].

[26] The formulation of the last quoted sentence seems to suggest a difference concerning 'the direction of causation' between the first two and "the third school of thought" in the climate change context. This is unfortunate, since the point was not a difference between schools, but the difference between causation and measurement within the third school. An "arrow of causation" would have the same direction for all schools.

Theoretical definitions usually cannot be shown to correspond to one interpretation or assessment approach of vulnerability as seen above for the IPCC definition. The conceptual literature distinguishes interpretations in terms of focal points, between the future and the present time aspect and related assessment methods, as well as between the stimulus and the entity and the related component system of the SES. All aspects are present in all interpretations, only the emphases differ. Theoretical definitions try to capture all aspects of the concept. Being short statements, they do not provide the information which aspects are emphasized in the interpretation of vulnerability a researcher has in mind.

Even when definitions of vulnerability use only one of the terms 'capacity' and 'exposure', which, qua focal point, could be associated with a δ-like and a γ-like interpretation of vulnerability, respectively, one cannot deduce that the definition corresponds to that interpretation. Of course, one has to pay attention when reading a definition that explicitly considers capacity apart from vulnerability, as often found in assessment guidelines and discussed in Section 5.2.2. However, a striking example that has nothing to do with this is a definition by Adger [1999]:

> *Social* vulnerability is the *exposure* of groups or individuals to stress as a result of social and environmental change, where stress refers to unexpected changes and disruption to livelihoods. [p. 249, our emphasis].

Theoretical definitions were seen to refine the ordinary language concept vulnerability. This certainly contributes to their being similar, despite being used with different interpretations. The similarity may actually blur or hide differences between interpretations and assessment approaches by suggesting that a common understanding has been reached. Here, the warning by Newell et al. [2005] to be "wary of superficial approaches to developing 'better communication' that only appear to remove conceptual confusion" [p. 301] seems appropriate. While the same terms are used in definitions of vulnerability, when addressing the different assessment approaches, "different speakers have different models and examples in mind" [Newell et al., 2005, p. 301].

Imagine a quiz that consists of associating a theoretical definition of vulnerability with the interpretation that it is used with. A person who is a little bit familiar with the literature on vulnerability to climate change would probably have great difficulties when presented only the exact wording of the definition. Adding the author's name would greatly simplify the task.

A further source of confusion for both theoretical definitions and assessment approaches of vulnerability is a certain arbitrariness arising in any of the classifications undertaken. The concept vulnerability has been decomposed, into components in theoretical definitions and into intermediate results in assessments. Different decompositions are not nested. The three concepts exposure, sensitivity and adaptive capacity, for example, cannot be sorted unambiguously into Chambers' decomposition of vulnerability into an external and an internal side. This does not prevent authors of conceptual work from doing so, and, as was seen for the example of sensitivity above, different authors choose different classifications.

Similarly, arbitrariness exists in relating observed phenomena and vulnerability components. Füssel and Klein [2006] provide the example of

> the vulnerability to flooding of a country that experiences significant internal migration from the highlands into the flood plains. This migration changes the exposure of certain population groups to flooding events. Aggregated to the country level, however, the effects of migration represent changes in the sensitivity of the population to flooding events. [p. 317].

Components of vulnerability distinguished at the theoretical level do not map unambiguously to measurements in assessments at the operational level. Again, the gap is confirmed. Operationally oriented documents such as assessment guidelines often do not even consider the most basic decompositions of vulnerability at a theoretical level. Very detailed subdivisions of vulnerability into factors to be considered in an assessment are presented without mentioning for example sensitivity by Twigg [2004], Davis et al. [2004] and Buckle et al. [2001].

Lastly, the interpretation of a case study in terms of the mathematical model of course is not free from arbitrariness either. An example is the ATEAM study, which by Ionescu [2009] has been interpreted to have no *impacts* function but compute harm simply from sensitivity. In the model presented here, where a *stimulus* function is used instead of the *impacts* function, the interpretation could be revised. Harm is combined out of climate change values and sensitivity, where the climate change part was included in the future evolution in the interpretation by Ionescu [2009].

5.5 Risk

In this section, the concept *risk* as used in the the domain of natural hazards and disaster management is related to the formal framework of vulnerability. It shall be seen that the concept risk plays a role that is rather similar to the role of vulnerability in the climate change context. The fact that both concepts occur as components in each other's definitions contributes to the confusion.

Natural hazards are one of the antecedent traditions of vulnerability [Adger, 2006], and similarities exist not least in the common focus on the social-ecological system. The fields of climate change adaptation and disaster risk management are growing closer as is documented e.g. by the Red Cross/Red Crescent [2008] Climate Guide. Efforts are made to bring together researchers and practitioners working in the context of vulnerability and hazard assessment to foster a "better integration of the concepts in order to strengthen the progress towards sustainability" as could be read in the conference announcement[27] of the conference 'SHIFT07 – Shift in Thinking – Perspectives of Vulnerability and Hazard Assessment'. Parts of this section were submitted for a Special Issue on this conference. In the article, the focus was on clarifying the relations between approaches to assessing risk and vulnerability, using a graphical framework based on the formal framework of vulnerability. Mathematical formulations were explicitly avoided as a response to a review of a former version of the article (see Section 5.6 for detail).

Here, the graphical framework is introduced by displaying the ordinary language concept risk as an example. Then, this framework is used to analyse risk assessment approaches using previous conceptual work and a case study in Section 5.5.2. Resulting similarities and differences in the assessment approaches of the two research communities are discussed in Section 5.5.3.

5.5.1 Graphical framework

Based on the formalization of vulnerability presented before, diagrams illustrate the elements of the mathematical model of vulnerability. Comparing to Diagram 5.3, the form stayed the same, but placeholders and mathematical symbols were replaced by words from ordinary language, and braces were replaced by arrows. This formerly formal framework was then presented as displaying *how vulnerability (or risk) is assessed* by representing steps that are undertaken. In fact, as seen before, at least for some approaches, the formalization primitives of vulnerability occured as elements in assessments, for example as intermediate results. Therefore, the formalization also shows the structure of assessments.

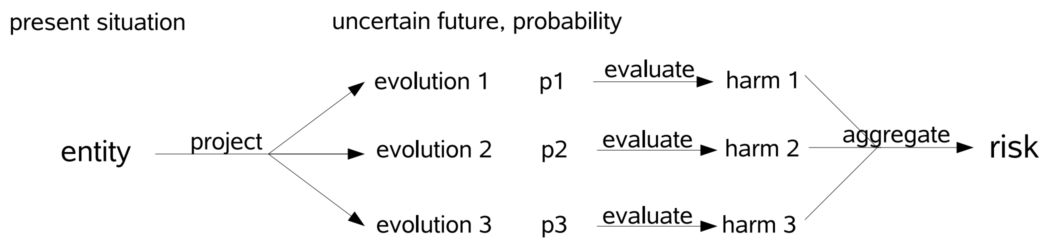
In ordinary language, just as vulnerability, risk expresses a possibility of future harm. The Oxford Dictionary of English [Soanes and Stevenson, 2005], for example, defines:

- a situation involving exposure to danger [...]
- the possibility that something unpleasant or unwelcome will happen [...] .

[27] http://www.pik-potsdam.de/events/scenario/conference-program.html

The first sentence mentions the 'present property' aspect by describing risk as a "situation", while the primitives of vulnerability, an uncertain future and a notion of negativity, are found in the second sentence. The circularity in definitions observed for 'vulnerable' and 'exposed' in this dictionary similarly applies to 'risk', which is here defined in terms of exposure. The definition of 'expose' ("cause someone to be vulnerable or at risk", see Section 5.1.1) is also based on the understanding of the word risk.

The graphical representation of the concept risk is the following:



The diagram illustrates which steps can be undertaken to measure risk. Arrows represent the steps, and their shafts are labelled with a name for the method applied in the step. The arrows point from the input that the method is applied upon to the results of the method. An arrow may branch out, meaning that a method produces several results from one input, and several arrows may converge into the same head, which means that a method is applied to several inputs and produces one result.

The added sketch of a probability distribution is included to represent the fact that risk assessments usually consider a probabilistic representation of the uncertain future. This is consistent with the concept risk as coined by Knight [1921] who distinguishes risk, in the sense of a measurable probability, from uncertainty where no probability can be measured. The interpretation of probability here must be an objective one, since for the subjective interpretation, probability could always be measured.

In scientific contexts, risk encounters similar conceptual difficulties as vulnerability. There are - no surprise here - various and quite different definitions of risk. They generally involve more mathematical terms than definitions of vulnerability, without however being mathematical definitions. A large (and much older) body of literature on risk and problems involved in measuring it [e.g., Knight, 1921; Adams, 1995] has not led to a general agreement on precise definitions of the terms used. In the context of natural hazards, 'risk' is defined in terms of 'probability', 'expectation' or 'probability and consequence', and is said to depend on 'hazard, exposure and vulnerability', or 'hazard and vulnerability' only. However, it seems to be recognized as one and the same concept more easily than vulnerability, and in spite of the Babylonian confusion that this component of risk is so generously certificated.

5.5.2 Risk assessment approaches

Given the very similar point of departure, the mathematical model of vulnerability lends itself to describing approaches to risk assessment. In the following, this is shown using the graphical representation. We analyze the pressure and release (PAR) model by Blaikie et al. [1994] (Section 5.5.2) and the Risk Triangle by Crichton [1999] (Section 5.5.2). Being proposed as conceptual models, these can and will be interpreted as rather general descriptions of *how risk can be assessed*. In the latter case, additionally to the short conceptual description, a risk assessment which operationalizes the approach is considered in Section 5.5.2: the Flood Risk Maps developed by the Institute for Environment and Sustainability at the Joint Research Centre [Barredo and de Roo, 2007].

The PAR model

The "pressure and release" model describes disaster as the intersection of processes creating vulnerability and exposure to hazards, described as pressures, while 'release' refers to reducing vulnerability. Risk is a "complex combination" of hazard and vulnerability, such that the result is nil if one of the components is nil. Also, risk is said to arise from the interaction of hazard and social conditions. This interchangeability of 'vulnerability' and (current) 'social conditions' in the description of risk suggests that Blaikie and colleagues use 'vulnerability' with a strong focus on the social system and on the present state. In fact, Blaikie et al. [1994] describe a progression of vulnerability from *root causes* (economic, demographic and political processes that give rise to vulnerability) via *dynamic pressures* (which 'translate' the effects of root causes into the vulnerability of unsafe conditions) to *unsafe conditions*: "the specific forms in which the vulnerability of a population is expressed in time and space in conjunction with a hazard" [p. 25]. From the point of view of the formalization, vulnerability is considered in its aggregated form while the primitives of vulnerability do not occur individually. Thus, in terms of Diagram 5.3, vulnerability is of type δ. As was mentioned in Section 5.4.3, the corresponding vulnerability interpretation originated here in the field of natural hazard risk. Theoretical definitions of vulnerability provided by Blaikie et al. [1994] were included in the analysis of theoretical definitions above. The table showed that they have the same structure as most other theoretical definitions.

The other component of risk, the hazard, can be graphically represented by assessment approach γ, specialised to the probabilistic case. Blaikie et al. describe the hazard as having "varying degrees of intensity and severity" [p. 21] and refer to statistical likelihoods of hazards obtained from long records. Here, the varying degrees of intensity are the descriptions of several possible futures, with attached probabilities derived from statistics. This is an example of a future projection step which concerns the stimulus only, identified as "exposure" in Section 5.2.2. The notion of harm is contained in 'severity' which can be evaluated for each hazard intensity inside the probabilistic uncertainty description. The risk-component hazard as described by Blaikie et al. [1994] thus corresponds to γ in Diagram 5.3.

Risk could be assessed by assessing the two components hazard and vulnerability and combining the results, as is also displayed by the formula $RISK = Hazard + Vulnerability$ contained in a diagram illustrating the PAR model [Blaikie et al., 1994, p.23]. The property that risk is nil when one of the components is further specifies the combination step: combination operators that do not have this property are not appropriate since they would violate this essential property of risk. This is an example of the benefits of expressing the structure of a concept mathematically: even if expressed in words rather than formulae, "the result is nil if one of the factors is nil" is a mathematical statement which makes assumptions explicit in a clear fashion. The formula chosen to represent this statement is slightly unfortunate because the symbol "+" is at first sight associated with addition, a combination that does *not* have this property.

To do the authors justice, it should be mentioned that Blaikie et al. [1994] do not want to actually assess risk in this way. They consider the PAR an oversimplified and static model, because it "suggests [...] that the hazard event is isolated and distinct from the conditions that create vulnerability." To overcome this separation of the hazard from social processes, the 'access model' is proposed. We shall, however, not consider this model here, but concentrate on the PAR model as a rough description that clarifies the concept risk.

The Risk Triangle

This approach is grounded in the insurance industry and extended to disaster management by Crichton [1999]. The triangle is used to illustrate that and how risk depends on three components: hazard, vulnerability and exposure. This mathematical metaphor constitutes a first step towards a mathematical model which helps stating assumptions clearly. Thinking of the size of risk as the area of the triangle specifies two properties of risk: it increases in all three of its components, that is, if one component increases while the other two are kept constant, risk increases, and it is nil if any of the components is nil. The triangle metaphor, however, cannot serve to actually compute risk from values of the three components. For example, if the sum of two sides does not exceed the length of the third side, no triangle area can be associated to the three lengths. A practical question would be how to reduce the results from the components' assessments to a length, that is, to a single number.

Again, risk can be assessed by assessing the components and then combining the results, and the properties assumed for the concept translate into conditions for the combination. While two of the components of risk used here have the same names as in the PAR model, this approach, unlike the PAR model, does not contain a δ part. It can be graphically represented in terms of the γ part of Diagram 5.3 only.

As before, the component hazard contains a probability distribution over possible futures: "probability and severity of natural hazards is now being modelled in very sophisticated ways" [Crichton, 1999, p. 103]. This concise description again includes a notion of harm in 'severity'. How the steps of providing possible evolutions and evaluating harm for each of them can be undertaken shall be seen in the case study example below, since Crichton does not go into detail on these steps. The other two components can be represented as further steps in the harm evaluation of each possible evolution, as will also be shown for the case study. The brief description by Crichton [1999] is very general: information about vulnerability and exposure comes from large databases.

The term exposure is used as is customary in the insurance context: the maximum exposure to losses is the value of the insured property. In ordinary language this would be referred to as 'exposed property'. Crichton [1999] admits that disaster management practitioners may find the way in which insurers consider exposure unfamiliar, especially because an insurer may reduce exposure by reducing the number of insured properties. While for a country this is not an option, Crichton proposes examples such as discouraging the development of housing and industry in hazard-prone areas.

The third component of risk, vulnerability, deserves a closer look: as an exceptional case in the literature, its definition by Crichton does not mention any uncertainty about the future. Vulnerability defined as "the extent to which [the property] *will* suffer damage or loss" [p. 102, our emphasis] seems to call for the extension "if the stimulus is present". It can be interpreted as a dose-response relationship between stimulus and harm and has been observed to correspond to *sensitivity* as used in the context of vulnerability to climate change, for example by Füssel and Klein [2006] (see page 104). In fact, so defined, vulnerability is a measuring function, used to transform stimulus values into harm values, which is precisely how *sensitivity* was described in the mathematical model for the case where the "joker function" represents the stimulus. Vulnerability thus plays a role like β from Diagram 5.3. Vulnerability in the insurance context is property related, an example would be whether building standards are satisfied. Crichton's example for the disaster context are more general "disaster preparedness measures" such as "contingency plans to help with rapid recovery using local institutions, government resources such as the army, and stockpiles of emergency food and shelters" [p. 102]. These hint at a component of type δ. However, in assessments, no δ measurement is present, as shall be seen from the case study that we now turn to.

Case study: Flood Risk Maps

The anaysis presented here is based on Genovese et al. [2007]. Digital maps of risks from inland river floods at the European scale that identify monetary/economic losses are produced by combining hazard, exposure and vulnerability where

- Hazard is the threatening natural event, including its probability/magnitude of occurence;
- Exposure is the values/humans that are present at the location related to a given event;
- Vulnerability is the lack of resistance to damaging/destructive forces (damage function) [Genovese et al., 2007, p. iii].

In the following, the three risk components used in the Flood Risk Maps will be spelt with a capital initial, to reduce the probability of confusion.

The **Hazard** is described by characteristics which make a flood event measurable, such as flood depth, discharge and duration. The description is spatially explicit: Flood hazard maps (FHM in the diagram below), in which the territory is divided into different flood hazard classes, are generated from catchment characteristics. A map can be seen as a spatially explicit description of the *stimulus* function.

These maps come with a probabilistic uncertainty description given in terms of return periods, derived from runoff statistics. The concept of return periods is a common but frequently misunderstood probabilistic description of floods [see, e.g., Pielke, 1999]. The *n-year flood* refers to a water level that has a probability of $1/n$ of being exceeded in any given year, where the probabilities are assumed to be constant and flood occurences independent for different years. The return period information relates to a distribution function in the sense of probability theory as follows. Say the water levels associated to a few values like $n = 100, 50, 20$ have been determined. If one uses these as the domain of a function which associates to each water level the probability of *non-exceedance*, that is, $1 - 1/n$ for the water level of the "n-year flood", this function can be seen as a discretized version of a distribution function, that is, values are given only for some water levels, not for all. Via this transformation, and the fact that distribution functions and probability measures are in one-to-one correspondence,[28] we can consider the information given in a return period description of floods as an uncertainty distribution. To make a long story short, the flood hazard maps with the return periods are an uncertainty distribution over (spatially explicit) stimulus values, corresponding to exposure in our mathematical model, here called Hazard. One could also consider the flood hazard maps themselves as the possible future evolutions, in this case given in terms of the stimulus only.

In further steps applied to each evolution, that is, to each flood hazard map, damages are computed using a combination of the other two components.

Spatially explicit **Exposure** information is obtained using land use classes from the Corine Land Cover project and a gridded population density of Europe dataset. Later, monetary damage values are associated to Exposure classes.

Vulnerability is given by depth-damage curves, specific for each country and each land use class. A curve expresses the degree of damage, a value in the interval $[0, 1]$, as a function of water depth. The curves are provided from results of another study.[29] Together, these curves can be viewed as a spatially explicit measuring function with the ordered target $[0, 1]$: to a tuple $(country, land\ use, water\ depth)$ this measuring function associates the value that the depth-damage curve belonging to the respective country and land use class outputs for the given water depth. Here, the country and land use information make the spatial extension: within the maps, this information is used to choose the right depth-damage function. The input of interest is the water-depth, that is, the stimulus value. If we consider the flood hazard maps the possible evolutions, the Vulnerability measuring function can thus be interpreted as a function of the evolution.

The combination of the two components Exposure and Vulnerability is given by a flood damage database. The authors mention explicitly that the Risk Triangle structure is maintained by attaching monetary land values to the land cover layer representing exposure, and building an information layer containing the "territorial-specific damage fraction (percent of damage as a function of flood depth)" [Genovese et al., 2007, p. 39] which represents Vulnerability. The database of resulting potential monetary loss occuring for a given water depth, specific per land use type and country, thus is a spatially explicit monetary dose-response function. In the mathematical model of vulnerability to climate change, this corresponds to the sensitivity measuring function, β. While in Section 5.2.2 it was stated that future

[28] Only in the "standard" approach to probability, but this is generally used.

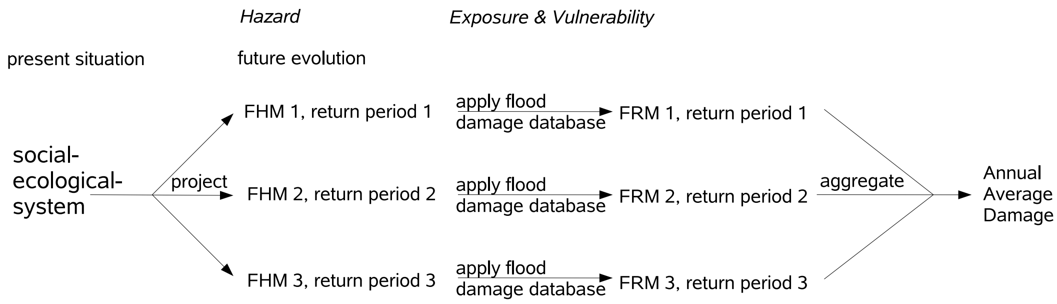
[29] The study by the Joint Research Centre is cited as "not for disclosure". Further analysis as to how far the social system was considered in producing these functions etc. thus is not feasible here.

sensitivity is of interest, here, this measuring function is assembled using past and present data, which can be seen as an approximation.

The harm evaluation is then carried out in a procedure too involved to be described here in a few words. However, in short, it can be described as functorial: the return periods are kept as they are, while "inside" the uncertainty distribution the depth-damage function obtained by combining Exposure and Vulnerability is applied to the water levels in each flood hazard map. The resulting damage maps, referred to as Flood Risk Maps (FRM in the diagram below), provide the levels of harm (still spatially explicit) for each return period.

Finally, a damage-probability curve is constructed from the damage maps for all return periods together. This is a curve which is approximated by the damage values for the given return periods and by linear interpolation between these. The "missing values" of the distribution function (modulo transformation) are filled in, so to say.

Annual Average Damage is computed by an aggregation, corresponding to probabilistic expectation. This can be interpreted as the aggregation step producing γ in Diagram 5.3. In conclusion, we have seen a γ-assessment of risk, graphically:



Two remarks will conclude this section. The case study illustrates the correspondence between Vulnerability in the Risk Triangle and sensitivity as defined by the IPCC. It even occurs explicitly in the texts, when Genovese et al. [2007] state that the Vulnerability component "of the risk assessment procedure is usually dedicated to describing how the exposed people, goods, infrastructures and services are *sensitive* to damage from a given extreme phenomenon." [p.28, our emphasis]. Both terms refer to a dose-response relationship between stimulus and harm.

Second and more importantly, the step in which the components Vulnerability and Exposure are combined reveals that the theoretical decomposition of risk into components is somewhat arbitrary. The potential monetary loss occuring for a given water depth, a natural element of the given computation, does not fit cleanly into one of the specified components of risk, but partly belongs to both Exposure and Vulnerability. Efforts have to be made to keep these components separate in a first step. If in a similar assessment monetary depth-damage curves were associated directly to the component Vulnerability, the component Exposure would become superfluous in the computation of risk. This confirms an observation for vulnerability from Section 5.4.5 equally for the concept risk: phenomena considered in assessments do not map unambiguously to components of a concept at the theoretical level.

5.5.3 Common structure – different terms, and vice versa

Given the social-ecological system as a common object of study in assessments of *vulnerability to climate change* and *risk of natural hazards*, and the fact that natural hazards and disaster management are one of the antecedent traditions of vulnerability [Adger, 2006], similarities were to be expected between the assessment approaches used by the different communities. In fact, it was shown that vulnerability and risk assessments are similar in structure, as revealed by the fact that risk assessments can be described as instances of the general mathematical model of vulnerability. Again, some assessments do not contain all parts mapped out in Diagram 5.3 and consequently different types of results are given the same name, here 'risk'.

The use of *different terms for the same assessment structures* may hide the similarity between assessment approaches and thus hinder the comparability of results. Also, benefitting from the work of the other community is rendered more difficult. Different terms and their various definitions can be a barrier to the communication between the climate change and the natural hazards communities.

The converse problem, using *the same terms for different assessment structures*, previously observed for vulnerability, was also confirmed for the concept *risk*. In the natural hazards community, the term 'risk' may refer to a result of type ε (PAR model), or to γ (Risk Triangle). The term 'vulnerability' is used for yet another element in assessments: in the risk triangle, vulnerability refers to a dose-response relationship, β. It also occurs for δ in the natural hazards community, as it did in the climate change community, where it was additionally used for ε.

Mapping terms used for elements in assessments of "possible future harm" to the placeholders in Diagram (5.3) and vice versa, one never obtains an injective map. This explains some of the confusion asserted in the scientific terminology, both within each of the two communities and between them. Moreover, risk occured in definitions of vulnerability (see Table, page 83), and vice versa, vulnerability in definitions of risk, as given by the PAR model, which certainly does not clarify matters.

Further, both concepts are decomposed into components in scientific approaches, and in both cases this decomposition was shown to involve a certain arbitrariness. Similarities are to be seen: for example, the decomposition of risk into hazard and vulnerability (the PAR model by Blaikie et al. [1994]) resembles the decomposition of vulnerability into an external and an internal side by Chambers [1989]. Even some of the components show similarities as obvious from the fact that components of risk can be mapped to the formalized components of vulnerability.

While in the context of risk, the approach producing a result of type γ is present, as seen in the Flood Risk Maps case study, here a δ approach is not found under this term. Approaches that focus on the social system and the present state here use the term vulnerability. The seminal work by Blaikie et al. [1994], coming from the disaster risk community, is one of the "antecedents" influential also on the context of vulnerability to climate change.

Of course, there is conceptual work on both concepts together. Having seen the different approaches to both concepts, the sensibility of statements made about the concepts may depend on the assessment approach considered. For example, when considering a γ approach to risk, but an ε approach to vulnerability, which includes a focus on the social system, a statement like "vulnerability as a research-organizing concept is thus more complex than risk ..." [Hochrainer et al., 2007] is warranted. However, when considering vulnerability as the dose-response relationship in the risk triangle, the above statement would not make sense. Therefore, care should be used when making such statements, and these should not be quoted out of context, because the explanation which approach to risk and vulnerability is taken is essential.

In conceptual work one finds seemingly contradictory statements, in which the concepts vulnerability and risk occur in different ways as components of each other. The mathematical model can in some cases help resolve the contradiction, as exemplified by the following two statements:

> Vulnerability [*here* ε] can [...] be explained by a combination of social factors [*here* δ] and environmental risk [*here* γ], where risk are those physical aspects of climate related hazards exogenous to the social system. [Adger, 1999, p. 252].

and

> In evaluating disaster risk [*here* ε], the social production of vulnerability [*here* δ] needs to be considered with at least the same degree of importance that is devoted to understanding and adressing natural hazards [*here* γ]. [Blaikie et al., 1994, p. 21]

These results suggest that, ideally, the conceptual barriers between the two communities could be overcome by setting aside the often emphasized conceptual differences. An agreement on the very general meaning of the terms vulnerability and risk, as provided by the everyday language concepts, can be taken for granted because otherwise there would have been no reason to use precisely these terms for the scientific concepts. This general meaning would serve as a basis for communication. Further, assessment approaches can be specified on a case to case basis. Discussing assessments instead of definitions given in natural language, one gains precision and explicitness. The graphical framework presented could support such discussions, and especially the comparison of assessments, also across the borders of the two communities. Whether such a 'return to basics' and giving up theoretical conceptual issues would be feasible, is a different story.

5.5.4 Remarks on formalization

Having presented a formal framework of vulnerability, shown its applicability to the concept risk, and applied it to clarify concepts, assessment approaches and associated interpretations, some remarks on formalization relating back to the formalizations of probability are due.

Comparing the formal framework of vulnerability to the mathematical models of probability, one can consider it as ranging between Kolmogorov's and de Finetti's formalizations in the following sense: Kolmogorov's formalization considers probability a primitive and thus does not explain its meaning. Here, vulnerability is decomposed into primitives, answering to the purpose of concept clarification. De Finetti, whose formalization of probability also aims at concept clarification goes a step further. He proposes an operational definition, while the mathematical models of vulnerability do not establish how to measure vulnerability. On the other hand, de Finetti formalizes one interpretation of probability. Here, the formal framework can be considered as rather similar to Kolmogorov's model of probability which is independent of interpretations of the concept. In fact, the formal framework has been used to discuss different interpretations.

While conceptual confusion is a common starting point in the formalizations of probability and vulnerability, the situations are rather different. The confusion Kolmogorov addressed with his axioms was mainly that within the mathematical discipline. For vulnerability, a mathematical theory does not exist. The confusion to be addressed concerns the meaning and the measurement of vulnerability. De Finetti's aim of formalizing probability was clarifying meaning and measurement of a concept, but the concepts themselves show an important difference. Vulnerability describes a property of people which is judged undesirable, probability describes information about uncertainty in a neutral fashion. An important question in studying vulnerability is "how does vulnerability come about?" because in the end the aim of understanding vulnerability is to reduce it. This question does not seem to make any sense in the context of probability.

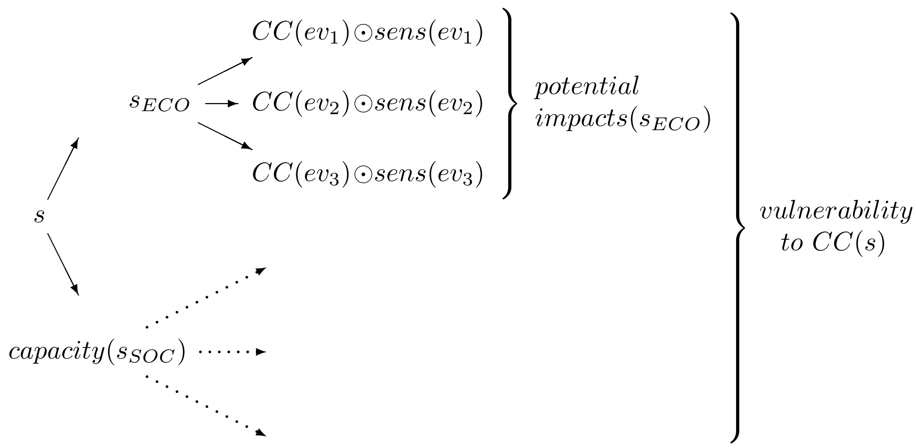
5.6 Different representations of the formalization

The formal framework of vulnerability to climate change has been presented in different forms, three of which were seen in this work. Mathematical formulae, embedded in text, were used in Sections 5.1 and 5.3. In the formulae, details that may seem insignificant to a mathematician, can make a difference for non-mathematicians reading them. A hint about notation that was perceived as unintuitive came from a non-mathematician in the FAVAIA group. The presentation by Ionescu [2009] denotes the set of harm values by V and the set of vulnerability values by W. Later presentations changed the notation to H for harm values and V for vulnerability values, to avoid confusion from this notation. Even if seemingly a very small difference, this change in notation comes a little bit closer to what Sarukkai [2001] states about "holding on to the semantics". H can be read as an abbreviation of the primitive harm, so that this notation recalls the primitive that has been translated into the measuring function taking values in H.

It was stated already that one cannot assume mathematics to function as lingua franca in the vulnerability research community. This at first sight detracts from (or in the worst case: eliminates) the benefits of formalization deriving from communication in mathematical language. On the other hand, formulae are not what constitutes mathematics as argued in Section 2.4. The formal framework represented by diagrams is still mathematical: it is precise and general and encourages to make assumptions explicit.

An interesting experience was made with a representation of the formal framework in terms of diagrams of the form of Diagram 5.3. These combined a graphical representation of some of the mathematical concepts with formulae for others. The basic idea of sketching a non-deterministic future evolution and the functorial application of measurements to the uncertainty distribution has been explained in Section 5.3.4. Curly braces were used to express an aggregation operation without restricting to specific functions. The symbols $\oplus, \otimes$ and $\odot$ were chosen purposefully to avoid symbols which are familiar from everyday use or primary school mathematics, in order to emphasize generality and avoid implicit assumptions. In fact, we wanted to avoid situations like the one sketched in Section 5.5.2, where in a diagram the symbol $+$ denotes a combination that, in the text two pages earlier, is specified to be nil if one of the components is nil.

An article presenting the formal framework and comparing the concepts risk and vulnerability using these hybrid diagrams has been rejected by the Journal Regional Environmental Change because "far too theoretical for REC". In the present work only the most general version of the diagrams in the article has been used, Diagram 5.3 with greek letters as placeholders to be filled in with the different terms used in the different assessment approaches. One example using words instead is the following diagram displaying the integrated assessment approach, where CC abbreviates climate change, *sens* sensitivity.



A comment about the article by an editor recalls the description of mathematics as a cultural anathema by Enzensberger [1999]:

> Either the mathematical formulations simply restate in mathematical terms rather trivial relationships, or I am missing something (probably the latter). So it may be a good paper in terms of theoretical mathematics, and if so, it should be submitted to an appropriate journal in that field.

Similarly, after a presentation given using diagrams only, the remark "thanks for not having shown formulas" was made. In both cases, the assumption that one did not or would not understand mathematics comes easily. Too easily, in fact, in the former example. It is rightly observed that some of the relationships stated are rather trivial. The fact that vulnerability describes a possibility of future harm is not a deep insight. However, precisely these "trivial relationships" were supposed to be made explicit as a clear basis for communication instead of jumping into vague descriptions of distinctions that are 'often difficult' to make.

The journal editor's quote further shows that it is still unclear "what mathematics is". While there are mathematical results related to the formal framework, for example on monadic dynamical systems in the work of Ionescu [2009], or the monads of finitely additive probabilities presented in the following chapter, the framework itself is not "theoretical mathematics". The plain mathematical description of structures found in vulnerability assessments does not contain mathematical results in the sense of theorems and proofs, which could be submitted to a mathematical journal.

The comment to the submitted version of the paper further contained the sentences: "And this raises the question: who is the audience for this paper? It certainly doesn't seem to be practitioners or researchers in the field." Because the formal framework is supposed to be helpful for researchers in the field, a further version of the paper was produced. This uses the diagrams as seen in this work in Section 5.5. These diagrams are closer to the "box-and-arrow diagrams" common for presentation of conceptual frameworks and frameworks for vulnerability assessment [e.g. O'Brien et al., 2007; Turner II et al., 2003]. This adapted version has been informally commented by a colleague who had himself been struggling with the too many definitions. Having found that the task of comparing (theoretical) definitions of vulnerability from the climate change and the natural hazards communities seems futile, this colleague stated that the framework was helpful to clarify ideas. This gives reason to hope that some of the clarification experienced by those carrying out the formalization process can be transferred.

That clarity can benefit from a mathematical description of concepts and assessment approaches was observed also in works about the concept risk. Both Blaikie et al. [1994] and Crichton [1999] use mathematical concepts to make their assumptions about the concept risk explicit. Even where used only as metaphors, mathematical concepts are helpful to illustrate properties of risk that would have been more difficult to state as concisely and clearly without recurring to mathematics.

"Selling" mathematics however remains a balancing act. The formal framework has from the beginning[30] been accompanied by justifications and words of caution, as for example the following.

> First, the framework could be perceived as being overly prescriptive, limiting the freedom and creativity of researchers to generate and pursue their own ideas on vulnerability. Second, it could be seen as being developed for illicit rhetorical purposes, namely to throw sand in the eyes of those unfamiliar with mathematical notation. Inspite of these caveats, we hope that the development of a formal framework of vulnerability will turn out to be a worthy undertaking, offering an opportunity for rigorous interdisciplinary research that can have important academic and social benefits. If this opportunity is seized by many, the risks represented by the two caveats will be minimised. [Ionescu et al., 2009, p. 2].

A hybrid presentation in another sense is in the making: a two tier paper which presents simple mathematical formulae side by side with the last generation of diagrams. The paper promises that the mathematical concepts used do not go beyond old high school acquaintances such as sets, functions or probability distributions, except for the carefully explained concept of a functor (for the simple examples of discrete sets and probability distributions only). Given that the mathematical concepts used are very basic, the following trick is rather of a psychological nature, trying to overcome a fear of not understanding mathematics: the paper purposefully indents the mathematical definitions and offers that they can be skipped as the information is also given in the surrounding text and the diagrams. Reactions remain

[30]The publication process of the quoted paper has been rather long, the quote can also be found in a working paper from 2005.

to be seen.[31] Before concluding this chapter, let us draw a link back to the previous chapter.

5.7 Probability, possibility and vulnerability

Throughout this chapter, the concepts probability and possibility occured in several places. This section collects some relations of these concepts to vulnerability which allow to draw conclusions on either side.

Probability can be used as a tool in vulnerability assessments, but it was seen that studies of vulnerability to climate change make less use of probabilistic methods than risk assessments in the context of natural hazards do. Also, studies of vulnerability to poverty use probability more frequently, as can be seen for example from the synthesis work of Calvo and Dercon [2005] which proposes a probabilistic setting. This difference can be traced back to the non-assignment of probabilities to the SRES scenarios discussed in Section 4.8. The assessments of vulnerability to climate change that use these in their description of the uncertain future not only find no default probabilities to use with the scenarios. Also, the debate, and especially the position of the authors of the scenarios, discourages from using probabilities.

It was further discussed (see Section 4.8) that frequentist probabilities for future scenarios are not appropriate in most cases. For studies of natural hazards and poverty, where the time horizons involved are generally (much) shorter, deducing probabilities from past data seems a more warranted approach. In climate change, one is left with subjective probabilities, or the belief aspect of probability. This implies subjectivity of vulnerability when using probability. While subjectivity of vulnerability measurements is criticized (seen in Section 5.4.4), the understanding seems to be that measurements are simply not good enough to measure vulnerability objectively. Subjectivity of the concept itself is not treated as frequently. As mentioned in Section 5.4.4, using admittedly subjective measurements of concepts which only have a subjective meaning in decisions with high stakes is very controversial.

The use or not use of probability in the context of climate change studies is, however, more directly linked to the concept vulnerability than shown so far in this work. Lempert et al. [2004], in their description of different frameworks for decision making, outline the Assess-Risk-of-Policy framework which considers different policy options and for each of these investigates the associated uncertainties. This framework, unlike the Predict-Then-Act framework which they contrast it with, does not necessarily make use of probability. Here, the concept vulnerability arises:

> . . . climate-change policy-makers should understand various sources of vulnerability to climate and focus on the technical and political actions that will reduce these vulnerabilities. These include improving the ability to respond to extreme climate events and increasing the efficiency of society's energy use regardless of any expectations about climate change. [Lempert et al., 2004, p. 4].

The phrases 'ability to respond' and 'regardless of any expectations about climate change', that is, about the future, show that the interpretation of vulnerability used here focuses on the entity and the present time. The association of this vulnerability interpretation and a 'no-probabilities'-approach is even more obvious in the work by Dessai and Hulme [2004]. These authors distinguish *top-down* from *bottom-up* assessments, which can be mapped to our γ- and δ-assessment approaches, respectively. They state a "case for probabilities" for γ-assessments but a case against them for δ-assessments. The δ-approach is described as

[31] The comment about the paper submitted to Regional Environmental Change also contained a categorical rejection of formalization:

> Do we really need such formalised definitions of terms that, in my view, can never (and should never) be completely formalised? A good analogy is the term "sustainability". After 20 years, it still doesn't have a single, formalised definition that is widely accepted. But this certainly doesn't prevent the term from being used. And most people certainly recognise an unsustainable society when they see it! (even if they can't give a precise definition).

In a scientific context, this seems best not to deal with at all.

"tremendously useful for understanding society's vulnerability to present-day climate and also the underlying causes of vulnerability", and its focus on people's capacity is highlighted. The authors state that not only probabilities, but even scenarios of climate change themselves are not needed for climate adaptation policy using this approach.

The different views on the usefulness of probability in climate change research can thus also be derived from the different interpretations of vulnerability. That in considering the concept vulnerability as of type δ, one is not concerned with probabilities can further be explained by returning to the duality of the concept possibility. This concept, inherent in vulnerability, was found present also in the concepts exposure and capacity, corresponding to a decomposition of vulnerability into an external and an internal side (see Section 5.2). To the concept possibility two aspects were attributed, a physical and a knowledge aspect. Exposure has both these aspects from the point of view of the entity: the stimulus has the physical possibility of occuring, but the entity does not know whether it will happen. Capacity on the other hand does not involve the knowledge aspect: a physical ability is considered. Knowledge about whether something will happen or not (even an action of the entity) is not taken into account in capacity assessments.[32] Since both subjective and frequentist probability are tools to describe information about uncertainty regarding the question whether or not something will happen (or has happened, or is true, etc.), they are not perceived as useful when dealing with this (only) physical possibility.

An investigation of the use of probability in vulnerability assessments has therefore revealed also that the dualities found for the concepts probability and possibility do not simply match. Probability is a tool in situations involving possibility with a knowledge aspect. As seen in de Finetti's work, this may be anything the reference person is uncertain about, that is, anything the reference person does not know. A merely physical aspect possibility seems not usefully described by probability. Since vulnerability in the δ-approach, respectively capacity as the δ component of (ε-)vulnerability, considers a physical possibility, its assessments do without probabilities. Different types of uncertainty have been described in the literature, some of which would relate to this discussion. For example, Dessai and Hulme [2004] discuss further distinctions of uncertainty in relation to climate change. Since there is a vast spectrum of literature on uncertainty that would go beyond the scope of this work, this chapter will be concluded here.

5.8 Conclusions

Conceptual clarification was the aim of using formalization in the context of vulnerability to climate change, where confusion is present in terminology and conceptual literature. As a starting point, vulnerability as used in ordinary language was analyzed, identifying as primitives an entity (and a stimulus), an uncertain future evolution of the entity (and the stimulus) under consideration, and a notion of harm. The mathematical definition of vulnerability, in which the defining concepts are the mathematical translations of these primitives, provides an abstract and precise description of what the word vulnerability means. Discussing properties of the mathematical elements of the definition, intuitions about the concept can be made explicit in a precise manner, as seen for example for the monotonicity condition for the aggregation function. Vulnerability was mathematically defined as an aggregate measure of possible future harm to an entity.

This definition was shown to be a common denominator of many slightly differing theoretical definitions of the scientific concept vulnerability, where the same primitives and structure are found. Refining the ordinary language concept, the definitions decompose vulnerability into related concepts such as exposure, susceptibility and capacity. These concepts, often left undefined, leave room for interpretation in the concept vulnerability itself. In the terminology, the scientific concept vulnerability is not introduded by starting out from

[32] Recall the statement that "what may or may not happen in the future" is not of primary interest for capacity by Kelly and Adger [2000] quoted on page 103.

intuitive primitives and defining all concepts unambiguously upon these and previously defined concepts. Theoretical definitions alone turned out insufficient to base a formalization of vulnerability and related scientific concepts upon them.

A more concrete understanding of the scientific concepts can be reached by considering which steps are undertaken in vulnerability assessments. The mathematical model of the scientific concept vulnerability, as derived by Ionescu [2009] from vulnerability assessments, refines the mathematical definition of ordinary language vulnerability. It was here adapted to take into account components of vulnerability found in the theoretical definitions. This second model was used to describe vulnerability and risk assessment approaches, revealing how these relate to each other. While one approach considers possible future evolutions of the social-ecological system explicitly, in another approach, measurements are carried out on the present state of the system. It was observed that the former approach predominantly focusses on the ecological and the latter on the social component of the SES. A third approach to assessing vulnerability comprises the first two and combines the results. In all three cases, the assessment result is referred to as 'vulnerability', producing the equivocality that causes a good part of the confusion in the terminology.

Interpretations of vulnerability that had been identified in the conceptual literature were related to these assessment approaches, revealing three distinctions of two interpretations each to describe the same distinction. In these interpretations, one time aspect of the *present property* vulnerability that consists in the *future possibility* of harm occurring is emphasized.

A suggestion derived from this explanation of the confusion in the vulnerability terminology is to refrain from proposing detailed theoretical definitions. Ironically, the confusion may be considered self-enhancing: aiming for clarity, most vulnerability assessments first provide theoretical definitions of the concepts used, adding new trees to the proverbial forest. An agreement on detailed theoretical definitions of concepts seems not only not feasible because all the terms used in this context are already overloaded with meanings, but also superfluous, given the imprecision inherent in natural language.

Spelling out the structure of concepts involved, one can consider the formal definition as *one* definition which *fits all cases*, an entity whose existence seems impossible to many researchers in the domain of vulnerability. The question posed by O'Brien et al. [2007] "whether it is possible to integrate [the] two interpretations into a comprehensive and formal framework for understanding vulnerability to climate change" [p. 84] however, deserves careful consideration. One framework has here beed used to clarify the two interpretations considered, but the phrase "understanding vulnerability to climate change" can and probably should be read as "understanding how vulnerability to climate change comes about". This question is not answered by the formal framework. It remains to the researchers working on vulnerability to answer this question. If the framework can help clarify ideas about concepts and previous assessments for these researchers, it has made a contribution.

As an outlook on further applications of this framework, let us mention the development of computational tools relating to it in the FAVAIA group. The basic architecture that can be used for computational vulnerability assessment by plugging in different models as well as evaluation and aggregation functions has the advantage that these elements are kept separate. The possibility of exchanging a part without having to reprogram everything allows a high degree of flexibility in assessments. A benefit of formalization stated by Ionescu et al. [2009], to "allow modellers to take advantage of relevant methods in applied mathematics, such as system theory and game theory" [p. 2], might be obtained.

The formalization of vulnerability has considered concepts only. Hinkel [2008] notes that a "motivation for formalising statements is to analyse the consequences of these statements" [p. 25]. The benefit of applying mathematical methods and results to vulnerability and thus contribute to a theory of vulnerability would presuppose first of all the translation of statements made about vulnerability. While probability theory before Kolmogorov's axioms was already a mathematical discipline, the theory of vulnerability as far as it exists to date is non-mathematical and mathematical results seem out of reach for now. A formalization of concepts is, however, a very first step towards aspiring mathematical results on vulnerability.

Chapter 6

Finitely additive probability monads

This chapter treats probability functors and monads. There exists a monad of probability measures, as shown by Giry [1981], and sketched in Section 6.3.

In climate change research, frequentist probabilities often are not avaliable (as discussed in Section 4.8), and in vulnerability assessments one might want to use probability assignments by experts. For the case where the expert's assignment is not specified on a σ-algebra, and is thus not a probability measure, it can be represented mathematically by a coherent probability assignment.

The formal framework of vulnerability introduced in Chapter 5 uses a functor to represent uncertainty. Therefore, a naturally arising question is whether the alternative mathematical models for probability seen in Chapter 4 also yield functors for application in the framework. They do, as will be shown throughout for finitely additive probabilities and in Section 6.7 for coherent probability assignments. The computational version of the formal framework of vulnerability further employs monadic dynamical systems (see Section 3.3). In order to use a monadic dynamical system, of course, a monad is needed. The next question therefore is: are there monads of finitely additive probabilities and/or of coherent probability assignments? The question is answered positively for finitely additive probabilities in this chapter. Given that finitely additive probabilities were shown to generalize the σ-additive Kolmogorov probabilities, the question arises whether such a monad can be defined as an extension to the Giry-monad. It will also be answered with 'yes' in this chapter.

The first monad of finitely additive probabilities (Section 6.2) considers finitely additive probabilities which are defined on the whole powerset of the underlying set. It is the technically least involved monad, and can be defined on the category $\mathbb{F}un$, because no measurability questions arise. However, it is not very closely related with the Giry monad. Therefore, Section 6.5 presents another monad, which is. This is defined on a category which generalizes the category $\mathbb{M}es$ of measurable sets and functions used by Giry. From the construction of this monad, the question arises whether this monad can be slimmed down for more generality. This question is addressed in Section 6.5.6. The chapter starts out with a section providing the necessary mathematical armamentarium.

6.1 Integration with respect to finitely additive probabilities

This section presents an integral for bounded real valued functions with respect to finitely additive probabilities. It is based on the chapter about integration with respect to finitely additive measures by Bhaskara Rao and Bhaskara Rao [1983], especially on their "S-integral" that is defined in terms of upper and lower sums, similar to the Riemann integral.

Throughout the section consider given a non-empty set Ω, an algebra $\mathcal{A}$ on Ω and a finitely additive probability P on $(\Omega, \mathcal{A})$. We denote finite partitions $\{A_1, \ldots, A_n\}$ of Ω in $\mathcal{A}$ by Z in this section. Let $\mathcal{Z}$ be the collection of all these Z and define a partial order on partitions in terms of refinement. $Z_1 \leq Z_2$ if Z_2 is a refinement of Z_1, that is, each set in Z_1 is a union of sets in Z_2.

Given a bounded function $f : \Omega \longrightarrow \mathbb{R}$ and a partition $Z = \{A_1, \ldots, A_n\}$ in $\mathcal{Z}$, lower and upper sums are defined as follows:

$$L(Z) = \sum_{i=1}^{n} (\inf_{\omega \in A_i} f(\omega)) P(A_i)$$
$$U(Z) = \sum_{i=1}^{n} (\sup_{\omega \in A_i} f(\omega)) P(A_i)$$

As f is bounded and P is a probability, these are real numbers.

The usual inequalities hold between the upper and lower sums of partitions $Z_1 \leq Z_2 \in \mathcal{Z}$, because each set in Z_1 is a union of sets in Z_2:

$$(\inf_{\omega \in \Omega} f(\omega)) \leq L(Z_1) \leq L(Z_2) \leq U(Z_2) \leq U(Z_1) \leq (\sup_{\omega \in \Omega} f(\omega))$$

In particular, for any two partitions Z_1, Z_2, it can easily be shown that $L(Z_1) \leq U(Z_2)$ by considering a common refinement, whose lower and upper sum lie between these. Therefore,

$$\underline{\int} f dP = \sup_{Z \in \mathcal{Z}} L(Z)$$

$$\overline{\int} f dP = \inf_{Z \in \mathcal{Z}} U(Z)$$

are well-defined numbers, because upper sums are bounded below (by any lower sum) and vice versa lower sums are bounded above.

From the above inequalities it follows that

$$\underline{\int} f dP \leq \overline{\int} f dP \tag{6.1}$$

Definition 6.1.1. A bounded function $f : \Omega \longrightarrow \mathbb{R}$ is *integrable*, or more precisely *integrable under* P, if the lower and the upper integral coincide, that is, if equality holds in (6.1). In that case, the integral of f with respect to P is defined to be

$$\int f dP = \underline{\int} f dP = \overline{\int} f dP.$$

Having defined the integral, we list some (mostly well-known) properties of the integral that are needed later.

For simple functions, the integral is the usual sum known from the Lebesgue integral definition.

Lemma 6.1.2. *For a simple function $f : \Omega \longrightarrow \mathbb{R}$ with the representation $f = \sum_{i=1}^{n} c_i \mathbb{1}_{A_i}$, the integral $\int f dP$ is equal to $\sum_{i=1}^{n} c_i P(A_i)$.*

Proof. Consider the partition $\{A_1, \ldots, A_n\}$, then on each A_i the function f is constant, with value c_i, wherefore $\inf_{\omega \in A_i} f(\omega) = \sup_{\omega \in A_i} f(\omega)$ and hence the lower and upper sum coincide. Both are equal to $\sum_{i=1}^{n} c_i P(A_i)$, which is also the value of upper and lower integral, since further refinement cannot change the values. □

In particular, the integral of a constant function with respect to a finitely additive probability is the constant itself, because a constant function is a simple function with only one element in the partition, the set Ω itself, and $P(\Omega) = 1$.

Example 6.1.3. *A special case of this integral is the Darboux integral which is equivalent to the very well-known Riemann integral. This means, a function is Darboux integrable if and only if it is Riemann integrable, and the integrals, if they exists, coincide. Let $\Omega = [0, 1]$ in $\mathbb{R}$, with $\mathcal{A} = \mathcal{J}[0, 1]$, the algebra generated by the intervals, and let the finitely additive probability P be the Lebesgue measure restricted to $\mathcal{A}$, weighting an interval with its length.*[1] *The above construction for this case produces the Darboux integral of a bounded function. For the Riemann integral, the construction involves the function value at a given point in each interval instead of supremum and infimum. Due to the equivalence, however, some authors define the Riemann integral using the construction seen here right away.*

Riemann-integrable functions are well studied, some characterisations are:

- *The set of discontinuity points of f is of Lebesgue measure zero.*
- *Šikić [1991] shows that a function $f : [a, b] \longrightarrow \mathbb{R}$ is Riemann integrable if an only if there are $\mathcal{F}$-simple functions which converge uniformly to f, where $\mathcal{F}$ is the algebra*

$$\mathcal{F} = \{A \subseteq [a, b] | A \text{ is Lebesgue measurable and } \Lambda(Fr(A)) = 0\}$$

where Λ denotes the Lebesgue measure and $Fr(A)$ is the frontier of the set A, that is, the intersection of the closures of A and its complement, $Fr(A) = Cl(A) \cap Cl(A^c)$.

In the general case, one cannot give a characterisation in terms of, for example, continuity because one does not necessarily have a topology on Ω. The set Ω is merely equipped with an algebra. While in measure theory the integral with respect to a σ-additive probability is by construction well-defined for all measurable functions, and non-measurable functions are not taken into consideration, here, the question which bounded functions are integrable is not unimportant. The following section provides some answers.

Integrability conditions

The definition of integrability with respect to a finitely additive probability is not very insightful. Two main questions are: given an $(\Omega, \mathcal{A})$ and a fixed P, which bounded functions $f : \Omega \longrightarrow \mathbb{R}$ are integrable under P? Which functions are integrable under any P?

Lemma 6.1.4. *Given $(\Omega, \mathcal{A}, P)$ and a bounded function $f : \Omega \longrightarrow \mathbb{R}$, the following are equivalent:*

(*i*) *f is integrable under P (in the sense of Definition 6.1.1).*

(*ii*) *For every $\varepsilon > 0$, there exists a finite partition $\{A_0, A_1, \ldots, A_n\}$ of Ω in $\mathcal{A}$, such that $P(A_0) < \varepsilon$ and for $i = 1, \ldots, n$ one has $|f(\omega) - f(\omega')| < \varepsilon$ for every $\omega, \omega' \in A_i$. For further reference, a partition of this kind will be called an ε-*partition* for f.*

(*iii*) *There exists a sequence of simple functions $f_n : \Omega \longrightarrow \mathbb{R}$ which converges to f in the following sense: for every $\varepsilon > 0$*

$$\lim_{n\to\infty} P^*(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) = 0.$$

where the set function $P^* : \mathfrak{P}(\Omega) \longrightarrow [0, 1]$ is defined by

$$P^*(B) = \inf\{P(A) | B \subseteq A, A \in \mathcal{A}\}$$

and called the outer probability *induced by P. This notion of convergence can be found as hazy convergence in the literature. It will be referred to as* hazy convergence under *P when the finitely additive probability is not fixed in the context.*

[1] We use the unit interval instead of an arbitrary interval $[a, b]$ in $\mathbb{R}$ because we have introduced the integral only for finitely additive probabilities, and we thus avoid having to renorm the Lebesgue measure.

Proof. We show the equivalence of condition (*ii*) with both other conditions.[2]
(*i*)⇒(*ii*) If f is integrable with respect to P, for any $\varepsilon > 0$ one finds a partition Z such that

$$\int f dP - L(Z) < \varepsilon^2/2 \text{ and } U(Z) - \int f dP < \varepsilon^2/2 \qquad \Rightarrow U(Z) - L(Z) < 2\varepsilon^2/2 = \varepsilon^2, \text{ but}$$

$$\begin{aligned} U(Z) - L(Z) &= \sum_{i=1}^{n} (\sup_{\omega \in A_i} f(\omega)) P(A_i) - \sum_{i=1}^{n} (\inf_{\omega \in A_i} f(\omega)) P(A_i) \\ &= \sum_{i=1}^{n} (\sup_{\omega \in A_i} f(\omega) - \inf_{\omega \in A_i} f(\omega)) P(A_i). \end{aligned}$$

This sum is bounded by ε^2. Let us divide its summands into two groups. By renumbering if necessary, let $A_1, \dots A_k$ be the sets in the partition for which $\sup_{\omega \in A_i} f(\omega) - \inf_{\omega \in A_i} f(\omega) < \varepsilon$ holds. Then $|f(\omega) - f(\omega')| < \varepsilon$ for every $\omega, \omega' \in A_i$ for $i = 1, \dots, k$.
Take all those A_j, $j = k+1, \dots, n$ for which this was not the case, then

$$\varepsilon \cdot \sum_{j=k+1}^{n} P(A_j) \leq \sum_{j=k+1}^{n} (\sup_{\omega \in A_j} f(\omega) - \inf_{\omega \in A_j} f(\omega)) P(A_j) < \varepsilon^2$$

where the last inequality holds because all summands in the original sum are non-negative, wherefore the sum of a part of them cannot exceed the whole sum. Thus, $\sum_{j=k+1}^{n} P(A_j) = P(\bigcup_{j=k+1}^{n} A_j) < \varepsilon$. Set $A_0 = \bigcup_{j=k+1}^{n} A_j$, then the partition that we looked for is given by $\{A_0, A_1, \dots, A_k\}$.

(*ii*)⇒(*i*) Let $\varepsilon > 0$. Given condition (*ii*), one can show that

$$\overline{\int} f dP - \underline{\int} f dP < \varepsilon.$$

Since ε is arbitrarily small, this shows the integrability of f with respect to P. Let $M = \sup_{\omega \in \Omega} |f(\omega)|$ and choose a finite partition $Z = \{A_0, A_1, \dots, A_n\}$ with $P(A_0) < \varepsilon/4M$ and $|f(\omega) - f(\omega')| < \varepsilon/2$ for every $\omega, \omega' \in A_i$ (which exists by (*ii*)). Then,

$$\begin{aligned} \overline{\int} f dP - \underline{\int} f dP \leq U(Z) - L(Z) &= \sum_{i=0}^{n} (\sup_{\omega \in A_i} f(\omega) - \inf_{\omega \in A_i} f(\omega)) P(A_i) \\ &\leq 2M \cdot P(A_0) + \varepsilon/2 \sum_{i=1}^{n} P(A_i) < \varepsilon/2 + \varepsilon/2 = \varepsilon. \end{aligned}$$

(*ii*)⇒(*iii*) Again, consider condition (*ii*) satisfied. For each n, pick an ε-partition for $\varepsilon = 1/n$ and denote it by $Z_n = \{A_0^n, \dots, A_{k_n}^n\}$. Then, $P(A_0^n) < 1/n$ and for all $i = 1, \dots, k_n$, one has $|f(\omega) - f(\omega')| < 1/n$ for all $\omega, \omega' \in A_i^n$. Fix an ω_i in each A_i^n and define the simple function f_n by

$$f_n = 0 \cdot \mathbb{1}_{A_0^n} + \sum_{i=1}^{k_n} f(\omega_i) \mathbb{1}_{A_i^n}.$$

These functions converge hazily under P to f. To see this, choose $\varepsilon > 0$ and $m \geq 1$ such that $1/m < \varepsilon$. For $n \geq m$, by construction, f_n and f differ less than $1/n < 1/m$ on all sets

[2]Most of this proof is adapted from [Bhaskara Rao and Bhaskara Rao, 1983, Theorems 4.4.7 and 4.5.7].

of the partition used to construct f_n, except on the set A_0^n. The set of ω for which f_n and f differ more than ε is thus contained in A_0^n.

$$P^*(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \varepsilon\}) \leq P^*(A_0^n) = P(A_0^n) < 1/n < \varepsilon$$

and therefore, condition (*iii*) holds.

(*iii*)⇒(*ii*) If condition (*iii*) is satisfied, there is a simple function f_n, say $f_n = \sum_{i=1}^m c_i \mathbb{1}_{A_i}$, such that $P^*(\{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \varepsilon/2\}) < \varepsilon' < \varepsilon$. Consider the subset of Ω on which f and f_n differ more than $\varepsilon/2$, and denote it $B = \{\omega \in \Omega : |f_n(\omega) - f(\omega)| > \varepsilon/2\}$. By the choice of f_n, we know that $P^*(B) < \varepsilon'$. Since $P^*(B) = \inf\{P(A)|B \subseteq A, A \in \mathcal{A}\} < \varepsilon'$, there is a $Z_0 \in \mathcal{A}$ with $B \subseteq Z_0$ and $P(Z_0) < \varepsilon$. For the other sets of the partition, we define $Z_i = A_i \cap B^c$. On each of these sets, we have $|f_n(\omega) - f(\omega)| \leq \varepsilon/2$ as well as $|f_n(\omega') - f(\omega')| \leq \varepsilon/2$. Since $f_n(\omega) = c_i = f_n(\omega')$, it follows that for all $i = 1 \dots m$,

$$\begin{aligned}|f(\omega) - f(\omega')| &\leq |f(\omega) - c_i| + |c_i - f(\omega')| \leq |f(\omega) - f_n(\omega)| + |f_n(\omega') - f(\omega')| \\ &\leq \varepsilon/2 + \varepsilon/2 = \varepsilon \text{ for all } \omega, \omega' \in Z_i.\end{aligned}$$

Therefore, (*ii*) is satisfied. □

This can be used to show a result "symmetric" to Lemma 6.1.2. Like for simple functions (and any finitely additive probability) the integral has a special form for concentrated probabilities and any *integrable* function.

Lemma 6.1.5. *Let P on $(\Omega, \mathcal{A})$ be the probability concentrated in a point $\omega^* \in \Omega$, that is $P(A) = \mathbb{1}_A(\omega^*)$. Then, if f is integrable under P, the integral is equal to the function value at that point: $\int f dP = f(\omega^*)$.*

Proof. If f is integrable under P, we have by Lemma 6.1.4 for each $\varepsilon > 0$ an ε-partition for f. That means, the function changes less than ε on all sets in the partition, except possibly on a set to which P gives less than ε probability. Considering the special form of P here, in any finite partition $\{A_0, A_1, \dots, A_n\}$ there will be one set A_j with $P(A_j) = 1$, namely the A_j which contains ω^*. Thus the constraint translates to: for any ε, there is, in $\mathcal{A}$, a set $A_\varepsilon \ni \omega^*$ such that $|f(\omega_1) - f(\omega_2)| < \varepsilon$ for any $\omega_1, \omega_2 \in A_\varepsilon$. Since $\omega^* \in A_\varepsilon$, one has $|f(\omega) - f(\omega^*)| < \varepsilon$ for any $\omega \in A_\varepsilon$. Then, letting Z_ε be the partition $\{A_\varepsilon, A_\varepsilon^c\}$, we have

$$\begin{aligned}L(Z_\varepsilon) &= (\inf_{\omega \in A_\varepsilon} f(\omega)) \cdot 1 + (\inf_{\omega \in A_\varepsilon^c} f(\omega)) \cdot 0 \geq f(\omega^*) - \varepsilon, \text{ and similarly,} \\ U(Z_\varepsilon) &= (\sup_{\omega \in A_\varepsilon} f(\omega)) \leq f(\omega^*) + \varepsilon\end{aligned}$$

As this holds for any $\varepsilon > 0$,

$$\underline{\int} f dP = \overline{\int} f dP = f(\omega^*) = \int f d(\mathbb{1}_{(.)}(\omega^*)) \text{ follows.}$$ □

Also, Lemma 6.1.4 yields integrability for all bounded functions, when the algebra used is the powerset of the underlying set:

Lemma 6.1.6. *Consider a finitely additive probability defined on $(\Omega, \mathfrak{P}(\Omega))$. Any bounded function f is integrable under P. Since this holds for all P, any bounded function is integrable under any P on $(\Omega, \mathfrak{P}(\Omega))$.*

Proof. One can show condition (*ii*) of Lemma 6.1.4 using the inverse image of the function. Since f is bounded, one finds an M such that $f(\Omega) \subseteq [-M, M]$. For any $\varepsilon > 0$ we need to construct an ε-partition. Choose an N such that $1/N < \varepsilon$ and partition $[-M, M]$ into pieces of length $1/N$. With $2M \cdot N$ parts, this partition is finite; denote it $\{Z_1, \dots, Z_K\}$. Consider the inverse images of the pieces, $f^{-1}(Z_i) \in \mathfrak{P}(\Omega)$ for $i = 1, \dots, K$. These are the desired ε-partition: on each $f^{-1}(Z_i)$ the function differs less than ε for any two points by construction. □

Some useful properties of the integral are the following.

Lemma 6.1.7. *Let $f : \Omega \longrightarrow \mathbb{R}$ and $g : \Omega \longrightarrow \mathbb{R}$ be bounded integrable functions.*

(i) *Given a continuous function $h : \mathbb{R} \longrightarrow \mathbb{R}$, the function $h \circ f$ is integrable.*

(ii) *The function $f + g$ is integrable, and $\int f + g\, dP = \int f\, dP + \int g\, dP$.*

(iii) *Given $\alpha \in \mathbb{R}$, the function $\alpha \cdot f$ is integrable and $\int \alpha f dP = \alpha \int f\, dP$.*

(iv) *The function $|f|$ is integrable, and $\left|\int f dP\right| \leq \int |f| dP$.*

Proof. Let us make the function explicit in the notation of lower and upper sums, writing $L(Z, f)$ for the lower sum $L(Z)$ as defined above.

(i) Since f is bounded, we can assume $f : \Omega \longrightarrow [-m, m]$. On this compact interval, the continuous function h is uniformly continuous. That means, for any $\varepsilon > 0$ there is a $\delta \leq \varepsilon$ such that for all $s, t \in [-m, m]$ with $|s - t| < \delta$, one has $|h(s) - h(t)| < \varepsilon$. Fix ε and a corresponding $\delta > 0$. Consider a δ-partition for f as in Lemma 6.1.4(ii), that is, a partition $\{A_0, A_1, \ldots, A_n\}$ with $P(A_0) < \delta$ and $|f(\omega) - f(\omega')| < \delta$ for all $\omega, \omega' \in A_i$ for $i = 1, \ldots, n$. Then one can conclude from uniform continuity that $|h \circ f(\omega) - h \circ f(\omega')| < \varepsilon$ for all $\omega, \omega' \in A_i$ for $i = 1, \ldots, n$ and $P(A_0) < \delta < \varepsilon$. Thus, we have found an ε-partition for the function $h \circ f$, which means that this function is integrable, because ε was arbitrary.

(ii) For the lower sums of the function $f + g$ one has

$$\begin{aligned} L(Z, f + g) &= \sum_{i=1}^{n} (\inf_{\omega \in A_i} (f + g)(\omega)) P(A_i) \\ &\geq \sum_{i=1}^{n} (\inf_{\omega \in A_i} f(\omega)) P(A_i) + \sum_{i=1}^{n} (\inf_{\omega \in A_i} g(\omega)) P(A_i) = L(Z, f) + L(Z, g). \end{aligned}$$

Similarly, $U(Z, f + g) \leq U(Z, f) + U(Z, g)$ for all partitions Z. Therefore,

$$\underline{\int} f dP + \underline{\int} g dP \leq \sup_{Z \in \mathcal{Z}} L(Z, f + g) \leq \inf_{Z \in \mathcal{Z}} U(Z, f + g) \leq \overline{\int} f dP + \overline{\int} g dP.$$

Since f and g are integrable, the left and right side have the same values. Thus,

$$\underline{\int} f + g\, dP = \overline{\int} f + g\, dP$$

and the claim follows.

(iii) This is trivial, as $\alpha \in \mathbb{R}$ can be factored out in any single formula in the construction, for negative α switching the upper and lower parts.

(iv) The function $|f|$ is integrable due to (i) because the absolute value function is continuous. Further, $\left|\int f dP\right| \leq \int |f| dP$ follows from the fact that

$$\left| \sum_{i=1}^{n} (\inf_{\omega \in A_i} (\omega)) P(A_i) \right| \leq \sum_{i=1}^{n} (\inf_{\omega \in A_i} |f(\omega)|) P(A_i)$$

and similarly with the supremum, by the triangle equality, for all sums involved in the integral construction. □

While for the Lebesgue intergal the pointwise limit of integrable functions is integrable because the pointwise limit of measurable functions is measurable, here this is not true in general. Examples for the Riemann integral, which can be considered a special case of our integral (via equivalence with the Darboux integral which is a special case, see Example 6.1.3) abound in the literature. The following one is taken from Gordon [2000].

Example 6.1.8. *Consider a listing $(r_k)_{k\in\mathbb{N}}$ of the rational numbers in $[0,1]$. Define the functions*

$$f_n(x) = \begin{cases} 1 & \text{if } x \in \{r_1, \ldots, r_n\} \\ 0 & \text{else} \end{cases} \quad \text{and } f(x) = \begin{cases} 1 & \text{if } x \in \mathbb{Q} \cap [0,1] \\ 0 & \text{else} \end{cases}$$

Then, $f_n \to f$ and all f_n are Riemann-integrable with an integral of 0, but f is the standard example of a function which is not Riemann-integrable. Its upper and lower sums are 1 and 0 respectively for any partition considered in the definition of the Riemann-integral.

However, uniform convergence preserves integrability. We denote uniform convergence by the usual double arrows.

Lemma 6.1.9. *If a sequence of integrable functions $f_n : \Omega \longrightarrow \mathbb{R}$ converges uniformly to a function $f : \Omega \longrightarrow \mathbb{R}$, denoted $f_n \rightrightarrows f$, then f is integrable and $\int f dP = \lim_{n\to\infty} \int f_n dP$.*

Proof. Fix $\varepsilon > 0$. Uniform convergence of the f_n to f implies that there is an $N \in \mathbb{N}$ such that for all $n \geq N$ and all $\omega \in \Omega$, $f_n(\omega) - \varepsilon \leq f(\omega) \leq f_n(\omega) + \varepsilon$. For a partition Z of Ω in $\mathcal{A}$ (where m denotes the number of the sets in the partition, because n is used as the function index),

$$L(Z,f) = \sum_{i=1}^{m} (\inf_{\omega\in A_i} f(\omega))P(A_i) \geq \sum_{i=1}^{m} (\inf_{\omega\in A_i} f_n(\omega) - \varepsilon)P(A_i) = L(Z, f_n) - \varepsilon \text{ and}$$
$$U(Z,f) = \sum_{i=1}^{m} (\sup_{\omega\in A_i} f(\omega))P(A_i) \leq \sum_{i=1}^{m} (\inf_{\omega\in A_i} f_n(\omega) + \varepsilon)P(A_i) = U(Z, f_n) + \varepsilon.$$

$$\text{Therefore,} \quad \underline{\int} f_n dP - \varepsilon \leq \sup_{Z\in\mathcal{Z}} L(Z,f) \leq \inf_{Z\in\mathcal{Z}} U(Z,f) \leq \overline{\int} f_n dP + \varepsilon.$$

$$\Rightarrow \underline{\int} f_n dP - \varepsilon \leq \underline{\int} f dP \leq \overline{\int} f dP \leq \overline{\int} f_n dP + \varepsilon$$

Since for all $\varepsilon > 0$, there exists an N such that this holds for all $n \geq N$, one concludes that

$$\underline{\int} f dP = \overline{\int} f dP \text{ and } \lim_{n\to\infty} \int f_n dP = \int f dP. \qquad \square$$

While the above results are standard in finitely additive integration, it is not very usual to consider many finitely additive probabilities on the same $(\Omega, \mathcal{A})$ at the same time. This is however necessary to construct a monad of finitely additive probabilities. We therefore provide the following result inspired at Lemma 6.1.4.

Lemma 6.1.10. *If a function $f : \Omega \longrightarrow [0,1]$ is integrable under all P on $(\Omega, \mathcal{A})$, then*

(*i*) *for every $\varepsilon > 0$, there exists a finite partition $\{A_1, \ldots, A_n\}$ of Ω in $\mathcal{A}$, such that for $i = 1, \ldots, n$ one has $|f(\omega) - f(\omega')| < \varepsilon$ for every $\omega, \omega' \in A_i$.*

(*ii*) *there is a sequence of simple functions $f_n : \Omega \longrightarrow [0,1]$ that converges uniformly to f.*

Proof.

(*i*) As integrability of a function with respect to a finitely additive probability P is equivalent to the existence of an ε-partition for each $\varepsilon > 0$, integrability with respect to *all* finitely additive probabilities implies that for a given $\varepsilon > 0$ one finds an ε-partition for any P. Choose an arbitrary P and consider an ε-partition, then one has

$$\{A_0, A_1, \dots, A_n\} \text{ such that for } i = 1, \dots, n, |f(\omega) - f(\omega')| < \varepsilon \text{ for all } \omega, \omega' \in A_i$$

Consider now any $\tilde{P}$ with the property that $\tilde{P}(A_0) = 1$. An ε-partition for this $\tilde{P}$ partitions A_0 further into sets $B_1, \dots, B_m$ with $|f(\omega) - f(\omega')| < \varepsilon$ for all $\omega, \omega' \in B_i, i = 1, \dots, m$. The partition of Ω given by $\{A_1, \dots, A_n, B_1, \dots, B_m\}$ is of the desired form.

(*ii*) A uniformly converging sequence can be constructed like the hazily converging sequence was constructed for Lemma 6.1.4. For each n consider a partition of the just established form for $\varepsilon = 1/n$, denote it by $\{Z_1, \dots, Z_{k_n}\}$, choose an ω_i in each Z_i and let

$$f_n = \sum_{i=1}^{k_n} f(\omega_i) \mathbb{1}_{Z_i}.$$

Then, for each $\varepsilon = 1/M$, there exists an N (in fact, any $N > M$ is fine) such that for all $n > N$, for all $\omega \in \Omega$:

$$\begin{aligned} |f_n(\omega) - f(\omega)| &= \left| \sum_{i=1}^{k_n} f(\omega_i) \mathbb{1}_{Z_i} - f(\omega) \right| = |f(\omega_{i^*}) - f(\omega)| \text{ for the } i^* \text{ with } \omega \in Z_{i^*} \\ &< 1/n < 1/N < 1/M = \varepsilon. \end{aligned}$$

□

From this Lemma follows

Corollary 6.1.11. *Since any bounded function $f : \Omega \longrightarrow \mathbb{R}$ is integrable under all P on $(\Omega, \mathfrak{P}(\Omega))$, there is a sequence of $\mathfrak{P}(\Omega)$-simple functions that converges uniformly to f.*

While possibly unintuitive this can be explained from the fact that $\mathfrak{P}(\Omega)$-simple functions need not look very simple.

6.2 Monad: finitely additive probabilities on powersets

The simplest monad of finitely additive probabilities is the one where we assume all finitely additive probabilities over a set to be defined on the whole powerset. This approach is found in the literature, the example of Dubins and Savage [1965] who use this assumption to avoid "tedious technical measurability difficulties" [p. 8] was given in Section 4.5.3. We define our monad on the category $\mathbb{F}un$ of sets and functions.

6.2.1 The functor

We denote the finitely additive probability endofunctor on the category $\mathbb{F}un$ by F. It associates to a set all finitely additive probabilities defined on its powerset. A task that will occur frequently is that of showing the equality of two finitely additive probabilities. This means showing the equality of two functions. Throughout the chapter, this will be done by proving equality for an arbitrary (and thus any) element in the domain of definition, that is, a set in the powerset or (later) in the algebra under consideration.

Notation 6.2.1. In the following, the notation, as far as the use of brackets is concerned, will be a hybrid between the standards in Category Theory which does not use brackets for function application and Probability Theory, which does. The probability of a set A is denoted by $P(A)$, and simple function application by something like $f(\omega)$. In order to avoid too many brackets hindering legibility, the application of functors and natural transformations will be denoted without brackets. This notation is introduced at first use.

Objects For an object Ω of the category $\mathbb{F}un$, the object $F\Omega$ is the set

$$F\Omega = \{P | P \text{ is a finitely additive probability on } (\Omega, \mathfrak{P}(\Omega))\}.$$

We write $F\Omega$ without brackets, but also without the space that can be used to denote functor application, to avoid notations like $\mu_{F\ \Omega}$ later, where Ω does not seem to belong to μ and F. Since all functors in this chapter are denoted by single capital letters of the Latin alphabet, and objects they are applied to are denoted by Greek letters or are pairs $(\Omega, \mathcal{A})$ etc., this should not cause confusion.

Arrows Given any function $f : \Omega \longrightarrow \Omega'$, one can map a finitely additive probability on the powerset of Ω to one on the powerset of Ω' by the standard procedure giving the induced measure, see Definition 3.2.4. Measurability is not an issue here, because the inverse image of any set in Ω' is an element of the powerset of Ω.

Thus, for P in $F\Omega$ and any set $A' \subseteq \Omega'$,

$$F f\, P(A') = P(f^{-1}(A')).$$

Here, we use the space to denote the application of F to f and that of Ff to P. Function application thus denoted associates to the left, that is FfP is as mentioned the lifted function Ff applied to P. The term without brackets is one finitely additive probability which is then applied to A. With brackets, $((Ff)(P))(A)$ would clutter more complicated formulae later. In cases where the finitely additive probability that Ff is applied to consists of several symbols, as for example $\mathbb{1}_{(.)}(\omega)$, brackets around it will be used of course: for example, $F f\, (\mathbb{1}_{(.)}(\omega))$.

To show that F is a functor, three conditions have to be checked.

Lemma 6.2.2.

(*i*) *$F f\, P \in F\Omega'$, that is, the definition really yields a finitely additive probability on $(\Omega', \mathfrak{P}(\Omega'))$,*

(*ii*) *F preserves identities, $F\, id_\Omega = id_{F\Omega}$, and*

(*iii*) *F preserves composition: for $f : \Omega \longrightarrow \Omega'$ and $g : \Omega' \longrightarrow \Omega''$, $F\,(g \circ f) = F\, g \circ F f$.*

Proof.

(*i*) One needs to check the axioms A1 and A2 on $(\Omega', \mathfrak{P}(\Omega'))$. Both are satisfied because the inverse image is well-behaved.

$$F f\, P(\Omega') = P(f^{-1}(\Omega')) = P(\Omega) = 1 \Rightarrow \text{ A1.}$$

Take two disjoint sets A' and B' in $\mathfrak{P}(\Omega')$, then

$$\begin{aligned} F f\, P(A' \cup B') &= P(f^{-1}(A' \cup B')) = P(f^{-1}(A') \cup f^{-1}(B')) \\ &= P(f^{-1}(A')) + P(f^{-1}(B')) \\ &= F f\, P(A') + F f\, P(B') \qquad \Rightarrow \text{ A2.} \end{aligned}$$

(*ii*) is easy: the inverse image of the identity function leaves subsets of Ω unchanged. Therefore, for any P in $F\Omega$ and all $A \subseteq \Omega$,

$$F\, id_\Omega\, P(A) = P(id_\Omega^{-1}(A)) = P(A) = (id_{F\Omega}(P))(A)$$

because the identity on the right hand side is the function on $F\Omega$ which maps each P to itself.

(*iii*) Take two functions $f : \Omega \longrightarrow \Omega'$ and $g : \Omega' \longrightarrow \Omega''$. Then, $F\,(g \circ f)$ is a function from $F\Omega$ to $F\Omega''$, thus for a $P \in F\Omega$, the finitely additive probability $F\,(g \circ f)\, P$ is defined on $\mathfrak{P}(\Omega'')$. For an $A'' \in \mathfrak{P}(\Omega'')$,

$$\begin{aligned} F\,(g \circ f)\, P(A'') &= P\big((g \circ f)^{-1}(A'')\big) = P\big(f^{-1}(g^{-1}(A''))\big) \\ &= \underbrace{F f\, P}_{\in F\Omega'}\big(g^{-1}(A'')\big) = F\, g\, \big(F f\, P)(A''\big) \\ &= (F\, g \circ F f)\, P(A''). \end{aligned}$$

□

All in all, F is an endofunctor on $\mathbb{F}un$.

6.2.2 The natural transformations

Together with the endofunctor F, two natural transformations $\eta : Id \longrightarrow F$ and $\mu : F^2 \longrightarrow F$ are required to construct a monad.

Definition 6.2.3. The transformation η associates to a set Ω the function $\eta_\Omega : \Omega \longrightarrow F\Omega$ defined by

$$\eta_\Omega\omega(A) = \mathbb{1}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \in A^c \end{cases} \quad \text{for any } A \subseteq \Omega.$$

Here, both brackets and spaces for the application of the natural transformation are left out. Due to the index that η has whenever it is applied, confusion should not arise between the object to which the transformation's arrow belongs, and the element of that object, which this arrow is then applied to.

Again, the term without brackets is a finitely additive probability, which is applied to the set A in the brackets, then. Each $\eta_\Omega\omega$ really is a finitely additive probability on $\mathfrak{P}(\Omega)$, because $\eta_\Omega\omega(\Omega) = 1$ since all $\omega \in \Omega$ are in Ω, and it is additive: for two disjoint sets, ω can be in at most one of them at the same time. To denote the finitely additive probability $\eta_\Omega\omega$ explicitly as a function on $\mathfrak{P}(\Omega)$, we write $\mathbb{1}_{(.)}(\omega)$, that is, $\mathbb{1}_{(.)}(\omega) = \eta_\Omega\omega$.

Next, one has to show that this transformation is natural, that is, $F f \circ \eta_\Omega = \eta_{\Omega'} \circ f$ for $f : \Omega \longrightarrow \Omega'$. Thus, for any $\omega \in \Omega$, we have to show

$$\begin{aligned} (F f \circ \eta_\Omega)(\omega) &= (\eta_{\Omega'} \circ f)(\omega) \text{ that is,} \\ F f\,(\eta_\Omega\omega) &= \eta_{\Omega'}(f(\omega)). \end{aligned}$$

Both sides in this equation yield a finitely additive probability on $(\Omega', \mathfrak{P}(\Omega'))$, whose equality is shown by application to $A' \in \mathfrak{P}(\Omega')$. Starting on the left hand side, one gets

$$\begin{aligned} F f\,(\eta_\Omega\omega)(A') &= F f\,(\mathbb{1}_{(.)}(\omega))(A') = \big(\mathbb{1}_{(.)}(\omega)\big)(f^{-1}(A')) \\ &= \begin{cases} 1 & \text{if } \omega \in f^{-1}(A') \\ 0 & \text{if } \omega \in f^{-1}((A')^c) \end{cases} = \begin{cases} 1 & \text{if } f(\omega) \in A' \\ 0 & \text{if } f(\omega) \in (A')^c \end{cases} \\ &= \mathbb{1}_{A'}(f(\omega)) \\ &= \eta_{\Omega'}(f(\omega))(A'). \end{aligned} \tag{6.2}$$

The naturality of this transformation is thus proven.

To define the transformation μ, first, we need the concept of *evaluation maps*[3]. For a fixed set $A \subseteq \Omega$, the evaluation map

$$\begin{aligned} &ev_A : F\Omega \longrightarrow [0,1] \text{ is defined by} \\ &ev_A(P) = P(A). \end{aligned}$$

When applying the functor twice, starting with a set Ω, one obtains the set of finitely additive probabilities defined on the powerset of $F\Omega$. These finitely additive probabilities over finitely additive probabilities are going to be denoted by Q to avoid confusion with the P on $(\Omega, \mathfrak{P}(\Omega))$. The idea behind the transformation μ is to "reduce such a Q to a P on $(\Omega, \mathfrak{P}(\Omega))$ by, for each $A \subseteq \Omega$, weigthing the results of all $P(A)$ with the probability that Q gives to the P's".

Definition 6.2.4. For a set Ω, the function $\mu_\Omega : F^2\Omega \longrightarrow F\Omega$ (as for the functor leaving out both brackets and spaces in the notation) is defined for each $A \in \mathfrak{P}(\Omega)$ as

$$\mu_\Omega Q(A) = \int ev_A dQ. \tag{6.3}$$

The integral, defined as in Section 6.1, is well-defined for each $A \subseteq \Omega$ and each Q that is a finitely additive probability on $(F\Omega, \mathfrak{P}(F\Omega))$ due to Lemma 6.1.6.

Also, the function on $\mathfrak{P}(\Omega)$ defined by (6.3) really is a finitely additive probability on $(\Omega, \mathfrak{P}(\Omega))$. For all $P \in F\Omega$, one has $ev_\Omega(P) = P(\Omega) = 1$. This implies that the evaluation map for Ω is the constant function with the value 1. Therefore,

$$\mu_\Omega Q(\Omega) = \int ev_\Omega dQ = \int 1 dQ = 1 \text{ because } Q \text{ is a finitely additive probability.} \Rightarrow \text{ A1}$$

Taking two disjoint sets A and B in $\mathfrak{P}(\Omega)$, observe that for each $P \in F\Omega$ one has $P(A \cup B) = P(A) + P(B)$, wherefore $ev_{A\cup B}(P) = ev_A(P) + ev_B(P)$. Linearity of the integral provides

$$\begin{aligned} \mu_\Omega Q(A \cup B) &= \int ev_{A\cup B} dQ = \int (ev_A + ev_B) dQ = \int ev_A dQ + \int ev_B dQ \\ &= \mu_\Omega Q(A) + \mu_\Omega Q(B) \qquad \Rightarrow \text{ A2.} \end{aligned}$$

Hence, axioms A1 and A2 are satisfied and $\mu_\Omega Q$ is an element of $F\Omega$ for each $Q \in F^2\Omega$.

Having found a transformation of the right type, its naturality has to be proven as a next step. For $f : \Omega \longrightarrow \Omega'$, the transformation needs to satisfy $Ff \circ \mu_\Omega = \mu_{\Omega'} \circ F^2 f$. First, we need the following lemma, based on a result shown and used by Giry [1981] for the case of σ-additive probabilities. For later reuse, it is stated in a more general version than needed in this section, where powersets instead of the general algebras $\mathcal{A}, \mathcal{A}'$ would suffice.

Lemma 6.2.5. *Given a finitely additive probability P on $(\Omega, \mathcal{A})$, an arrow $f : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ and a function $g' : \Omega' \longrightarrow \mathbb{R}$ with a sequence of simple functions $g'_n \rightrightarrows g'$ that converges uniformly to it,*

$$\int g' d(Ff\,P) = \int g' \circ f dP.$$

Proof. First of all, the integral $\int g' d(Ff\,P)$ exists and is equal to the limit of the integrals of the simple functions due to Lemma 6.1.9. Consider one of the g'_n, and assume its

[3] The notation ev does not have anything to do with the evolutions in the formal framework or vulnerability, neither is this an evaluation in the sense of the harm measurement there.

representation is $\sum_{i=1}^{n} c_i \cdot \mathbb{1}_{A'_i}$ for some real numbers c_i and sets $A'_i \in \mathcal{A}'$.

$$\begin{aligned}\int g'_n d(Ff\,P) &= \int \sum_{i=1}^{n} c_i \cdot \mathbb{1}_{A'_i} d(Ff\,P) = \sum_{i=1}^{n} c_i \cdot \int \mathbb{1}_{A_i} d(Ff\,P) = \sum_{i=1}^{n} c_i \cdot Ff\,P(A'_i) \\ &= \sum_{i=1}^{n} c_i \cdot P(f^{-1}(A'_i)) = \sum_{i=1}^{n} c_i \cdot P(\{\omega \in \Omega | f(\omega) \in A'_i\}) = \sum_{i=1}^{n} c_i \cdot \int \mathbb{1}_{A'_i}(f(.))dP \\ &= \int \sum_{i=1}^{n} c_i \mathbb{1}_{A'_i}(f(.))dP = \int g'_n(f(.))dP = \int g'_n \circ f dP.\end{aligned}$$

This can be used (see $(*)$) to show the equality of the integrals with g' in place of g'_n.

$$\int g' d(Ff\,P) = \lim_{n\to\infty} \int g'_n d(Ff\;P) \stackrel{(*)}{=} \lim_{n\to\infty} \int g'_n \circ f dP = \int g' \circ f dP$$

where the last equality follows from the uniform convergence $(g'_n \circ f) \rightrightarrows g' \circ f$ (justified next) and again Lemma 6.1.9 which allows to switch integral and uniform limits. The last uniform convergence is easy to show: given an $\varepsilon > 0$, for all $\omega \in \Omega$

$$|(g'_n \circ f)(\omega) - (g' \circ f)(\omega)| = |g'_n(f(\omega)) - g'(f(\omega))| < \varepsilon \text{ for all } n \geq N \text{ for some } N$$

due to the uniform convergence $g'_n \rightrightarrows g'$. □

For naturality of μ, the equality $Ff \circ \mu_\Omega = \mu_{\Omega'} \circ F^2 f$ has to hold for each $Q \in F^2\Omega$. That is,

$$\begin{gathered}(Ff \circ \mu_\Omega)(Q) = (\mu_{\Omega'} \circ F^2 f)(Q) \text{ or} \\ Ff(\mu_\Omega Q) = \mu_{\Omega'}(F^2 f\,Q)\end{gathered}$$

needs to be shown. Since both sides produce a finitely additive probability on $(\Omega', \mathfrak{P}(\Omega'))$, we need to show equality for application to $A' \subseteq \Omega'$.

$\mu_\Omega Q$ is a finitely additive probability on $(\Omega, \mathfrak{P}(\Omega))$, therefore

$$Ff\,(\mu_\Omega Q)(A') = \mu_\Omega Q(f^{-1}(A')) = \int ev_{f^{-1}(A')} dQ$$

On the righthand side, we have $F^2 f\,Q$ which is a finitely additive probability on $(F\Omega', \mathfrak{P}(F\Omega'))$ to which the respective $\mu_{\Omega'}$ is applied.

$$\mu_{\Omega'}(F^2 f\,Q)(A') = \int ev_{A'} d(F^2 f\,Q).$$

Since $F^2 f$ is $F(Ff)$, we can apply Lemma 6.2.5 to $\int ev_{A'} d(F(Ff)Q)$, because the existence of a sequence of simple functions converging uniformly to $ev_{A'}$ is guaranteed by Corollary 6.1.11. One obtains

$$\mu_{\Omega'}(F^2 f\,Q)(A') = \int ev_{A'} d(F(Ff)\,Q) = \int ev_{A'} \circ Ff\,dQ = \int ev_{f^{-1}(A')} dQ,$$

because the integrand in the second last integral is the function

$$P \stackrel{Ff}{\longmapsto} Ff\;P \stackrel{ev_{A'}}{\longmapsto} Ff\;P(A') = P(f^{-1}(A'))$$

that maps P to $P(f^{-1}(A'))$ as does the function $ev_{f^{-1}(A')}$. So, the equality of the two sides and thus the naturality of μ is shown. Now that all the ingredients to a monad are given, the interaction between the transformations has to be checked. This is done in the following two sections.

6.2.3 Unitarity of η

We have to show that $\mu\circ\eta_F = id_F = \mu\circ F\eta$, that is, for a set Ω, both $\mu_\Omega\circ\eta_{F\Omega}$ and $\mu_\Omega\circ F\eta_\Omega$, when applied to a $P \in F\Omega$, have to return exactly this P. Application to a finitely additive probability P yields

$$(\mu_\Omega \circ \eta_{F\Omega})P = \mu_\Omega(\eta_{F\Omega}P) \text{ on the left, and}$$
$$(\mu_\Omega \circ F\,\eta_\Omega)\,P = \mu_\Omega(F\,\eta_\Omega\,P) \text{ on the right side.}$$

Equality is checked by application to $A \subseteq \Omega$.

Starting on the left hand side, the arrow of the transformation η that is associated with the object $F\Omega$ is a function from $F\Omega$ to $F^2\Omega$. It is given by $\eta_{F\Omega}P(B) = \mathbb{1}_B(P)$ for sets $B \in \mathfrak{P}(F\Omega)$.

$$\mu_\Omega(\eta_{F\Omega}P)(A) = \mu_\Omega(\mathbb{1}_{(.)}(P))(A) = \int ev_A d(\mathbb{1}_{(.)}(P)) = ev_A(P) = P(A)$$

The second last equality here follows from the fact that ev_A is integrable under any Q on $(F\Omega, \mathfrak{P}(F\Omega))$ by Lemma 6.1.6 and that then we can use Lemma 6.1.5.

On the right hand side, we have the finitely additive probability $F\,\eta_\Omega\,P$ defined on $\mathfrak{P}(F\Omega)$ by $F\,\eta_\Omega\,P(B) = P(\eta_\Omega^{-1}(B))$.

$$\mu_\Omega(F\,\eta_\Omega\,P)(A) = \int ev_A d(F\,\eta_\Omega P) = \int ev_A \circ \eta_\Omega dP$$

where Lemma 6.2.5 was used to give the last equality and was allowed to be used because a sequence of simple functions that converge uniformly to ev_A exists. Now, the integrand $ev_A \circ \eta_\Omega$, when applied to an ω, evaluates the finitely additive probability $\mathbb{1}_{(.)}(\omega)$ at $A \subseteq \Omega$, which means it is the function $\omega \longmapsto \mathbb{1}_A(\omega)$. Since $\int \mathbb{1}_A dP = P(A)$ because $\mathbb{1}_A$ is a (very) simple function, we conclude that

$$\mu_\Omega(F\,\eta_\Omega\,P)(A) = P(A) \text{ as well.}$$

Hence,

$$\mu_\Omega(\eta_{F\Omega}P)(A) = P(A) = id_{F\Omega}(P)(A) = \mu_\Omega(F\,\eta_\Omega\,P)(A)$$

for all $P \in F\Omega$ and all $A \in \mathfrak{P}(\Omega)$, so that the first monad property is proven.

6.2.4 Associativity of μ

Here we have to show $\mu\circ\mu_F = \mu\circ F\,\mu$, which, for an object Ω becomes $\mu_\Omega\circ\mu_{F\Omega} = \mu_\Omega\circ F\,\mu_\Omega$. Both sides, when applied to a finitely additive probability R in $F^3\Omega$ must yield the same $P \in F\Omega$, a finitely additive probability on $(\Omega, \mathfrak{P}(\Omega))$.

$$(\mu_\Omega \circ \mu_{F\Omega})(R) = (\mu_\Omega \circ F\,\mu_\Omega)(R) \text{ that is,}$$
$$\mu_\Omega(\mu_{F\Omega}R) = \mu_\Omega(F\,\mu_\Omega\,R)$$

has to be shown. Again, we apply both sides to an $A \subseteq \Omega$. On the left,

$$\mu_\Omega(\mu_{F\Omega}R)(A) = \int ev_A d(\mu_{F\Omega}R).$$

Consider a sequence of simple functions $f_n \rightrightarrows ev_A$, which exists due to Corollary 6.1.11. As $\mu_{F\Omega}R$ is a finitely additive probability on $F\Omega$,

$$\int ev_A d(\mu_{F\Omega}R) = \lim_{n\to\infty} \int f_n d(\mu_{F\Omega}R)$$

by Lemma 6.1.9. Taking a closer look at $\int f_n d(\mu_{F\Omega}R)$ for a fixed n, with $f_n = \sum_{i=1}^m c_i \cdot \mathbb{1}_{B_i}$, one sees that

$$\int f_n d(\mu_{F\Omega}R) = \sum_{i=1}^m c_i \cdot \mu_{F\Omega}R(B_i) = \sum_{i=1}^m c_i \cdot \int ev_{B_i} dR = \int \Big(\sum_{i=1}^m c_i \cdot ev_{B_i}\Big) dR \tag{6.4}$$

where we have used the definition of the integral, the definition of μ and the linearity of the integral.

To proceed, we need another

Lemma 6.2.6. *For $f_n \rightrightarrows ev_A$, define a sequence of functions $g_n : F^2\Omega \longrightarrow [0,1]$ by*

$$g_n(Q) = \int f_n dQ.$$

Let $g : F^2\Omega \longrightarrow [0,1]$ be given by $g(Q) = \int ev_A dQ$. Then, $g_n \rightrightarrows g$ on $F^2\Omega$.

Proof. Given uniform convergence of $f_n \rightrightarrows ev_A$, we can fix $\varepsilon > 0$ and N such that for all $n \geq N$ and all $P \in F\Omega$ one has $|f_n(P) - ev_A(P)| < \varepsilon$. Then, for $n \geq N$ also

$$\begin{aligned} |g_n(Q) - g(Q)| &= \left|\int f_n dQ - \int ev_A dQ\right| = \left|\int f_n - ev_A dQ\right| \leq \int |f_n - ev_A| dQ \\ &\leq \sup_{P \in F\Omega} |f_n(P) - ev_A(P)| < \varepsilon, \end{aligned}$$

where the first $\leq$-step uses Lemma 6.1.7 (*iv*) and the second one is correct because Q is a probability and thus has a mass of 1 to distribute. Since the bound holds for all $Q \in F^2\Omega$, this shows the uniform convergence $g_n \rightrightarrows g$. □

With the notation from this lemma, we can rewrite the last integrand in (6.4), the function $\sum_{i=1}^m c_i ev_{B_i}$ on $F^2\Omega$. Applied to a Q, it yields

$$\sum_{i=1}^m c_i \cdot ev_{B_i}(Q) = \sum_{i=1}^m c_i \cdot Q(B_i) = \int f_n dQ = g_n(Q). \tag{6.5}$$

Substitute this back into (6.4) to obtain

$$\int f_n d(\mu_{F\Omega}R) = \int g_n dR.$$

Recalling that

$$\mu_\Omega(\mu_{F\Omega}R)(A) = \lim_{n\to\infty} \int f_n d(\mu_{F\Omega}R)$$

one obtains

$$\mu_\Omega(\mu_{F\Omega}R)(A) = \lim_{n\to\infty} \int g_n dR$$

for the left side. By the uniform convergence $g_n \rightrightarrows g$ and Lemma 6.1.9, this is equal to $\int g dR$ which the right side can be reduced to. The right side in the equation, applied to A, is

$$\mu(F\mu_\Omega\, R)(A) = \int ev_A d(F\mu_\Omega R) = \int ev_A \circ \mu_\Omega dR \tag{6.6}$$

by an application of Lemma 6.2.5 in the last equality. The last integrand in (6.6) associates with every $Q \in F^2\Omega$ the value $ev_A(\mu_\Omega Q) = \mu_\Omega Q(A) = \int ev_A dQ = g(Q)$. Therefore, it is equal to the function g as defined in Lemma 6.2.6 and

$$\mu(F\mu_\Omega\, R)(A) = \int g dR = \lim_{n\to\infty} \int g_n dR = \mu_\Omega \circ \mu_{F\Omega} R(A), \text{ q.e.d.}$$

In conclusion, the following theorem is proven:

Theorem 6.2.7. *(F, η, μ) as defined above is a monad on the category $\mathbb{F}un$.*

The underlying idea in proving this may have been hidden between the different steps and lemmata. It is an adaptation of the procedure of measure theoretic induction. For simple functions the equalities to be shown are mostly simple manipulations, such as in formula (6.4). For the general case, these were substituted by a sequence of simple functions converging uniformly to them, which exists easily on the powerset as the underlying algebra. Since under uniform convergence limits and integrals may be switched, the results follow for all functions. The uniformly converging sequences will not be as obviously existent for general algebras later.

6.3 Giry's monad of probability measures

This section sketches the monad defined by Giry [1981] and discusses how it relates to the monad from the previous section.

6.3.1 Sketch of the monad

The underlying category for Giry's monad is $\mathbb{M}es$, the category of measurable sets and measurable functions. Objects in this category are sets equipped with σ-algebras, $(\Omega, \mathcal{F})$. Arrows are measurable functions between these, that is, an arrow $f : (\Omega, \mathcal{F}) \longrightarrow (\Omega', \mathcal{F}')$ is a function from Ω to Ω', such that $f^{-1}(A') \in \mathcal{F}$ for all $A' \in \mathcal{F}'$. Identities are the usual identity functions (measurable with respect to any σ-algebra on Ω) and composition is function composition, which preserves measurability. The category properties are thus inherited immediately from these properties in $\mathbb{F}un$.

The probability functor, denoted Π, is defined as follows. To a set with a σ-algebra it associates a set

$$\Pi_\Omega = \{P | P \text{ is a probability measure on } (\Omega, \mathcal{F})\} \tag{6.7}$$

and a σ-algebra, $\mathcal{B}_\Omega$, defined as the initial σ-algebra (see Definition 3.1.10) of the evaluation maps. These are here defined on Π_Ω: for $A \in \mathcal{F}$,

$$ev_A : \Pi_\Omega \longrightarrow [0,1] \text{ is again defined by } ev_A(P) = P(A).$$

Now, $(ev_A)_{A\in\mathcal{F}}$ is a family of functions from Π_Ω to the measurable space $([0,1], \mathcal{B}[0,1])$, where $\mathcal{B}[0,1]$ denotes the Borel-σ-algebra on $[0,1]$. The initial σ-algebra is the smallest σ-algebra on Π_Ω which makes all evaluation maps measureable.

More precisely,

Definition 6.3.1. The probability functor Π associates to a measurable space $(\Omega, \mathcal{F})$

$$\begin{aligned} \Pi(\Omega, \mathcal{F}) &= (\Pi_\Omega, \mathcal{B}_\Omega) \text{ where} \\ \Pi_\Omega &= \{P | P \text{ is a } \sigma\text{-additive probability on } (\Omega, \mathcal{F})\} \text{ and} \\ \mathcal{B}_\Omega &= \sigma\left(\bigcup_{A\in\mathcal{F}} ev_A^{-1}(\mathcal{B}[0,1])\right). \end{aligned}$$

To a measurable function $f : (\Omega, \mathcal{F}) \longrightarrow (\Omega', \mathcal{F}')$, it associates the measurable function $\Pi f : \Pi(\Omega, \mathcal{F}) \longrightarrow \Pi(\Omega', \mathcal{F}')$ that maps any $P \in \Pi_\Omega$ to the induced probability under f:

$$\Pi f\, P(A') = P(f^{-1}(A')) \text{ for each } A' \in \mathcal{F}'. \tag{6.8}$$

Further, the natural transformations η and μ are defined by

$$\eta_\Omega\omega(A) = \mathbb{1}_A(\omega) \text{ for } A \in \mathcal{F}, \text{ and}$$
$$\mu_\Omega Q(A) = \int ev_A dQ \text{ for } A \in \mathcal{F} \text{ and } Q \text{ a probability measure on } (\Pi_\Omega, \mathcal{B}_\Omega),$$

where $\int$ is the Lebesgue integral. Giry shows that (Π, η, μ) is a monad and that the arrows of the Kleisli category of this monad are transition probabilities of a Markov process, that is, a stochastic process in continuous time with the Markov property: given the present state the probability for the future evolution does not depend on the past evolution. In the monadic dynamical system, this generalizes the example of the Markov chain given in Section 5.1.2.

6.3.2 Π and F

The functors Π and F apply to different categories, $\mathbb{M}es$ and $\mathbb{F}un$, respectively. While objects in $\mathbb{F}un$ are sets, an object in $\mathbb{M}es$ has a set, but together with this set there is a σ-algebra of its subsets. That means, each object in $\mathbb{F}un$ can appear as the set in many objects in $\mathbb{M}es$, which differ in the σ-algebra.

Arrows in both categories are functions. An arrow in $\mathbb{F}un$, $f : \Omega \longrightarrow \Omega'$ may or may not be an arrow between two objects $(\Omega, \mathcal{F})$ and $(\Omega', \mathcal{F}')$ in $\mathbb{M}es$, depending on whether the function is $\mathcal{F}$-$\mathcal{F}'$-measurable.

A comparison of the functors Π and F is nonetheless not as far fetched as it may seem, because the "lacking" σ-algebra of the $\mathbb{F}un$-objects is provided by F itself. This was of course already suggested by the similarity in the definitions. The result of applying the functors to an object in both cases has a set.

$$F\Omega = \{P|P \text{ is a finitely additive probability on } (\Omega, \mathfrak{P}(\Omega))\}$$
$$\Pi_\Omega = \{P|P \text{ is a } \sigma\text{-additive probability on } (\Omega, \mathcal{F})\}$$

The powerset is a σ-algebra, so in the case where $\mathcal{F}$ is the powerset of Ω, one can conclude $F\Omega \supset \Pi_\Omega$ because any σ-additive probability is a finitely additive probability. It was seen that for uncountable Ω, countably additive (probability) measures can be defined only if discrete. That is, the choice of $\mathcal{F}$ as the powerset for the Giry monad would restrict the set of measures, wherefore this case is not of much interest. Even if Ω is finite or countably infinite, the set Π_Ω is very big as can be seen even for a simple example. Consider an Ω with two elements, say $\Omega = \{a, b\}$. Probabilities on the powerset of this Ω (with four elements) are determined by the values they give to a and b and these have to add up to one. The set of all these is $\Pi_\Omega = \{P|P(a) = p, P(b) = 1 - p, p \in [0, 1]\}$. This means there are as many of these P as real numbers in $[0, 1]$. Therefore, using the powerset as the algebra to go with the set of probability measures would lead one into the same trouble at the next level of functor application, that is, Π^2. The smaller σ-algebra $\mathcal{B}_\Omega$ on Π_Ω that Giry uses is essential to her monad construction.

To obtain a monad that better corresponds to the one by Giry, we will in the next but one section define another monad of finitely additive probabilities which applies to a supercategory of $\mathbb{M}es$ and "extends" Π. The next section provides the preparations for this.

6.4 Finitely measurable functions

Measurability of functions allows to transfer measures from one measurable space to another one. Since we want to do this here with finitely additive probabilities, we introduce a similar concept for functions between sets which are equipped merely with algebras, and call this *finite measurability.*

Definition 6.4.1. Given two sets with algebras, $(\Omega, \mathcal{A})$ and $(\Omega', \mathcal{A}')$, a function $f : \Omega \longrightarrow \Omega'$ is *finitely measurable*, or more precisely $\mathcal{A}$-$\mathcal{A}'$-finitely measurable, if

$$f^{-1}(A') \in \mathcal{A} \text{ for all } A' \in \mathcal{A}'.$$

Every measurable function is therefore finitely measurable but not vice versa. When wanting to emphasize that the stronger condition is meant, we attach the qualifyier "countable" to measurability in the usual sense. Unfortunately, finite measurability lacks most of the nice properties of measurability, which is probably why the concept does not appear in the literature. As mentioned in Section 3.2, measure theory is carried out only for measurable spaces and measurable functions, while algebras and finitely additive functions are not considered. However, here this concept provides useful in some steps towards a monad of finitely additive probabilities that can generalize the monad of σ-additive probabilities by Giry.

Finitely measurable functions from $(\Omega, \mathcal{A})$ to $([0,1], \mathcal{J}[0,1])$ are of special interest for our monad. Simple functions for algebras are defined analogously to Definition 3.2.6. Some properties of finite measurability are the following.

Lemma 6.4.2. *Consider an* $(\Omega, \mathcal{A})$ *and* $\mathbb{R}$ *equipped with* $\mathcal{J}[0,1]$*, and* $\mathcal{A}$-$\mathcal{J}[0,1]$*-finite measurability.*

(*i*) *Indicator functions of sets in* $\mathcal{A}$ *are finitely measurable.*

(*ii*) *Simple functions are finitely measurable.*

(*iii*) *For* $\alpha \in \mathbb{R}$*, the function* αf *is finitely measurable if* f *is (assuming* α *is such that also* $\alpha f : \Omega \longrightarrow [0,1]$*).*

Proof.

(*i*) $\mathbb{1}_A^{-1}(J)$ is in $\mathcal{A}$ for any $J \in \mathcal{J}[0,1]$ because

$$\mathbb{1}_A^{-1}(J) = \begin{cases} \emptyset & \text{if } 0 \notin J, 1 \notin J \\ A & \text{if } 0 \notin J, 1 \in J \\ A^c & \text{if } 0 \in J, 1 \notin J \\ \Omega & \text{if } 0 \in J, 1 \in J \end{cases}$$

and since $\mathcal{A}$ is an algebra, with $A \in \mathcal{A}$ all the results are in $\mathcal{A}$.

(*ii*) A simple function is of the form $f = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$ for $A_i \in \mathcal{A}$. The inverse image of a set $J \in \mathcal{J}[0,1]$ is

$$\left(\sum_{i=1}^n c_i \mathbb{1}_{A_i}\right)^{-1}(J) = \left\{\omega \in \Omega \,\middle|\, \sum_{i=1}^n c_i \mathbb{1}_{A_i}(\omega) \in J\right\} = \bigcup_{i: c_i \in J} A_i.$$

Being a finite union of sets in $\mathcal{A}$, this is a set in $\mathcal{A}$.

(*iii*) For $\alpha \in \mathbb{R}$ such that αf has the range $[0,1]$, it suffices to show that the inverse image of an interval is in $\mathcal{A}$, due to Lemma 3.1.8. This is the case, because

$$(\alpha f)^{-1}([a,b]) = \{\omega \in \Omega \,|\, \alpha f(\omega) \in [a,b]\} = \{\omega \in \Omega \,|\, f(\omega) \in [a/\alpha, b/\alpha]\}.$$

This is in $\mathcal{A}$ whenever f is $\mathcal{A}$-$\mathcal{J}[0,1]$-finitely measurable. □

One of the major differences with countable measurability is this:

Lemma 6.4.3. *Pointwise limits of finitely measurable functions need not be finitely measurable.*

Proof. Consider $\Omega = [0,1)$ and the algebra generated by the half-open intervals $[a,b)$ with $a,b \in [0,1], a < b$, denoted by $\mathcal{A}$. This $\mathcal{A}$ is less similar to $\mathcal{J}[0,1]$ than it might seem: It does not contain points, because of the constraint that $a < b$ and because, with complements and finite intersections and unions only, one obtains half-open intervals or disjoint unions of these only. A sequence of finitely measurable functions converging to a function that is not finitely measurable can be constructed by considering indicator functions of half-open intervals converging to the indicator function of a point. Let $f_n : \Omega \longrightarrow \{0,1\}$ and equip the target with the algebra $\{\emptyset, \{0\}, \{1\}, \{0,1\}\}$. Let $f_n = \mathbb{1}_{[a,a+1/n)}$, then f_n is finitely measurable for each n. For example, the inverse image $f_n^{-1}(\{1\}) = [a, a+1/n)$ is an element of $\mathcal{A}$ for any n, and similarly for the three other sets. However, the f_n converge to the function $f = \mathbb{1}_{\{a\}}$, which is not finitely measurable, because $f^{-1}(\{1\}) = \{a\} \notin \mathcal{A}$. □

6.4.1 Connections between integrability and finite measurability

As the Lebesgue integral is by definition applied to measurable functions, the question how integrability relates to measurability is not an issue. Here, the integral is defined without any measurability assumptions, but the concept of finite measurability will be used in the construction of the monad. Therefore, we need some results on how finite measurability relates to integrability.

For the integral from Section 6.1, $\mathcal{A}$-$\mathcal{J}[0,1]$-finite measurability guarantees integrability:

Lemma 6.4.4. *If $f^{-1}(J) \in \mathcal{A}$ for all $J \in \mathcal{J}[0,1]$, the function $f : \Omega \longrightarrow [0,1]$ is integrable in the sense of Definition 6.1.1.*

Proof. We show that condition (ii) of Theorem 6.1.4 is satisfied; integrability follows from the equivalence in the theorem. Given $\varepsilon > 0$, one can construct a partition of Ω in $\mathcal{A}$ with the desired property via the inverse image mapping. Divide $[0,1]$ into intervals of length $< \varepsilon$. Say, $[0, \frac{1}{N}), [\frac{1}{N}, \frac{2}{N}), \ldots, [\frac{N-1}{N}, 1]$ for an $N > \frac{1}{\varepsilon}$. The intervals are elements of $\mathcal{J}[0,1]$, thus their inverse images $f^{-1}\big([\frac{i-1}{N}, \frac{i}{N})\big) =: A_i$ are in $\mathcal{A}$ by assumption. The A_i partition Ω, because the inverse image is well-behaved. And, by definition, $f(\omega) \in [\frac{i-1}{N}, \frac{i}{N})$ (respectively in $[\frac{N-1}{N}, 1]$ if $i = N$) for all $\omega \in A_i$, which means that $|f(\omega) - f(\omega')| \leq \frac{1}{N} < \varepsilon$ for all $\omega, \omega' \in A_i$ for all $i = 1, \ldots, N$. Thus (ii) is satisfied. □

6.5 Monad: finitely additive probabilities on fixed algebras

Given that any σ-additive probability is a finitely additive probability, the question arises whether a monad of finitely additive probabilities can be defined on a suitable category containing $\mathbb{M}es$ in a way that is somehow compatible with Giry's monad of probability measures. General algebras instead of the powerset as a default choice have to be allowed as domains of definition of finitely additive probabilities. Since the definition of (interesting) σ-additive probabilities on big sets presupposes smaller σ-algebras than the powerset, also on sets of finitely additive probabilities, the default choice of the powerset as algebra must be dropped. This section presents a monad of finitely additive probabilities which applies to a "finite" analogue and at the same time supercategory of the category $\mathbb{M}es$. Notation is used as in section 6.2.

6.5.1 The category $\mathbb{F}Mes$

As the underlying category, we use a "finite version of" the category of measurable spaces with measurable functions: an object is a set equipped with an algebra and arrows are finitely measurable functions. This allows to "map" probabilities from one object to another using the well-known construction of induced measures. The category is denoted $\mathbb{F}Mes$.

Objects An object in the category $\mathbb{F}Mes$ is a set equipped with an algebra of subsets. It is denoted $(\Omega, \mathcal{A})$, where Ω is a non-empty set and $\mathcal{A}$ is an algebra of subsets of Ω.

Arrows An arrow $f : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ is a function $f : \Omega \longrightarrow \Omega'$ that satisfies the following condition

$$f^{-1}(A') \in \mathcal{A} \text{ for all } A' \in \mathcal{A}'. \tag{6.9}$$

The source of such an arrow is the object $(\Omega, \mathcal{A})$, the target $(\Omega', \mathcal{A}')$. In the terminology of Section 6.4, an arrow in this category is a finitely measurable function.

Identities The identity arrow $id_{(\Omega,\mathcal{A})} : (\Omega, \mathcal{A}) \longrightarrow (\Omega, \mathcal{A})$ associated to an object $(\Omega, \mathcal{A})$ is the identity function on the set Ω. As the inverse image of any subset under the identity is that subset itself, this function satisfies (6.9) and is thus an arrow of $\mathbb{F}Mes$.

Composition In this category, composition is defined as the composition of functions. For two arrows $f : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ and $g : (\Omega', \mathcal{A}') \longrightarrow (\Omega'', \mathcal{A}'')$, the resulting function $g \circ f : \Omega \longrightarrow \Omega''$ is an arrow $g \circ f : (\Omega, \mathcal{A}) \longrightarrow (\Omega'', \mathcal{A}'')$ of $\mathbb{F}Mes$ because

$$(g \circ f)^{-1}(A'') = f^{-1}(g^{-1}(A'')) \in \mathcal{A} \text{ for all } A'' \in \mathcal{A}''$$

since $g^{-1}(A'') \in \mathcal{A}'$ for all $A'' \in \mathcal{A}''$ and $f^{-1}(A') \in \mathcal{A}$ for all $A' \in \mathcal{A}'$.

Because composition is defined as composition of functions, $\mathbb{F}Mes$ inherits the two properties required for a category from the category of $\mathbb{F}un$ of sets and functions: composition is associative and it has the identity arrows as units. That is, for arrows $f : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$, $g : (\Omega', \mathcal{A}') \longrightarrow (\Omega'', \mathcal{A}'')$ and $h : (\Omega'', \mathcal{A}'') \longrightarrow (\Omega''', \mathcal{A}''')$,

$$h \circ (g \circ f) = (h \circ g) \circ f \quad \text{and}$$
$$id_{(\Omega,\mathcal{A})} \circ f = f = f \circ id_{(\Omega,\mathcal{A})}.$$

Therefore, $\mathbb{F}Mes$ is a category.

It contains $\mathbb{M}es$ as a subcategory: any object in $\mathbb{M}es$ is an object in $\mathbb{F}Mes$, because any σ-algebra is an algebra, and an arrow of $\mathbb{M}es$ is also an arrow of $\mathbb{F}Mes$, since measurability implies finite measurability. It is also a full subcategory as in Definition 3.3.2, because in the case where the algebras of two objects in $\mathbb{F}Mes$ are σ-algebras, i.e. the object is an object of $\mathbb{M}es$ also, finite measurability corresponds precisely to measurability, which means both categories contain the same arrows between the objects.

6.5.2 The functor

The finitely additive probability endofunctor on the category $\mathbb{F}Mes$ is denoted by G. It associates to each object $(\Omega, \mathcal{A})$ an object $G(\Omega, \mathcal{A})$ and to each arrow $f : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ an arrow $Gf : G(\Omega, \mathcal{A}) \longrightarrow G(\Omega', \mathcal{A}')$, defined as follows.

Objects Fix an object $(\Omega, \mathcal{A})$. Then $G(\Omega, \mathcal{A})$ is a pair $(\Psi, \mathcal{B})$ of a set and an algebra of subsets of this set. The set is the obvious set of all finitely additive probabilities on the given object,

$$\Psi = \{P | P \text{ is a finitely additive probability on } (\Omega, \mathcal{A})\}.$$

For the definition of the algebra $\mathcal{B}$, we recur to the initial σ-algebra of the evaluation maps from definition 6.3.1. This choice will be motivated in Section 6.5.5. The algebra of the object $G(\Omega, \mathcal{A})$ is

$$\mathcal{B} = \sigma\left(\bigcup_{A\in\mathcal{A}} ev_A^{-1}(\mathcal{B}[0,1])\right)$$

Notation 6.5.1. In the following, the notation $(\Psi, \mathcal{B})$ will be used for the object $G(\Omega, \mathcal{A})$ and extended in the obvious way, for example, $G(\Omega', \mathcal{A}') = (\Psi', \mathcal{B}')$ and so on.

Arrows For arrows $f : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$, we define $Gf : G(\Omega, \mathcal{A}) \longrightarrow G(\Omega', \mathcal{A}')$ as a function $Gf : \Psi \longrightarrow \Psi'$ that takes finitely additive probabilities on $(\Omega, \mathcal{A})$ into finitely additive probabilities on $(\Omega', \mathcal{A}')$ and satifies condition (6.9) for the respective algebras $\mathcal{B}$ and $\mathcal{B}'$. For $P \in \Psi$,

$$Gf\, P(A') = P(f^{-1}(A')) \text{ for each } A' \in \mathcal{A}'.$$

The three properties shown for the functor F in Lemma 6.2.2 are satisfied for G as well. That is, $Gf\, P$ is an element of Ψ', and G preserves identities and composition. Except for a substitution of the algebras $\mathcal{A}$ and $\mathcal{A}'$ instead of the powersets, the proofs are identical to the proof of that lemma.

To establish that G is a functor, one needs to show in addition that Gf is an arrow of the category $\mathbb{F}Mes$, that means $(Gf)^{-1}(B') \in \mathcal{B}$ for any $B' \in \mathcal{B}'$ needs to hold. Here, one can make use of Lemma 3.1.8 and the remark following it. The algebra $\mathcal{B}'$ on Ψ' is defined as

$$\mathcal{B}' = \sigma\Big(\bigcup_{A' \in \mathcal{A}'} ev_{A'}^{-1}(\mathcal{B}[0,1]) \Big).$$

Therefore, it suffices to show

$$(Gf)^{-1}(B') \in \mathcal{B} \text{ for } B' \in \bigcup_{A' \in \mathcal{A}'} ev_{A'}^{-1}(\mathcal{B}[0,1]).$$

For such a B' there is an $A' \in \mathcal{A}'$, such that $B' = ev_{A'}^{-1}(C)$ for some Borel set $C \subseteq [0,1]$. Thus,

$$(Gf)^{-1}(B') = (Gf)^{-1}\big(ev_{A'}^{-1}(C)\big).$$

Now, $ev_{A'}^{-1}(C)$ is a set of the form $\{P' \in \Psi' | P'(A') \in C\}$. Its inverse image under Gf consists of those $P \in \Psi$ that Gf maps to a P' in this set:

$$\begin{aligned}(Gf)^{-1}(ev_{A'}^{-1}(C)) &= \{P \in \Psi | Gf\, P(A') \in C\} \\ &= \{P \in \Psi | P(f^{-1}(A')) \in C\}.\end{aligned}$$

Since sets of the form

$$\{P \in \Psi | P(A) \in D\} \text{ for some } A \in \mathcal{A} \text{ and some Borel set } D$$

are in $\mathcal{B}$, and $f^{-1}(A') \in \mathcal{A}$ because f is an arrow of $\mathbb{F}Mes$, also $\{P \in \Psi | P(f^{-1}(A')) \in C\} \in \mathcal{B}$. Therefore, Gf is an arrow of $\mathbb{F}Mes$, and in conclusion G is a functor.

6.5.3 The natural transformations

The natural transformations are defined very similarly to those of the functor F, with restricted domains of definition. Showing that the functions defined for each object qualify as the transformations η and μ involves showing that they are arrows of $\mathbb{F}Mes$.

For an object $(\Omega, \mathcal{A})$ of $\mathbb{F}Mes$, we define a function $\eta_{(\Omega,\mathcal{A})} : \Omega \longrightarrow \Psi$. This means, for $\omega \in \Omega$, $\eta_{(\Omega,\mathcal{A})}\omega$ needs to be a finitely additive probability on $(\Omega, \mathcal{A})$. Obviously, the chosen finitely additive probability is again the concentrated probability function, here restricted to the given algebra. We overload the indicator notation,

$$\eta_{(\Omega,\mathcal{A})}\omega(A) = \mathbb{1}_A(\omega) \text{ defined for } A \in \mathcal{A}.$$

$\eta_{(\Omega,\mathcal{A})}\omega$ is a finitely additive probability on $\mathcal{A}$ by the same arguments as above for F, and the notation $\mathbb{1}_{(.)}(\omega)$ will also be used as before.

Showing finite measurability of $\eta_{(\Omega,\mathcal{A})}$ from $(\Omega,\mathcal{A})$ to $G(\Omega,\mathcal{A}) = (\Psi,\mathcal{B})$ is the step that differs from the analogous section for the functor F, where this was not necessary. Again we use Lemma 3.1.8 to show that for all $B \in \mathcal{B}$ the inverse image $\eta^{-1}_{(\Omega,\mathcal{A})}(B)$ is in $\mathcal{A}$: it suffices to consider sets B in the generator of $\mathcal{B}$, that is, sets $B = ev_A^{-1}(C)$ with $A \in \mathcal{A}$ and $C \in \mathcal{B}[0,1]$.

$$\begin{aligned}
\eta^{-1}_{(\Omega,\mathcal{A})}(B) &= \eta^{-1}_{(\Omega,\mathcal{A})}(ev_A^{-1}(C)) = \{\omega|\eta_{(\Omega,\mathcal{A})}\omega \in ev_A^{-1}(C)\} \\
&= \{\omega|\eta_{(\Omega,\mathcal{A})}\omega \in \{P \in \Psi|P(A) \in C\}\} = \{\omega|\eta_{(\Omega,\mathcal{A})}\omega(A) \in C\} \\
&= \begin{cases} \emptyset & \text{if } 0 \notin C, 1 \notin C \\ A & \text{if } 0 \notin C, 1 \in C \\ A^c & \text{if } 0 \in C, 1 \notin C \\ \Omega & \text{if } 0 \in C, 1 \in C. \end{cases}
\end{aligned} \tag{6.10}$$

Since $\mathcal{A}$ is an algebra, and $A \in \mathcal{A}$ was assumed, this is in all four cases an element of $\mathcal{A}$, wherefore (6.9) holds. In conclusion, $\eta_{(\Omega,\mathcal{A})} : \Omega \longrightarrow \Psi$ defines an arrow of the category $\eta_{(\Omega,\mathcal{A})} : (\Omega,\mathcal{A}) \longrightarrow G(\Omega,\mathcal{A})$ for each object and hence is a transformation of the desired type, $\eta : Id \longrightarrow G$.

To prove naturality of η, it suffices to look back at equation (6.2) in Section 6.2.2. Given an arrow $f : (\Omega,\mathcal{A}) \longrightarrow (\Omega',\mathcal{A}')$, the equation $Gf\,(\eta_{(\Omega,\mathcal{A})}\omega)(A') = \eta_{(\Omega',\mathcal{A}')}(f(\omega))(A')$ holds for all $A' \in \mathcal{A}'$ by exactly the same reasoning as applied to F and $A' \in \mathfrak{P}(\Omega')$ there. It follows that $Gf \circ \eta_{(\Omega,\mathcal{A})} = \eta_{(\Omega',\mathcal{A}')} \circ f$ and hence that η is a natural transformation.

Similarly, μ is defined as before for F, but the finitely additive probabilities it applies to are defined on $\mathcal{B}^2$, no longer on the powerset over the set of finitely additive probabilities. Therefore, the integrability of the evaluation maps cannot be taken for granted here. In addition, the finite measurability of μ has to be established.

We denote the object obtained by applying G to $G(\Omega,\mathcal{A}) = (\Psi,\mathcal{B})$ by $(\Psi^2,\mathcal{B}^2)$ or $G^2(\Omega,\mathcal{A})$ and elements of Ψ^2 again by Q.

$$\begin{aligned}
\Psi^2 &= \{Q|Q \text{ is a finitely additive probability on } (\Psi,\mathcal{B})\} \text{ and} \\
\mathcal{B}^2 &= \sigma\Big(\bigcup_{B\in\mathcal{B}} ev_B^{-1}(\mathcal{B}[0,1])\Big) \qquad \text{where for each } B \in \mathcal{B}, \\
&\quad ev_B \text{ is the evaluation map } ev_B : \Psi^2 \longrightarrow [0,1],\ ev_B(Q) = Q(B).
\end{aligned}$$

For an object of $\mathbb{F}Mes$, $(\Omega,\mathcal{A})$, a function $\mu_{(\Omega,\mathcal{A})} : \Psi^2 \longrightarrow \Psi$, that "collapses a Q into a P", is now defined for $A \in \mathcal{A}$, but by the same formula as above,

$$\mu_\Omega Q(A) = \int ev_A dQ. \tag{6.11}$$

The integral is still well-defined for each $A \in \mathcal{A}$, because $\mathcal{J}[0,1] \subset \mathcal{B}[0,1]$, which means that the inverse images of sets in $\mathcal{J}[0,1]$ under every ev_A are in $\mathcal{B}$ by definition of $\mathcal{B}$. Integrability follows from Lemma 6.4.4.

That the function on $\mathcal{A}$ defined in (6.11) is a finitely additive probability on $(\Omega,\mathcal{A})$ follows like for the functor F (see after Definition 6.2.4).

To show that $\mu_{(\Omega,\mathcal{A})}$ is finitely measurable, we use the following lemma. The first and last statement were "for free" in the case of probabilities defined on the powersets; the second statement was already proven there in Lemma 6.2.6, and is here stated again only for completeness because the domain of definition of the functions changed.

Lemma 6.5.2.

(*i*) *For each* $A \in \mathcal{A}$ *there exists a sequence of simple functions* $f_n : \Psi \longrightarrow [0,1]$ *that converges to* ev_A uniformly *over* Ψ.

(*ii*) *For a sequence* $f_n \rightrightarrows ev_A$*, the functions* $g_n : \Psi^2 \longrightarrow [0,1]$ *defined by* $g_n(Q) = \int f_n dQ$ *converge uniformly to the function* $g : \Psi^2 \longrightarrow [0,1]$ *defined by* $g(Q) = \int ev_A dQ$.

(*iii*) *Given a finitely additive probability* R *on* $(\Psi^2, \mathcal{B}^2)$*, the integrals in the following formula are well-defined and* $\lim_{n\to\infty} \int g_n dR = \int g dR$.

Proof.

(*i*) The sequence of functions is constructed via ev_A's inverse image. As in the proof of Lemma 6.4.4, take a partition of $[0,1]$ into N intervals, $[0, \frac{1}{N}), [\frac{1}{N}, \frac{2}{N}), \ldots, [\frac{N-1}{N}, 1]$, for some $N \in \mathbb{N}$ and consider their inverse images under the given function ev_A. We denote these by $B_i, i = 1, \ldots, N$. Recall that all $B_i \in \mathcal{B}$ by definition of $\mathcal{B}$ and that the B_i partition Ψ. Now, fix a value in each little interval in $[0,1]$, for simplicity, we take the middle point $\xi_i := (i - \frac{1}{2})/N$. Define the following simple function on Ψ:

$$f(P) = \sum_{i=1}^{N} \xi_i \cdot \mathbb{1}_{B_i}(P).$$

If this procedure is carried out for each $n \in \mathbb{N}$ with $N = n$, one obtains a sequence of simple functions $f_n(P) = \sum_{i=1}^{n} \xi_i^n \cdot \mathbb{1}_{B_i^n}(P)$. By construction, this sequence converges uniformly to ev_A on Ψ: for any $\varepsilon > 0$ there is an $M \in \mathbb{N}$ such that for all $n \geq M$ and all $P \in \Psi$

$$|ev_A(P) - f_n(P)| < \varepsilon \text{ because}$$
$$|ev_A(P) - f_n(P)| = |P(A) - \xi_i^n| \text{ if } P \in B_i^n.$$

But $P \in B_i^n$ means that $P \in ev_A^{-1}\Big([\frac{i-1}{n}, \frac{i}{n})\Big)$ which in turn means that $P(A) \in [\frac{i-1}{n}, \frac{i}{n})$. Hence, the distance between $P(A)$ and ξ_i^n is always less than $1/n$. So, any choice of $M > 1/\varepsilon$ yields the result.

(*ii*) The proof can be copied from that of Lemma 6.2.6. The only change is the domain of definition of the functions g_n and g. Since all integrals are well-defined, the proof does not change.

(*iii*) One only needs to show that $\int g_n dR$ is well-defined for all n. Then Lemma 6.1.9 does the rest: since $g_n \rightrightarrows g$, Lemma 6.1.9 guarantees that $\int g dR$ is well-defined and that $\lim_{n\to\infty} \int g_n dR = \int g dR$.

Consider, for a fixed n, $g_n(Q) = \int f_n dQ$. To simplify the above notation, let $f_n = \sum_{i=1}^{m} c_i \cdot \mathbb{1}_{B_i}$ for sets $B_i \in \mathcal{B}$ and values $c_i \in [0,1]$. Since f_n is simple, its integral with respect to Q is

$$\int f_n dQ = \sum_{i=1}^{m} c_i \cdot Q(B_i) = \sum_{i=1}^{m} c_i \cdot ev_{B_i}(Q)$$

and thus the function g_n can be written as $g_n = \sum_{i=1}^{m} c_i \cdot ev_{B_i}$. Every ev_{B_i} is integrable with respect to R (in fact, with respect to any finitely additive probability on $\mathcal{B}^2$) by definition of $\mathcal{B}^2$ and Lemma 6.4.4: again, inverse images of $\mathcal{J}[0,1]$-sets are in $\mathcal{B}^2$. From the existence of $\int ev_{B_1} dR, \ldots, \int ev_{B_m} dR$ follows that of

$$\sum_{i=1}^{m} c_i \cdot \int ev_{B_i} dR = \int \sum_{i=1}^{m} c_i \cdot ev_{B_i} dR = \int g_n dR$$

by linearity of the integral (Lemma 6.1.7 (*ii*) and (*iii*)). □

For finite measurability of μ, it needs to be shown that $\mu^{-1}_{(\Omega,\mathcal{A})}(B) \in \mathcal{B}^2$ for all $B \in \mathcal{B}$. As before, using Lemma 3.1.8, it suffices to show this for B in the generator, that is, show $\mu^{-1}_{(\Omega,\mathcal{A})}(ev_A^{-1}(C)) \in \mathcal{B}^2$ for Borel sets C and $A \in \mathcal{A}$. Now,

$$\begin{aligned} \mu^{-1}_{(\Omega,\mathcal{A})}(ev_A^{-1}(C)) &= \{Q \in \Psi^2 | \mu_{(\Omega,\mathcal{A})}Q \in (ev_A^{-1}(C))\} \\ &= \{Q \in \Psi^2 | \mu_{(\Omega,\mathcal{A})}Q(A) \in C\} \\ &= \{Q \in \Psi^2 | \int ev_A dQ \in C\} \\ &= \{Q \in \Psi^2 | g(Q) \in C\} = g^{-1}(C) \end{aligned} \tag{6.12}$$

with g defined as in Lemma 6.5.2. Consider also the corresponding g_n for a sequence $f_n \rightrightarrows ev_A$ as in the lemma.

The task has been reduced to showing that $g^{-1}(C) \in \mathcal{B}^2$ for Borel sets $C \subseteq [0,1]$, in other words that g is a (finitely) measurable function from $(\Psi^2, \mathcal{B}^2)$ to $([0,1], \mathcal{B}[0,1])$. With the definition of the algebra $\mathcal{B}^2$ as a σ-algebra (and the Borel sets being a σ-algebra, of course), we can use the standards from measure theory. The functions g_n are measurable: they can be written as $g_n(Q) = \sum_{i=1}^m c_i ev_{B_i}(Q)$ for some natural number m and $c_i \in [0,1]$ and sets $B_i \in \mathcal{B}^2$ (all deriving from the f_n under consideration), as shown in equation (6.5); all ev_{B_i} are measurable by definition of $\mathcal{B}^2$ and linear combinations of measurable functions to $\mathbb{R}$ with the Borel sets are measurable (Lemma 3.2.5). With the g_n, also their pointwise (because even uniform) limit g is measurable due to the same lemma. Therefore, each $\mu_{(\Omega,\mathcal{A})}$ is an arrow of our category $\mathbb{F}Mes$, and hence μ is a transformation.

The proof of naturality of μ is again analogous to the proof for the functor F at the end of Section 6.2.2. Lemma 6.2.5 was shown in the generality needed to apply it here as well. Given an arrow $f : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$ of $\mathbb{F}Mes$, for each $Q \in \Psi^2$ and each $A' \in \mathcal{A}'$,

$$Gf\,(\mu_{(\Omega,\mathcal{A})}Q)(A') = \mu_{(\Omega',\mathcal{A}')}(G^2 f\, Q)(A')$$

because both sides are equal to $\int ev_{f^{-1}(A')} dQ$. The proof can be obtained by simple substitutions of the algebras for the powersets and G for F at the end of Section 6.2.2 and will therefore not be repeated here.

6.5.4 Monad properties

This short section generalizes the corresponding results from Sections 6.2.3 and 6.2.4.

For all objects $(\Omega, \mathcal{A})$ of $\mathbb{F}Mes$,

$$\mu_{(\Omega,\mathcal{A})} \circ \eta_{G(\Omega,\mathcal{A})} \;= id_{G(\Omega,\mathcal{A})} = \mu_{(\Omega,\mathcal{A})} \circ G\,\eta_{(\Omega,\mathcal{A})}$$

holds because for $P \in \Psi$ and all $A \in \mathcal{A}$

$$\mu_{(\Omega,\mathcal{A})}(\eta_{G(\Omega,\mathcal{A})}\,P)(A) = P(A) = id_{G(\Omega,\mathcal{A})}(P)(A) = \mu_{(\Omega,\mathcal{A})}(G\,\eta_{(\Omega,\mathcal{A})}\,P)(A)$$

follows from almost the same steps as made in Section 6.2.3. For the left side, integrability of ev_A with respect to the finitely additive probability $(\mathbb{1}_{(.)}(P))$ on $\mathcal{B}$ is guaranteed by Lemma 6.4.4 in this case, which allows to conclude that the integral is the function value at P by Lemma 6.1.5. On the right hand side, one is allowed to exchange the functor application with the integration (Lemma 6.2.5) because Lemma 6.5.2 ensures the existence of simple functions that uniformly converge to ev_A here.

Similarly, $\mu_{(\Omega,\mathcal{A})} \circ \mu_{G(\Omega,\mathcal{A})} = \mu_{(\Omega,\mathcal{A})} \circ G\mu_{(\Omega,\mathcal{A})}$ for all objects $(\Omega, \mathcal{A})$ of $\mathbb{F}Mes$. The proof given in Section 6.2.4 can be adapted step by step to G, because Lemma 6.5.2 provides the neccesary results for G.

In conclusion, also the following theorem holds:

Theorem 6.5.3. (G, η, μ) *as defined in this section are a monad on the category* $\mathbb{F}Mes$.

6.5.5 G extends Giry's Π

In what sense is G an extension of Giry's monad? Given an object $(\Omega, \mathcal{F})$ of $\mathbb{M}es$ and hence also of $\mathbb{F}Mes$, let us compare what the functors associate to it.

$$\Pi(\Omega, \mathcal{F}) = (\Pi_\Omega, \mathcal{B}_\Omega)$$
$$G(\Omega, \mathcal{F}) = (\Psi, \mathcal{B})$$

That Π_Ω is a subset of Ψ is clear because any σ-additive probability qualifies also as a finitely additive probability. The σ-algebras have the following relation: they are both defined as the smallest σ-algebra that makes the evaluation maps of all $A \in \mathcal{F}$ measurable, however, the sources of the evaluation maps are not the same in the two cases. To make this explicit in the notation, let

$$ev_A : \Pi_\Omega \longrightarrow [0,1] \text{ and}$$
$$\tilde{ev}_A : \Psi \longrightarrow [0,1]$$

be the evaluation maps for the two cases. Then, ev_A is the restriction of $\tilde{ev}_A$ to Π_Ω.

Consider a Borel-set $C \subseteq [0,1]$.

$$ev_A^{-1}(C) = \{P \in \Pi_\Omega | P(B) \in C\}$$
$$\tilde{ev}_A^{-1}(C) = \{P \in \Psi | P(B) \in C\} = \{P \in \Pi_\Omega | P(B) \in C\} \cup \{P \in \Psi \backslash \Pi_\Omega | P(B) \in C\}$$

so that each set in the generator of $\mathcal{B}_\Omega$ is the restriction of the corresponding set in the generator of $\mathcal{B}$ to those elements that are in Π_Ω. That is, one can consider $\mathcal{B}$ an extension of $\mathcal{B}_\Omega$ in some sense.

6.5.6 Monad with a smaller algebra

If one considers the functor G on the category $\mathbb{F}Mes$ independently from the Giry monad, the construction of the σ-algebra $\mathcal{B}$ on the set of finitely additive probabilities, Ψ, seems rather far-fetched. When defining a functor on $\mathbb{F}Mes$, one needs to provide an algebra with the set of finitely additive probabilities on a given object $(\Omega, \mathcal{A})$. For finitely additive probabilities only, without considering σ-additivity, again, the easiest and probably most natural choice of an algebra on Ψ would be its powerset. One can define a functor H on $\mathbb{F}Mes$, which associates to an $(\Omega, \mathcal{A})$ the object $(\Psi, \mathfrak{P}(\Psi))$, and treats arrows just as G did. Most finite measurability questions for H are easy to solve, because the inverse image of anything is always contained in the powerset on the source of the function under consideration. However, defining $\eta_{(\Omega,\mathcal{A})} : (\Omega, \mathcal{A}) \longrightarrow (\Psi, \mathfrak{P}(\Psi))$ as the concentrated probability

$$\eta_{(\Omega,\mathcal{A})}\omega(A) = \mathbb{1}_A(\omega) \text{ for } A \in \mathcal{A}, \tag{6.13}$$

it is not necessarily a finitely measurable function. One would have to show that for any $B \in \mathfrak{P}(\Psi)$, the inverse image under $\eta_{(\Omega,\mathcal{A})}$ is an element of the algebra $\mathcal{A}$, that is, $\eta^{-1}_{(\Omega,\mathcal{A})}(B) \in \mathcal{A}$. However, in particular if $\mathcal{A}$ is a small algebra, one can certainly find sets $B \in \mathfrak{P}(\Psi)$ such that $\eta^{-1}_{(\Omega,\mathcal{A})}(B) = \{\omega \in \Omega | \mathbb{1}_{(.)}(\omega) \in B\} \notin \mathcal{A}$. That is, the scheme used to construct the monad G does not make H a monad.

A more interesting question is how small the algebra on Ψ may be so that one still obtains a monad on $\mathbb{F}Mes$. Sticking to the construction using the inverse images under the evaluation maps of an algebra on $[0,1]$, because it has worked fine so far, two questions come up. Can a smaller algebra be chosen on $[0,1]$? And can the initial σ-algebra be reduced to an analogously defined "initial algebra"?

We answer the first question positively: In the σ-additive case the choice of the Borel-σ-algebra on $[0,1]$ is so to say automatic. The "natural" choice of an algebra on $[0,1]$ that need not be a σ-algebra might be $\mathcal{J}[0,1]$. Using this in the same construction with the initial σ-algebra is sufficient to build a monad. To see this, consider yet another similar functor, an endofunctor J on $\mathbb{F}Mes$ defined by

$$\begin{aligned} &J(\Omega,\mathcal{A}) = (\Psi,\mathcal{C}), \text{ where} \\ &\quad \mathcal{C} = \sigma\Big(\bigcup_{A\in\mathcal{A}} ev_A^{-1}(\mathcal{J}[0,1])\Big) \\ &\text{and for } f:(\Omega,\mathcal{A}) \longrightarrow (\Omega',\mathcal{A}') \\ &\quad Jf:\Psi \longrightarrow \Psi' \text{ defined as before by} \\ &\quad\quad Jf\,P(A') = P(f^{-1}(A')) \text{ for } A' \in \mathcal{A}'. \end{aligned}$$

All elements of $\mathcal{C}$ are also elements of $\mathcal{B}$, the algebra in $G(\Omega,\mathcal{A})$, but not vice versa, because the inclusion $\mathcal{J}[0,1] \subset \mathcal{B}[0,1]$ is strict. $\mathcal{C}$ is a rougher algebra than $\mathcal{B}$ which "sees more sets". There is nothing new in the proof that J is a functor, it follows analogously to the proof for G, with the small substitution that, in the proof of finite measurability presented on page 144, the Borel sets (C and D in that formulation) now are sets in $\mathcal{J}[0,1]$.

The definition of the transformation η is analogous to that for G as well. For each object $(\Omega,\mathcal{A})$, the function $\eta_{(\Omega,\mathcal{A})}$ associates to $\omega \in \Omega$ the concentrated probability in ω, defined on $\mathcal{A}$, so that $\eta_{(\Omega,\mathcal{A})}\omega(A) = \mathbb{1}_A(\omega)$. The finite $\mathcal{C}$-$\mathcal{J}[0,1]$-measurability also follows just as in formula (6.10), where again the general Borel-set is substituted as above. Obviously, naturality is also proven like for G, that is, like for F, except that only $A' \in \mathcal{A}'$ are of interest, in equation (6.2), Section 6.2.2.

The transformation μ deserves a closer look. For an object $(\Omega,\mathcal{A})$, the set Ψ is the same as for G, the set of all finitely additive probabilities on $(\Omega,\mathcal{A})$. However, since the algebra on Ψ is different, the set of "finitely additive probabilities over finitely additive probabilities" is not the same as before. $J^2(\Omega,\mathcal{A}) = (\tilde{\Psi}^2,\mathcal{C}^2)$ with

$$\begin{aligned} &\tilde{\Psi}^2 = \{Q|Q \text{ is a finitely additive probability on } (\Psi,\mathcal{C})\} \text{ and} \\ &\mathcal{C}^2 = \sigma\Big(\bigcup_{B\in\mathcal{C}} ev_B^{-1}(\mathcal{J}[0,1])\Big) \qquad \text{where for each } B \in \mathcal{C}, \\ &\quad ev_B \text{ is the evaluation map } ev_B : \tilde{\Psi}^2 \longrightarrow [0,1],\ ev_B(Q) = Q(B). \end{aligned}$$

Now, for an object $(\Omega,\mathcal{A})$, the function $\mu_{(\Omega,\mathcal{A})} : \tilde{\Psi}^2 \longrightarrow \Psi$ is defined as the well-known integral over the evaluation map for each $A \in \mathcal{A}$:

$$\mu_{(\Omega,\mathcal{A})}Q(A) = \int ev_A dQ. \tag{6.14}$$

While rougher than $\mathcal{B}$, the algebra $\mathcal{C}$ still contains all inverse images of $\mathcal{J}[0,1]$-sets under the ev_A, wherefore with Lemma 6.4.4, the integral is well-defined. The definition also delivers a finitely additive probability, shown as for the first functor, F. The finite measurability of the functions $\mu_{(\Omega,\mathcal{A})}$ is the interesting point: for G, we applied "measure theory standards", making use of the fact that $\mathcal{B}[0,1]$ is a σ-algebra. This was not used before that point. That it is sufficient to consider sets in the generator, that is, the application of Lemma 3.1.8 and Remark 3.1.9, does not presuppose any structure on the generator. Integrability was shown to follow from $\mathcal{A}$-$\mathcal{J}[0,1]$-finite measurability in Lemma 6.4.4. The requirement of $\mathcal{A}$-$\mathcal{B}[0,1]$-finite measurability would be stronger because condition (6.9) is stated for all sets in the second algebra, but it is not necessary for the desired integrablity. In fact, in the monad proof for G, we used the fact that $\mathcal{J}[0,1] \subset \mathcal{B}[0,1]$ to apply the thus stated lemma.

In the proof that for G the transformation μ is finitely measurable, "measure theory standards" were used for convenience. Limits of measurable functions are measurable. For J, analogously to the proof for G, we arrive at the task of showing that $g^{-1}(C)$ is a set in $\mathcal{C}^2$ for any $C \in \mathcal{J}[0,1]$ (after formula (6.12)). This g is an integral, $g(Q) = \int ev_A dQ$ and we have a sequence g_n converging uniformly to g. Moreover, each g_n is of the form

$g_n(Q) = \sum_{i=1}^{m} c_i \cdot ev_{B_i}(Q)$ for some $m \in \mathbb{N}$ and $c_i \in [0,1]$ and sets $B_i \in \mathcal{C}^2$. We know that the ev_{B_i} are $\mathcal{C}^2$-$\mathcal{J}[0,1]$-finitely measurable by definition of $\mathcal{C}^2$. Since with $\mathcal{B}$ instead of $\mathcal{C}$ this finite measurability was countable measurability, or measurability in the usual sense, the measurability of sums and limits followed.

Here, we are in an intermediate case between finite and countable measurability. The algebra on the source of the function is a σ-algebra, the one on the target is not. While this does not allow to simply use the results from measure theory, it allows to mimick their proofs, because the σ-algebra on the source is actually the important one.

We need to show that

Lemma 6.5.4. *Given a (countably!) measurable space $(\Omega, \mathcal{F})$ and two finitely measurable functions $f, g : (\Omega, \mathcal{F}) \longrightarrow ([0,1], \mathcal{J}[0,1])$, their sum is $\mathcal{F}$-$\mathcal{J}[0,2]$-finitely measurable.*

Proof. We need to show that the inverse image under $f+g$ of any set in $\mathcal{J}[0,2]$ is an element of $\mathcal{F}$. Considering the generator $\{[0,a],[0,a)|a \in [0,2]\}$ of $\mathcal{J}[0,2]$, it suffices to show this for intervals of this form. That is, one needs to show that

$$(f+g)^{-1}([0,a]) \in \mathcal{F} \text{ and } (f+g)^{-1}([0,a)) \in \mathcal{F}.$$

Consider the second statement.

$$\begin{aligned}(f+g)^{-1}([0,a)) &= \{\omega \in \Omega | (f+g)(\omega) < a\} = \{\omega \in \Omega | f(\omega) + g(\omega) < a\} \\ &= \{\omega \in \Omega | f(\omega) < a - g(\omega)\}\end{aligned}$$

Since the rational numbers are dense in the real numbers (in any interval), for any $x < y \in [0,2]$, one finds an $r \in [0,2] \cap \mathbb{Q}$ with $x < r < y$. Therefore,

$$\begin{aligned}(f+g)^{-1}([0,a)) &= \{\omega \in \Omega | \exists r \in [0,2] \cap \mathbb{Q} : f(\omega) < r \text{ and } r < a - g(\omega)\} \\ &= \bigcup_{r \in [0,2] \cap \mathbb{Q}} \{\omega \in \Omega | f(\omega) < r\} \cap \{\omega \in \Omega | g(\omega) < a - r\} \\ &= \bigcup_{r \in [0,2] \cap \mathbb{Q}} f^{-1}([0,r)) \cap g^{-1}([0,a-r))\end{aligned}$$

The two inverse images in the last line are elements of $\mathcal{F}$ for any r under consideration. In the case that $a-r$ is negative, the interval reduces to the empty set, whose inverse image is the empty set. Since $\mathcal{F}$ is a σ-algebra, also the countable union in the last line is an element of $\mathcal{F}$, which was to be shown. For the first statement, $(f+g)^{-1}([0,a]) \in \mathcal{F}$, we can make use of complements, because the inverse image is well-behaved:

$$(f+g)^{-1}([0,a]) = (f+g)^{-1}((a,2]^c) = \big((f+g)^{-1}((a,2])\big)^c = \big(\{\omega \in \Omega | (f+g)(\omega) > a\}\big)^c$$

The set $\{\omega \in \Omega | (f+g)(\omega) > a\}$ can be shown to be an element of $\mathcal{F}$ exactly as before, wherefore also $(f+g)^{-1}([0,a]) \in \mathcal{F}$. □

The lemma was stated for $\mathcal{F}$-$\mathcal{J}[0,2]$-finite measurability for simplicity. For the functions considered here, evaluation maps of sets in a finite partition, the sum $g_n(Q) = \sum_{i=1}^{m} c_i \cdot ev_{B_i}(Q)$ takes values in the interval $[0,1]$ wherefore we can conclude $\mathcal{F}$-$\mathcal{J}[0,1]$-finite measurability for the functions g_n.

Also their limit is finitely measurable:

Lemma 6.5.5. *Given a (countably!) measurable space $(\Omega, \mathcal{F})$ and finitely measurable functions $g_n : (\Omega, \mathcal{F}) \longrightarrow ([0,1], \mathcal{J}[0,1])$ with a pointwise limit $g : \Omega \longrightarrow [0,1]$, the function g is $\mathcal{F}$-$\mathcal{J}[0,1]$-finitely measurable.*

Proof. We have to show that $g^{-1}((a,b)) \in \mathcal{F}$ for any interval $(a,b) \in [0,1]$. That $g_n^{-1}((a,b)) \in \mathcal{F}$ for any interval $(a,b) \in [0,1]$ is guaranteed by assuming finite measurability of these functions. Now, $g^{-1}((a,b))$ can be written as

$$\bigcup_{n=1}^{\infty} \bigcup_{K=1}^{\infty} \bigcap_{k=K}^{\infty} g_k^{-1}((a+\frac{1}{n}, b-\frac{1}{n})) \tag{6.15}$$

as will be justified below. Then, since all $g_k^{-1}((a+\frac{1}{n}, b-\frac{1}{n})) \in \mathcal{F}$ by assumption, and $\mathcal{F}$ is a σ-algebra, the claim follows. The form (6.15) of $g^{-1}((a,b))$ is correct because $\omega \in g^{-1}((a,b))$ means that $g(\omega) \in (a,b)$. Using the pointwise convergence, this is the case if and only if there are $n, K \in \mathbb{N}$ such that $g_k(\omega) \in (a+\frac{1}{n}, b-\frac{1}{n})$ for all $k \geq K$. This means there are $n, K \in \mathbb{N}$ with $\omega \in g_k^{-1}\big(a+\frac{1}{n}, b-\frac{1}{n}\big)$ for all $k \geq K$, which can be written as $\omega \in \bigcap_{k=K}^{\infty} g_k^{-1}((a+\frac{1}{n}, b-\frac{1}{n}))$ for the n, K considered. Therefore $\omega \in g^{-1}((a,b))$ means

$$\omega \in \bigcup_{n=1}^{\infty} \bigcup_{K=1}^{\infty} \bigcap_{k=K}^{\infty} g_k^{-1}((a+\frac{1}{n}, b-\frac{1}{n})) \text{ and vice versa.} \qquad \square$$

In conclusion, together with the ev_{B_i}, also the functions g_n and g are $\mathcal{C}^2$-$\mathcal{J}[0,1]$-finitely measurable, so that J just like G is a monad with the corresponding natural transformations because further in the proof of naturality of μ and of the monad properties themselves nothing changes.

The second question must be left open. Using the algebra on Ψ defined by

$$\mathcal{D} = \mathfrak{a}\Big(\bigcup_{A\in\mathcal{A}} ev_A^{-1}(\mathcal{J}[0,1])\Big),$$

one can still prove functoriality for a map K that associates $(\Psi, \mathcal{D})$ to $(\Omega, \mathcal{A})$, and that maps functions as $F, \ldots, J$ did. The analogous η and μ can be defined. With this $\mathcal{D}$, each evaluation map is an arrow in $\mathbb{F}Mes$ from $(\Psi, \mathcal{D})$ to the object $([0,1], \mathcal{J}[0,1])$, because for any set $J \in \mathcal{J}[0,1]$, by definition $ev_A^{-1}(J) \in \mathcal{D}$. Still, the evaluation maps for all sets in $\mathcal{A}$ are integrable with respect to any finitely additive probability that one can define on $\mathcal{D}$, so, $\mu_{(\Omega,\mathcal{A})}$ is well-defined. But is it an arrow of the category? While finite measurability here shows its weaknesses, a counterexample for which one finds a $\mu_{(\Omega,\mathcal{A})}$ that is not finitely measurable, has also not been identified (yet?).

More generally speaking, the question whether the sum of two finitely measurable functions (to $\mathbb{R}$ equipped with the algebra generated by the intervals) is open. Together with the fact that pointwise limits of finitely measurable functions are not necessarily themselves finitely measurable, it seems difficult if not impossible to show finite measurability for integrals, which contain a limiting operation (supremum and infimum) in the construction.

6.6 Monad on $\mathbb{F}un$ with arbitrary algebras?

To generalize the monad of finitely additive probabilities on powersets but remain on the category $\mathbb{F}un$ and avoid finite measurability, one could consider a construction in which the algebra that the finitely additive probability is defined on is not fixed with the set in the object, but comes with the finitely additive probability itself. This approach produced some intermediate results, but a monad was not obtained for now.

Algebra functor

As a first step, here one needs to map algebras from one set to another. A functor for algebras of subsets, denoted S, can be constructed as follows.

Given a set Ω, the set $S\,\Omega$ contains all algebras of subsets of Ω. A function $f : \Omega \longrightarrow \Omega'$ is lifted to $S\,f : S\,\Omega \longrightarrow S\,\Omega'$ via the definition: if $\mathcal{A} \in S\,\Omega$ then $S\,f\mathcal{A} = \mathcal{B}$, where $\mathcal{B} \subseteq \mathfrak{P}(\Omega')$, and $B \subseteq \Omega'$ is in $\mathcal{B}$ if $f^{-1}(B) \in \mathcal{A}$.

This defines an algebra on Ω'. First, $\mathcal{B} \in S\,\Omega'$ because Ω' itself is always included in $\mathcal{B}$; its inverse image is Ω, which is in $\mathcal{A}$. Second, the inverse image is well-behaved: if $B \in \mathcal{B}$, then $f^{-1}(B) \in \mathcal{A}$, and since this is an algebra $(f^{-1}(B))^c \in \mathcal{A}$, thus, $f^{-1}(B^c) = (f^{-1}(B))^c \in \mathcal{A}$ which means that $B^c \in \mathcal{B}$. Similarly, $\mathcal{B}$ is closed under intersection by mapping everything backwards. Thus, $\mathcal{B}$ is an algebra of subsets of Ω'.

For functoriality, identities and composition have to be preserved. For identities this is obvious, because the inverse image of the identity is the identity. For composition, consider f as above and $g : \Omega' \longrightarrow \Omega''$. We need to show $S\,g \circ S\,f = S\,(g \circ f)$. For the left hand side, define $\mathcal{C}$, algebra on Ω'' by taking all those $C \subseteq \Omega''$ that satisfy $g^{-1}(C) \in \mathcal{B}$ where $\mathcal{B}$ is constructed as above. Then $C \in \mathcal{C}$ means $g^{-1}(C) = B \in \mathcal{B}$ and by definition of $\mathcal{B}$, $f^{-1}(B) \in \mathcal{A}$. Together, $f^{-1}(g^{-1}(C)) \in \mathcal{A}$. But,

$$\begin{aligned} f^{-1}(g^{-1}(C)) &= f^{-1}\big(\{\omega' \in \Omega' | g(\omega') \in C\}\big) \\ &= \{\omega \in \Omega | f(\omega) \in \{\omega' \in \Omega' | g(\omega') \in C\}\} \\ &= \{\omega \in \Omega | f(\omega) = \omega' \text{ and } g(\omega') \in C\} \end{aligned}$$

For the right hand side, define $\mathcal{C}'$ by the above construction for $g \circ f$, that is, $C \in \mathcal{C}'$ if $(g \circ f)^{-1}(C) \in \mathcal{A}$.

$$\begin{aligned} (g \circ f)^{-1}(C) &= \{\omega \in \Omega | (g \circ f)(\omega) \in C\} \\ &= \{\omega \in \Omega | f(\omega) = \omega' \text{ and } g(\omega') \in C\} \end{aligned}$$

Thus, $\mathcal{C} = \mathcal{C}'$, meaning that composition is preserved.

Let us provide an intuition about a mapped algebra, considering an algebra $\mathcal{A}$ on Ω and a function f as above. In the "worst case", $\mathcal{B} = S\,f(\mathcal{A})$ is the trivial algebra $\{\emptyset, \Omega'\}$. Let $Im(f)$ be the image of Ω under f in Ω', that is, $Im(f) = \{f(\omega) | \omega \in \Omega\}$. The mapped algebra $\mathcal{B}$ automatically contains the powerset of $\Omega' \backslash Im(f)$, because the inverse image of any set contained in the complement of the image is the empty set, which is contained in $\mathcal{A}$.

If $\mathcal{A}$ is finitely generated, one can consider the finitely many constituents of the generator, say $\{C_1, \dots, C_m\}$. Applying f to these may not preserve their empty intersections: in general, $f(C_i) \cap f(C_j) \neq \emptyset$ for $i \neq j$. Then, all that can be said is that the inverse image of the union of the constituents is in $\mathcal{A}$. In $\mathcal{B}$, constituents whose images under f overlap cannot be distinguished any more. However, when the constituents can be sorted into groups with non-overlapping images under f, each group gives rise to a set in $\mathcal{B}$. In this case, one can imagine the mapped algebra as of the following form: if there are k sets stemming from such a group of constituents in $\mathcal{A}$, take $k + 1$ copies of $\mathfrak{P}(\Omega' \backslash Im(f))$. Keep one copy as it is, and to each other copy associate one of the k sets. To get the sets in the algebra, join the constituent-group-set to all sets in the copy, build the "cartesian union" as it were.

In comparison with a finitely measurable function $f : (\Omega, \mathcal{A}) \longrightarrow (\Omega', \mathcal{A}')$, the algebra $S\,f(\mathcal{A})$ is not smaller than $\mathcal{A}'$. The mapped algebra contains the given one because by finite measurability of f, one has $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{A}'$, but $\mathcal{A}'$ need not contain all sets in Ω' for which this holds. For example, it need not contain the powerset of the complement of $Im(f)$.

Finitely additive probability functor

Given the algebra functor, one can define a functor of finitely additive probabilities that generalizes F. The category on which it is defined is $\mathbb{F}un$. To emphasize the underlying set of the algeba that a finitely additive probability is defined on, we will speak of "a finitely additive probability over Ω", for example. Our new functor, denoted L, associates to a set all finitely additive probabilities on all algebras of subsets. To make the algebra explicit in the notation, the functor produces pairs: finitely additive probabilities and algebras, which are their domains of definition, that is

$$L\,\Omega = \{(P, \mathcal{A}) | \mathcal{A} \in S\,\Omega \text{ and } P : \mathcal{A} \longrightarrow [0,1] \text{ is a finitely additive probability on } (\Omega, \mathcal{A})\}.$$

L maps functions by the usual procedure, taking as algebra on the target set the mapped algebra constructed previously. That is, $L\,f : L\,\Omega \longrightarrow L\,\Omega'$ is defined, for $f : \Omega \longrightarrow \Omega'$, by

$$\begin{aligned} L\,f(P, \mathcal{A}) &= (P', \mathcal{A}') \text{ with } \mathcal{A}' = S\,f\mathcal{A}, \text{ and} \\ P'(A') &= P(f^{-1}(A')) \text{ for all } A' \in \mathcal{A}'. \end{aligned} \qquad (6.16)$$

That this is a functor can again be shown exactly as for F, substituting the powersets by the respective algebras. No measurability questions arise.

A monad (L, η, μ) on the category $\mathbb{F}un$ was not obtained. The unsuccessful attempts at constructing it will not be detailed here. The problem, in short, is rather μ than η. For the latter, the usual concentrated probability can be defined on the powerset (for lack of another algebra that would be preferrable). μ, however, is not as obvious. For an object Ω of $\mathbb{F}un$, the object $L^2\,\Omega$ is the set of all finitely additive probabilities over $L\,\Omega$, and these finitely additive probabilities may be defined on any algebra of subsets of $L\,\Omega$. When μ_Ω is applied to an element $(Q, \mathcal{B})$ of $L^2\,\Omega$, the result must be a finitely additive probability over Ω. It is not clear which algebra this finitely additive probability should be defined on. For the simple case of a discrete Q that gives positive mass to finitely many values, the intersection of the algebras of the finitely additive probabilities that get these positive weights suggests itself. It is however unclear how to extend such an idea to the general case, especially since a finitely additive probability with infinitely many possible values (no matter whether countably or uncountably many) is not determined by the probabilities assigned to these values. Recall the uniform probability over $\mathbb{N}$ from Section 4.4. Considering intersections of many algebras, one might all too often end up with the trivial algebra $\{\emptyset, \Omega\}$.

Constructing the algebra to define $\mu_\Omega Q$ upon via integrability of evaluation maps also failed: considering $\mu_\Omega Q(A)$ for all $A \in \mathfrak{P}(\Omega)$ to be $\int ev_A dQ$ if the evaluation map is integrable and undefined otherwise, does not work, because the set of the $A \subseteq \Omega$ for which this integral is well-defined is not necessarily an algebra.

Even including an undefined element $\perp$ directly in the functor did not solve the problem. The undefined element "destroys" information, so that diagrams do not necessarily commute depending on whether information is destroyed at an earlier or later point in the composition of several operations. All in all, the trouble is immediately suggested by a missing "natural" candidate for the algebra on which to define a finitely additive probability resulting from the application of μ. Any construction that assumes extra information is bound not to yield a *natural* transformation.

6.7 Coherent assignments and monads

As an outlook, let us consider also the further generalization of finitely additive probabilities, coherent probability assignments, in relation to functors and monads. A very similar construction to the algebra functor allows to show that coherent probability assignments have functorial structure. A functor of coherent probability assignments T can be constructed on the category $\mathbb{F}un$ as follows.

To an object of $\mathbb{F}un$, that is, a set Ω, T associates all coherent probability assignments on sets $\mathcal{E}$ of subsets $E \subseteq \Omega$ of Ω. For notational convenience, we will denote coherent assignments here as total functions on the powerset of Ω which take values in $[0,1] \cup \perp$, with the interpretation that any set which is not assigned a probability value by the assignment is mapped to the undefined element $\perp$. That is, a coherent probability assignment is a function

$$\alpha : \mathfrak{P}(\Omega) \longrightarrow [0,1] \cup \perp$$

such that the restriction of α to the set on which it takes values in $[0,1]$, that is $\mathcal{E} = \alpha^{-1}([0,1])$, can be extended to a finitely additive probability on $\mathfrak{a}(\mathcal{E})$, the generated algebra. As before, we will also take the domain of definition into the notation. Then, the functor associates to a set Ω

$$\begin{aligned} T\,\Omega = \{(\alpha, \mathcal{E}) | \alpha : \mathfrak{P}(\Omega) \longrightarrow [0,1] \cup \perp,\, \mathcal{E} = \alpha^{-1}([0,1]) \\ &\text{and } \alpha \text{ is a coherent probability assignment, that is,} \\ &\alpha\,|_{\mathcal{E}} \text{ can be extended to a finitely additive probability} \\ &P : \mathfrak{a}(\mathcal{E}) \longrightarrow [0,1]\}. \end{aligned}$$

To an arrow of $\mathbb{F}un$, that is, a function $f : \Omega \longrightarrow \Omega'$, the functor T associates the assignment obtained by checking for all sets in $\mathfrak{P}(\Omega')$ whether their inverse image is assigned a probability by α.

$$T\,f\,\alpha(A') = \alpha(f^{-1}(A')) \text{ for all } A' \in \mathfrak{P}(\Omega'). \tag{6.17}$$

The so defined $T\,f\,\alpha$ is a coherent probability assignment on Ω': consider any finitely additive probability P on $\mathfrak{a}(\mathcal{E})$ that α extends to. Mapping this finitely additive probability with the functor L constructed above produces a finitely additive probability $L\,f\,(P, \mathfrak{a}(\mathcal{E}))$ over Ω'. The here obtained $T\,f\,\alpha$ extends to the function in the pair $L\,f\,(P, \mathfrak{a}(\mathcal{E}))$. This can be seen by rewriting also finitely additive probabilities with associated algebras as functions from $\mathfrak{P}(\Omega)$ to $[0,1] \cup \perp$ as done with coherent probability assignments. That is, in a redundant but here useful notation, a finitely additive probability is a pair $(P, \mathcal{A})$ such that $P : \mathfrak{P}(\Omega) \longrightarrow [0,1] \cup \perp$ with $P^{-1}([0,1]) = \mathcal{A}$ and $P\,|_{\mathcal{A}}$ is a finitely additive probability according to definition 4.4.1. Since the information about the algebra is given by P, we will write the application of $L\,f$ to P only. This is then given precisely by the adapted formula (6.17)

$$L\,f\,P(A') = P(f^{-1}(A')) \text{ for all } A' \in \mathfrak{P}(\Omega'). \tag{6.18}$$

Returning to the finitely additive probability that extends α, those sets A' whose inverse image is in $\mathfrak{a}(\mathcal{E})$ are assigned the probability of this inverse image as before, and those for which this is not the case are assigned $\perp$. Since these were not considered at all before, this yields the same finitely additive probability on $S\,f\mathcal{A}$ as the definition in (6.16). Further, since the definitions in (6.17) and (6.18) coincide, and $P\,|_{\mathcal{E}} = \alpha\,|_{\mathcal{E}}$, also $L\,f\,P\,|_{\mathcal{F}} = T\,f\,\alpha\,|_{\mathcal{F}}$, where $\mathcal{F}$ is the set of all those sets in $\mathfrak{P}(\Omega')$ whose inverse image under f is in $\mathcal{E}$, that is, $\mathcal{F} = \{A' \in \mathfrak{P}(\Omega') | f^{-1}(A') \in \mathcal{E}\}$.

T is a functor because it obviously preserves identities, and as the inverse image is well-behaved it also preseves composition: for $f : \Omega \longrightarrow \Omega'$ and $g : \Omega' \longrightarrow \Omega''$ and $A'' \subseteq \Omega''$,

$$\begin{aligned} T\,(g \circ f)\,\alpha(A'') &= \alpha((g \circ f)^{-1}(A'')) = \alpha(f^{-1}(g^{-1}(A''))) \\ &= (T\,f\,\alpha)(g^{-1}(A'')) = (T\,g \circ T\,f)\,\alpha(A''). \end{aligned}$$

However, the question whether it yields a monad is again an open question. A coherent assignment could be viewed as the collection of all the finitely additive probabilities on the generated algebra $\mathfrak{a}(\mathcal{E})$ that it extends to. Instead of a function to $[0,1] \cup \perp$, a coherent probability assignment could be written as a function with interval values: assign to each set in $\mathfrak{a}(\mathcal{E})$ the interval of values that any finitely additive probability in the collection assigns to it. This is an interval because convex combinations of coherent assignments are coherent assignments. A theory that could be useful here is presented by Weichselberger [2001], where probability is generalized to a function that takes interval values. However, this has not been further pursued. Given that for the functor L a monad has not been found, it seems unlikely to find a monad for T. The close connection to L exposed above suggests that the problems encountered in the attempt to construct such a monad for L can be expected here as well. Also, identifying a coherent probability assignment with the set of its extensions would not always allow to re-extract the original assignment from this set alone. In the case where more sets than the elements of $\mathcal{E}$ (which were assigned probabilities by the original α) have a unique value, one would need the additional information which sets were in $\mathcal{E}$. This may not be a major problem, but the extra-information to be considered could render the construction of a monad more complicated.

6.8 Conclusions and outlook

This chapter has introduced several monads of finitely additive probabilities. First, a monad on the category $\mathbb{F}un$ was presented that considers finitely additive probabilities defined on the powerset of a given set. In this case, no measurability questions arise, which yields a comparatively easy proof.

Since probability measures defined on the powerset of infinite sets can be discrete probability measures only, and this would exclude many interesting cases, a monad of finitely additive probabilities which in a certain sense extends Giry's monad of probability measures was presented next. It lives on a category that generalizes the category of measurable spaces and functions $\mathbb{M}es$, denoted $\mathbb{F}Mes$. The objects of this category are sets with algebras, its arrows are functions satisfying an equivalent of measurability in terms of algebras, introduced as finite measurability for the construction of this monad. The concept of finite measurability helped to prove that there exists a monad of finitely additive probabilities defined on fixed algebras, but in itself it left some questions open.

In order to extend Giry's monad, the construction of this second monad leaned on her construction, for example, in the choice of the algebra to accompany the set of finitely additive probabilities on a given set. In a first attempt, we simply copied the σ-algebra used by Giry. If the criterion of extending her monad is dropped, the assumptions made for this algebra can be slightly weakened, as was shown next. However, it also became obvious that the concept of finite measurability is not powerful enough to carry the construction applied here by itself: a σ-algebra was needed to make a function containing an integral an arrow of the category of finitely measurable sets and functions. Whether a more general monad of finitely additive probabilities on the category $\mathbb{F}Mes$ can be obtained in some other way is an open question.

Generalizing further again, the question whether monads of finitely additive probabilities defined on arbitrary algebras exist on the category $\mathbb{F}un$ was considered. While a functor is rather straightforward, a monad was not obtained here, leaving further questions open. The same holds for coherent probability assignments, since these further generalize finitely additive probabilities defined on arbitrary algebras.

Interesting mathematical questions for further investigation are brought up by returning to the motivation of this chapter, the monadic dynamical systems used in the formal framework of vulnerability. A closer look at the Kleisli-categories of all the monads found could be taken in further work. Giry relates the Kleisli-arrows to Markov processes. Since finitely addivite Markov processes are less well studied than the countably additive counterpart, their study via category theory could be an interesting endeavour. Markov processe are also closely related to conditional expectation, which was not treated in this work. It would be interesting how the Kleisli-arrows of the monads presented here relate to different concepts of conditional expectation and conditional probability in the finitely additive setting that can be found already in the literature [see, e.g. Coletti and Scozzafava, 2002, and references therein]. However, this topic was considered beyond the scope of this work.

Further results from category theory about probability measures presented by Giry [1981] could also be examined in the finitely additive setting. An "important problem" investigated by her is that of finding a probability measure on the sample space compatible with given transition probabilities of a Markov process. While there is a lot of work on stochastic processes using probability measures, for finitely additive probabilities mainly processes in discrete time seem to be considered in the literature [see, e.g., Dubins and Savage, 1965]. The category theoretical approach might yield some interesting results for continuous time processes. Last but not least, Giry [1981] quotes an unpublished paper by Lawvere from 1962, stating that "most problems in probability and statistics theory can be translated in terms of diagrams of these Kleisli categories." [p. 68] A further exploration of what the combination of finitely additive probability and category theory has to offer seems worthwhile.

Chapter 7

Conclusions and outlook

Many of the conceptual questions are, unfortunately, inexhaustible if one wishes to examine them thoroughly; and the worst thing is that, often, they are also rather boring unless one has a special interest in them. [de Finetti, 1974, p. 14]

This work is subdivided into two parts, related in that the first part provides the motivation for the second one. In this first part, the method of formalization was discussed and applied with a special interest in the two concepts probability and vulnerability. While probability is a concept with a history of formalizations and thus serves as an example, the application of the method formalization to vulnerability is new. It yields the formal framework of vulnerability which motivated this dissertation.

Formalization was introduced as translation into mathematics. Therefore, first of all, mathematics had to be awarded the status of a language. The characteristics of this language, differentiating it from natural languages, are precision and consistency in the definition of concepts, and the abstract and therefore general nature of concepts used and of solutions found. The exact nature of mathematical definitions implies that mathematical concepts do not carry implicit connotations which differ with the speaker's background, as is the case for many concepts in the vulnerability terminology. It therefore suggests itself as a lingua franca for communication between different scientific disciplines, and can be hoped to enhance clarity in the interdisciplinary field of vulnerability research. The symbolic notation used in mathematics, which constitutes another difference to natural languages, entails that operations can be performed on the symbols. Fundamental for the use of the computer, mathematics here has an advantage over natural languages. However, the concise notation makes communication difficult for those who are not trained to use it. Mathematics is not generally "spoken" by the target audience of the formal framework.

Hoping to nevertheless reap benefits of formalization for the confused terminology of vulnerability, the mathematical models of concepts were re-translated into diagrams using natural language words, disguising to some extent the underlying mathematics. A mathematical model of vulnerability as used in ordinary language and a model of the scientific concept constitute the formal framework of vulnerability. Even if rather simple, in its generality the first model was able to embody the common structure of many theoretical definitions of the scientific concept. The refined mathematical model for the scientific concept had a similar effect for assessment approaches and interpretations of vulnerability. Several distinctions of interpretations made in the conceptual literature were ascribed to one pair of assessment approaches, each described by one part of the mathematical model of the scientific concept. Approaches to assessing risk of natural hazards, interpreted as instances of the formal framework, were shown to have a similar structure to vulnerability assessments while terms are used differently. The revealed non-injectivity of any mapping between terms used and elements of assessments explained the terminology confusion both within and between the two research communities.

Relating the levels of meaning and measurement of 'possible future harm', a mismatch between theoretical and operational definitions of vulnerability was discussed. The missing connection between theoretical defintions and interpretations of the concept motivates a suggestion for work on the concept: decrease the emphasis on theoretical definitions, as it were, shift the focus away from the trees in order to see the forest.

The concepts vulnerability and risk have in common an uncertain future evolution of the entity under consideration. While in risk assessments probability is generally used to describe uncertainty about future events, this is much less the case in vulnerability assessments. Two reasons for this were discussed. First, the focus on abilities of people to act in one vulnerability assessment approach entails a disinterest in the description of the uncertain future, and hence probability is not needed. Second, the SRES scenarios, constituting the foundation of the description of the uncertain future in many assessments, are explicitly provided without probabilities. The view of probability motivating this choice, roughly speaking, reduces the concept from a description of uncertainty to a supposed prediction tool. The debate about the use of probability reveals that this concept still is not clear.

A chapter was dedicated to probability to discuss this concept as an example of formalization. Here, not only different interpretations exist, but also different mathematical models compete. The less general standard model based on Kolmogorov's axioms is perceived as mathematically more convenient: being based on measure theory, it can use many of its results. On the other hand, it excludes examples which from a probabilistic point of view seem no less intuitive than others, as was seen for uniform distributions, which can be defined on finite and uncountably infinite sample spaces, but not on countably infinite ones. Such examples can be included when the axiom of countable additivity is disregarded, as is the case for finitely additive probabilities. These are again generalized by coherent probability assignments, for which the domain of definition need not be an algebra.

Formalization was seen to be no miracle cure for confused concepts: interpretations of probability are not unambiguously linked to a mathematical model of it. Also, mathematical models are related to the concept, its meaning and its measurement in rather different ways, explained from the intention pursued with a formalization. A mathematical model may be motivated mathematically, as Kolmogorov's axioms were. They build a foundation to turn a previously confused half-mathematical research area into a full-fledged mathematical discipline. The concept probability is taken as a primitive, implying that neither its meaning nor how to measure it is explained by the mathematical model. A mathematical model may on the other hand be motivated by the attempt at claryfying concepts, as in de Finetti's approach, where mathematics is perceived as a tool, not a goal in itself. The operational definition of (subjective) probability provided allows to measure probabilities and illustrates the meaning of the concept.

Discussing mathematics as a language, beneficial for concept clarification in interdisciplinary contexts, illustrating with the example probability, applying the translation method to the example vulnerability, reaping some benefits by sorting out definitions, interpretations and relations with the concept risk and the concept probability, and finally reducing the mathematical appearance of the results for communication with non-mathemematicians, the first part of this work is a work *about* mathematics.

The second part is a work *in* mathematics in a more restricted sense. It combines elements from the two fields of probabilitity theory and category theory, motivated by the first part. Where probability is used in climate change research, often a frequentist interpretation is not warranted. Considering a subjective interpretation instead, probabilities are for example elicited from experts. Here, rather than using probability measures as in the standard model, the more general mathematical model of coherent probability assignments suggests itself. The idea was therefore to make more general models of probability, such as finitely additive probability and coherent probability assignments, available for use in the formal framework of vulnerability.

The generalized uncertainty description in the framework uses category theoretical concepts. A functor describes a set of possibilities to which further information may be attached, and allows the application of functions to the possibilities themselves, maintaining the extra information. It was shown that both finitely additive probabilities and coherent probability assignments form functors.

In the computational version of the framework, monadic dynamical systems are used to compute the future description in a very general fashion. These allow iteration involving functorial descriptions of uncertainty when a monad is provided. Therefore, several monads of finitely additive probabilities were identified. These are a monad of finitely additive probabilities defined on powersets, an extension of the monad of probability measures proposed by Giry [1981] and small generalizations of this. For the second class of monads, an auxiliary concept to deal with functions defined on sets which are equipped with algebras had to be introduced. This 'finite measurability' was useful enough to support proofs of some monads, but was seen to have serious deficiencies when compared to measurability in the usual sense. This is probably why it is not encoutered in the measure theory literature. Coherent probability assignments unfortunately where not proven to yield a monad, which leaves scope for further work at the intersection between probability and category theory.

As an outlook, let us state some interesting questions raised in this work or arising from it. This concerns both conceptual questions and mathematical questions which could not also be treated here. For example, the question whether "vulnerability does not exist", in the sense that one can only consider the subjective vulnerability as viewed by a fixed reference person, was mentioned. While subjectivity of vulnerability measurements is discussed as a drawback of these, inherent subjectivity of the concept could be investigated. A mathematical further question that was mentioned is whether monadicity allows to derive results on the existence of a finitely additive probability on the sample space from given finitely additive transition probabilities, especially considering a continuous time system.

This work has been a theoretical work. Practical questions from the climate change decision making context would provide further objects of study on the borders between vulnerability and probability. A topic of current interest is the issue of learning. Adaptive management and social learning are frameworks discussed in the context of adaptation decision making. In a probabilistic context, updating one's views about the future with new information is studied, for example, under the heading of Bayesian learning. A generalization of coherent probability assignments, coherent conditional assignments as studied by Coletti and Scozzafava [2002], allow an updating procedure for probabilities that is more generally applicable than Bayes' theorem which is traditionally used. Integrating these probabilistic ideas into the formal framework of vulnerability might enable it to describe further approaches in climate change decision making, possibly opening up further areas where mathematics can provide some clarity.

Bibliography

Adams, J., 1995. *Risk*. UCL Press Limited, London.

Adger, W. N., 1999. Social Vulnerability to Climate Change and Extremes in Coastal Vietnam. *World Development*, 27(2), pp. 249–269.

Adger, W. N., 2006. Vulnerability. *Global Environmental Change*, 16, pp. 268–281.

Aldous, D. and Diaconis, P., 1986. Shuffling cards and stopping times. *American Mathematical Monthly*, 93(5), pp. 333–347.

Banach, S., 1930. Über additive Massfunktionen in abstrakten Mengen. *Fundamenta Mathematicae*, 15, pp. 97–101.

Banach, S. and Kuratowski, C., 1929. Sur une généralisation du problème de la mesure. *Fundamenta Mathematicae*, 14, pp. 127–131.

Barredo, J. I. and de Roo, A., 2007. Map of high flood risk areas in Europe. Institute for Environment and Sustainability, Joint Research Centre of the European Commission, LB-78-07-414-EN-C, Ispra.

Bernoulli, J., 2005. On the Law of Large Numbers. Ars Conjectandi. Basileae, Impensis Thurnisiorum, Fratrum, 1713. Pars Quarta tradens Usum & Applicationem Praecedentis Doctrinae in Civilibus, Moralibus & Oeconomicis. Translated into English by Oscar Sheynin.
URL `http://www.sheynin.de/download/bernoulli.pdf`

Beutelspacher, A., 1996. *"In Mathe war ich immer schlecht . . . "*. Vieweg, Braunschweig, Wiesbaden.

Bhaskara Rao, K. and Bhaskara Rao, M., 1983. *Theory of Charges*. Pure and Applied Mathematics. Academic Press, London.

Billingsley, P., 1979. *Probability and Measure*. John Wiley & Sons, New York, Toronto, London.

Bird, R. and de Moor, O., 1997. *Algebra of Programming*. International Series in Computer Science. Prentice Hall, Hemel Hempstead.

Blaikie, P., Cannon, T., Davis, I., and Wisner, B., 1994. *At Risk - natural hazards, people's vulnerability and disasters*. Routledge, London and New York.

Bogachev, V., 2006. *Measure Theory*. Springer, Berlin.

Bohle, H. G., Downing, T. E., and Watts, M. J., 1994. Climate change and social vulnerability. Toward a sociology and geography of food insecurity. *Global Environmental Change*, 4(1), pp. 37–48.

Boldrighini, C., 2009. Doeblin and Kolmogorov: The Mathematisation of Probability in the Thirties. In: M. Emmer (ed.), *Mathematics and Culture VI*, pp. 247–260. Springer, Berlin, Heidelberg.

Borkar, V., Konda, V., and Mitter, S., 2004. On De Finetti coherence and Kolmogorov probability. *Statistics & Probability Letters*, 66, pp. 417–421.

Botts, T., 1969. Probability Theory and the Lebesgue Integral. *Mathematics Magazine*, 42, pp. 105–111.

Brooks, N., 2003. Vulnerability, risk and adaptation: A conceptual framework. *Tyndall Center Working Paper*, 38.

Brooks, N., Adger, W. N., and Kelly, P. M., 2005. The determinants of vulnerability and adaptive capacity at the national level and the implications for adaptation. *Global Environmental Change*, 15, pp. 151–163.

Buckle, P., Marsh, G., and Smale, S., 2001. Assessing Resilience & Vulnerability: Principles, Strategies & Actions. Guidelines (Emergency Management Australia).

Burton, I., Kates, R. W., and White, G. F., 1993. *The Environment as Hazard*. The Guilford Press, New York, London, second edition.

Calvo, C. and Dercon, S., 2005. Measuring Individual Vulnerability. *Department of Economics Discussion Paper Series*, 229.
URL `http://ideas.repec.org/p/oxf/wpaper/229.html`

Cannon, T., 1994. Vulnerability Analysis and the Explanation of 'Natural' Disasters. In: A. Varley (ed.), *Disasters, Development, and Environment*, pp. 13–30. John Wiley & Sons, Chichester.

Carter, T. R., Jones, R. N., Lu, X., Bhadwal, S., Conde, C., Mearns, L. O., ONeill, B. C., Rounsevell, M. D. A., and Zurek, M. B., 2007. New Assessment Methods and the Characterisation of Future Conditions. In: M. Parry, O. Canziani, J. Palutikof, P. van der Linden, and C. Hanson (eds.), *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 133–171. Cambridge University Press, Cambridge.

Chambers, R., 1989. Vulnerability, Coping and Policy - Introduction. *IDS Bulletin - Institute of Development Studies*, 20(2), pp. 1–7.

Coletti, G. and Scozzafava, R., 2002. *Probabilistic Logic in a Coherent Setting*, volume 15 of *Trends in Logic*. Kluwer Academic Publishers, Dordrecht, Boston, London.

Courant, R. and Robbins, H., 1941. *What is Mathematics*. Oxford University Press, London, New York, Tokyo. Ninth Printing 1958.

Crichton, D., 1999. The Risk Triangle. In: J. Ingleton (ed.), *Natural Disaster Management*. Tudor Rose, London.

Cutter, S. L., 1993. *Living with Risk*. Edward Arnold, London.

Cutter, S. L., 1996. Vulnerability to Environmental Hazards. *Progress in Human Geography*, 20(4), pp. 529–539.

Cutter, S. L., Mitchell, J. T., and Scott, M. S., 2000. Revealing the Vulnerability of People and Places: A Case Study of Georgetown County, South Carolina. *Annals of the Association of American Geographers*, 90(4), pp. 713–737.

Davis, I., Haghebaert, B., and Peppiatt, D., 2004. Social Vulnerability & Capacity Analysis (VCA): An Overview. Discussion paper prepared for the ProVention Consortium Workshop at IFRC Geneva on May 25-26.

de Finetti, B., 1930a. A proposito dell'estensione del teorema delle probabilità totali alle classi numerabili. *Rendiconti del R. Istituto Lombardo di Scienze e Lettere*, 63, pp. 901–905.

de Finetti, B., 1930b. Ancora sull'estensione alle classi numerabili del teorema delle probabilità totali. *Rendiconti del R. Istituto Lombardo di Scienze e Lettere*, 63, pp. 1063–1069.

de Finetti, B., 1930c. Sui passaggi al limite nel calcolo delle probabilità. *Rendiconti del R. Istituto Lombardo di Scienze e Lettere*, 63, pp. 155–166.

de Finetti, B. (ed.), 1972. *Probability, Induction and Statistics.* John Wiley & Sons, London, New York, Sydney, Toronto.

de Finetti, B., 1974. *Theory of Probability, Vol. 1.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, London, New York, Sydney, Toronto. English Translation of: Teoria Delle Probabilità, Einaudi, Torino, 1970.

de Finetti, B., 1975. *Theory of Probability, Vol. 2.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, London, New York, Sydney, Toronto. English Translation of: Teoria Delle Probabilità, Einaudi, Torino, 1970.

de Finetti, B., 1981. *Wahrscheinlichkeitstheorie.* R. Oldenbourg Verlag, Wien, München. German Translation of: Teoria Delle Probabilità, Einaudi, Torino, 1970.

de Finetti, B., 1995. The logic of probability. *Philosophical Studies*, 77, pp. 181–190. Translated by R. B. Angell from *La logique de la probabilité*, Actualités Scientifiques et Industrielles, 391, Actes du Congrès International de Philosophie Scientifique, Sorbonne, Paris, 1935, IV. Induction et Probabilité, Herrnann et Cle, Éditeurs, Paris, 1936, pp. 31-39.

de Finetti, B., 2008. *Philosophical Lectures on Probability.* Springer, Dordrecht. Collected, edited and annotated by Alberto Mura.

Dessai, S. and Hulme, M., 2004. Does climate adaptation policy need probabilities? *Climate Policy*, 4, pp. 107–128.

Diaconis, P., 2007a. History of Probability from Greek and Talmudic Sources to Pascal, Fermat and Bernoulli. Lecture Notes Theory of Probability, Lecture 1. Visited Aug. 15, 2008.

Diaconis, P., 2007b. Meaning of Probability. Lecture Notes Theory of Probability, Lecture 6. Visited Aug. 15, 2008.

Diaconis, P., 2007c. Subjective Probability. Lecture Notes Theory of Probability, Lecture 7. Visited Aug. 15, 2008.

Dow, K., 2003. Exploring Differences in Our Common Future(s): the Meaning of Vulnerability to Global Environmental Change. *Geoforum*, 23(3), pp. 117–136.

Dubins, L. E., 1974. On Lebesgue-like extensions of finitely additive measures. *The Annals of Probability*, 2, pp. 456–463.

Dubins, L. E. and Savage, L. J., 1965. *How to Gamble if you Must: Inequalities for Stochastic Processes.* McGraw Hill, New York.

Dudley, U., 1997. Is Mathematics Necessary? *The College Mathematics Journal*, 28(5), pp. 360–364.

Elstrodt, J., 1999. *Maß- und Integrationstheorie*. Springer, Berlin, Heidelberg, New York.

Enzensberger, H. M., 1999. *Zugbrücke außer Betrieb - Die Mathematik im Jenseits der Kultur. Drawbridge Up: Mathematics - A Cultural Anathema*. A K Peters, Natick, Massachusetts.

Eriksen, S. H. and Kelly, P. M., 2007. Developing Credible Vulnerability Indicators for Climate Adaptation Policy Assessment. *Mitigation and Adaptation Strategies for Global Change*, 12(4).

Fishburn, P. C., 1986. The Axioms of Subjective Probability. *Statistical Science*, 1(3), pp. 335–345.

Franklin, J., 2001. *The Science of Conjecture, Evidence and Probability Before Pascal*. John Hopkins Press, Baltimore.

Fréchet, M., 1930a. Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'événements. *Rendiconti del R. Istituto Lombardo di Scienze e Lettere*, 63, pp. 899–900.

Fréchet, M., 1930b. Sur l'extension du théorème des probabilités totales au cas d'une suite infinie d'événements. *Rendiconti del R. Istituto Lombardo di Scienze e Lettere*, 63, pp. 1058–1062.

Füssel, H.-M., 2007. Vulnerability: a generally applicable conceptual framework for climate change research. *Global Environmental Change*, 17(2), pp. 155–167.

Füssel, H.-M. and Klein, R. J. T., 2006. Climate change vulnerability assessments: an evolution of conceptual thinking. *Climatic Change*, 75(3), pp. 301–329.

Galavotti, M. C., 2008. De Finetti's Philosophy of Probability. In: A. Mura (ed.), *Philosophical Lectures on Probability*, pp. xv–xxii. Springer, Dordrecht.

Gale, D. and Shapley, L., 1962. College admission and the stability of marriage. *American Mathematical Monthly*, 69, pp. 9–14.

Gallopín, G. C., 2006. Linkages between vulnerability, resilience, and adaptive capacity. *Global Environmental Change*, 16(3), pp. 293–303.

Gangopadhyay, S. and Rao, B. V., 1997. Some Finitely Additive Probability: Random Walks. *Journal of Theoretical Probability*, 10(3), pp. 643–657.

Geluck, P., 2008. *God save the Cat*. Rue Elise Editions.

Genovese, E., Lugeri, N., Lavalle, C., Barredo, J. I., Bindi, M., and Moriondo, M., 2007. An assessment of weather-related risks in Europe. Preliminary Report at M18: maps of flood and drought risks. *Deliverable to the ADAM project*, DA2.1.

Giry, M., 1981. A categorial approach to probability theory. In: B. Banaschewski (ed.), *Categorical Aspects of Topology and Analysis*, volume 915 of *Lecture Notes in Mathematics*, pp. 68–85. Springer, Berlin.

Glare, P. G. W. (ed.), 1982. *Oxford Latin Dictionary*. Clarendon Press, Oxford.

Good, I., 1977. Book Review: Theory of probability, Volume I, Bruno de Finetti, John Wiley & Sons, New York, 1974, xix + 300 pp., $22.50. Theory of probability, Volume II, Bruno de Finetti, John Wiley & Sons, New York, 1975, xviii + 375 pp., $29.50. *Bulletin of the American Mathematical Society*, 83(1).

Gordon, R. A., 2000. A Convergence Theorem for the Riemann Integral. *Mathematics Magazine*, 73(2), pp. 141–147.

Grübler, A. and Nakićenović, N., 2001. Identifying dangers in an uncertain climate. *Nature*, 412, p. 15.

Hacking, I., 1975. *The Emergence of Probability. A philosophical study of early ideas about probability, induction and statistical inference.* Cambridge University Press, Cambridge.

Hàjek, A., 2007. Interpretations of Probability. Stanford Encyclopedia of Philosophy. URL http://plato.stanford.edu/entries/probability-interpret

Heath, D. and Sudderth, W., 1976. De Finetti's Theorem on Exchangeable Variables. *The American Statistician*, 30(4), pp. 188–189.

Henshaw, J. M., 2006. *Does measurement measure up?: how numbers reveal and conceal the truth.* The John Hopkins University Press, Baltimore, Maryland.

Hinkel, J., 2008. *Transdisciplinary Knowlegde Integration. Cases from Integrated Assessment and Vulnerability Assessment.* Ph.D. thesis, Wageningen University, Wageningen, The Netherlands.

Hinkel, J. and Klein, R. J. T., 2007. Integrating knowledge for assessing coastal vulnerability. In: L. M. Fadden, R. J. Nicholls, and E. Penning-Rowsell (eds.), *Managing Coastal Vulnerability.* Earthscan, London.

Hochrainer, S., Mechler, R., Pflug, G., Linnerooth-Bayer, J., Aaheim, A., and Mestl, H., 2007. Preliminary report on Europes financial and economic vulnerability to meteorological extremes. *Deliverable to the ADAM project*, DA2.3.

Hofmann, M., 2007. Resilience. A formal approach to an ambiguous concept. *FAVAIA Working Paper 5, Potsdam Institute for Climate Impact Research, Potsdam, Germany.* URL http://www.pik-potsdam.de/favaia/pubs/diplomarbeit_mhofmann.pdf

Huygens, C., 1657. *De Ratiociniis in Ludo Aleae.* Den Haag.

Ionescu, C., 2009. *Vulnerability Modelling and Monadic Dynamical Systems.* Ph.D. thesis, Freie Universität Berlin.
URL http://www.diss.fu-berlin.de/diss/receive/FUDISS_thesis_000000008403

Ionescu, C. and Botta, N., 2008. Modeling Versus Formalization. Accepted for publication in the Proceedings of the 98th Dahlem Workshop of Dahlem Conferences: 'Is There a Mathematics of Social Entities?'.

Ionescu, C., Klein, R. J. T., Hinkel, J., Kavi Kumar, K. S., and Klein, R., 2009. Towards a formal framework of vulnerability to climate change. *Environmental Modelling and Assessment*, 14(1), pp. 1–16.

Ionescu, C., Klein, R. J. T., Hinkel, J., Kumar, K. S. K., and Klein, R., 2005. Towards a formal framework of vulnerability to climate change. *NeWater Working Paper 2 and FAVAIA Working Paper 1.*
URL http://www.pik-potsdam.de/favaia/pubs/ionescu_etal_2005c.pdf

Jacod, J. and Protter, P., 2000. *Probability Essentials.* Universitext. Springer, Berlin, Heidelberg.

Janssen, M. A. and Ostrom, E., 2006. Resilience, vulnerability and adaptation: A cross-cutting theme of the International Human Dimensions Programme on Global Environmental Change. *Global Environmental Change*, 16(3), pp. 237–239. Editorial.

Jeffrey, R., 2004. *Subjective Probability. The Real Thing.* Cambridge University Press, Cambridge.

Jones, R. N., 2001. An environmental risk assessment/management framework for climate change impact assessments. *Natural Hazards*, 23(2-3), pp. 197–230.

Justens, D., 2008. *La Mathématique du Chat*. Delagrave Édition, Paris.

Kahneman, D., P., S., and Tversky, A. (eds.), 1982. *Judgment Under Uncertainty. Heuristics and Biases*. Cambridge University Press, Cambridge.

Kates, R. W., 1985. The Interaction of Climate and Society. In: R. W. Kates, J. H. Ausubel, and M. Berberian (eds.), *Climate Impact Assessment: Studies of the Interaction of Climate and Society*, SCOPE Report 27. John Wiley & Sons, Chichester.

Kelly, P. M. and Adger, W. N., 2000. Theory and practice in assessing vulnerability to climate change and facilitating adaptation. *Climatic Change*, 47, pp. 325–352.

Knight, F., 1921. *Risk, uncertainty and profit*. Harper & Row, New York.

Kolmogorov, A. N., 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Erg. Math., Vol. 2, No. 3. Springer, Berlin.

Kolmogorov, A. N., 1956. *Foundations of the Theory of Probability*. Chelsea Publishing Company, New York. English translation of: Grundbegriffe der Wahrscheinlichkeitsrechnung, Berlin, 1933.

Kolmogorov, A. N., 1991. Entry: Probability, *mathematical*. In: *Matematicheskaia entsiklopediia. English. Encyclopedia of Mathematics*. Kluwer Academic Publishers, Dordrecht. English translation of: Matematicheskaya Enziklopedia, vol. 1, Sovietskaya Enziklopedia, Moscow, 1977.

Kuban, R. and MacKenzie-Carey, H., 2001. Community-Wide Vulnerability and Capacity Assessment (CVCA). Government of Canada; Office of Critical Infrastructure Protection and Emergency Preparedness. Catalogue No.: D-82-68/2002E-IN; ISBN: 0-662-32220-7.

Kyburg, H. E. and Smokler, H. E., 1980. *Studies in Subjective Probability*. Robert E. Krieger Publishing Company, Huntington, New York, second edition. Original edition 1964.

Landhuis, E., 2004. Lifelong debunker takes on arbiter of neutral choices. Magician-turned-mathematician uncovers bias in a flip of a coin. Stanford Report, June 7, 2004. URL `http://news-service.stanford.edu/news/2004/june9/diaconis-69.html`

Lempert, R., Nakićenović, N., Sarewitz, D., and Schlesinger, M. E., 2004. Characterizing climate change uncertainties for decision-makers. *Climatic Change*, 65, pp. 1–9.

Liverman, D., 1990. Vulnerability to global environmental change. In: R. Kasperson, K. Dow, D. Golding, and J. Kasperson (eds.), *Understanding Global Environmental Change: The Contributions of Risk Analysis and Management*, pp. 27–44. Clark University, Worcester, MA.

Loeve, M., 1977. *Probability Theory*. Number 45 in Graduate Texts in Mathematics. Springer, New York, fourth edition.

Luers, A. L., 2005. The surface of vulnerability: an analytical framework for examining environmental change. *Global Environmental Change*, 15(3).

Luers, A. L., Lobell, D. B., Sklar, L. S., Addams, C. L., and Matson, P. A., 2003. A method for quantifying vulnerability, applied to the agricultural system of the Yaqui Valley, Mexico. *Global Environmental Change*, 13(4), pp. 255–267.

Maĭstrov, L. E., 1974. *Probability Theory. A historical sketch*. Academic Press, New York and London. English translation of: Teoriiá Veroiátnosteĭ Istoricheskiĭ Ocherk, published by Izdatel'stvo Nauka, Moscow, 1967. Translated and edited by Samuel Kotz.

Markowich, P. A., 2007. *Applied Partial Differential Equations: A Visual Approach.* Springer, Berlin.

McCarthy, J. J., Canziani, O. F., Leary, N. A., Dokken, D. J., and White, K. S. (eds.), 2001. *Climate Change 2001: Impacts, Adaptation and Vulnerability.* Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.

Metzger, M. J. and Schröter, D., 2006. Towards a spatially explicit and quantitative vulnerability assessment of environmental change in Europe. *Regional Environmental Change*, 6(4), pp. 201–216.

Moench, M. and Dixit, A. (eds.), 2007. *Working with the Winds of Change. Toward Strategies for Responding to the Risks Associated with Climate Change and other Hazards.* ProVention Consortium; Institute for Social and Environmental Transition–International; Institute for Social and Environmental Transition–Nepal, second edition.

Morgan, M. G. and Henrion, M., 1990. *Uncertainty. A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis.* Cambridge University Press, New York, Port Chester, Melbourne, Sydney.

Murakami, H., 2009. *What I talk about when I talk about running.* Vintage Books, London.

Nakićenović, N., Davidson, O., Davis, G., Grübler, A., Kram, T., Rovere, E. L. L., Metz, B., Morita, T., Pepper, W., Pitcher, H., Sankovski, A., Shukla, P., Swart, R., and Dadi, R. W. Z., 2000. IPCC Special Report. Emissions Scenarios. Summary for Policymakers. Published for the Intergovernmental Panel on Climate Change. ISBN: 92-9169-113-5. URL `http://www.ipcc.ch/pdf/special-reports/spm/sres-en.pdf`

Nakićenović, N. and Swart, R. (eds.), 2000. *Special Report on Emissions Scenarios.* Intergovernmental Panel on Climate Change Special Report. Cambridge University Press.

Newell, B., Crumbley, C. L., Hassan, N., Lambin, E. F., Pahl-Wostl, C., Underdal, A., and Wasson, R., 2005. A conceptual template for integrative human-environment research. *Global Environmental Change*, 15(4), pp. 299–307.

Neyman, J. (ed.), 1950. *First Course in Probability and Statistics.* Henry Holt, New York.

O'Brien, K., Eriksen, S., Nygaard, L. P., and Schjolden, A., 2007. Why different interpretations of vulnerability matter in climate change discourses. *Climate Policy*, 7(1), pp. 73–88.

O'Brien, K., Leichenko, R., Kelkar, U., Venema, H., Aandahl, G., Tompkins, H., Javed, A., Bhadwal, S., Barg, S., Nygaard, L., and West, J., 2004. Mapping vulnerability to multiple stressors: Climate change and globalization in India. *Global Environmental Change*, 14(4), pp. 303–313.

Onions, C. T. (ed.), 1959. *The Oxford Universal Dictionary.* Oxford University Press, Oxford, (revised) third edition.

Parry, M., Canziani, O., Palutikof, J., van der Linden, P., and Hanson, C. (eds.), 2007. *Climate Change 2007: Impacts, Adaptation and Vulnerability.* Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge.

Pielke, R. A. J., 1999. Nine Fallacies of Floods. *Climatic Change*, 42, pp. 413–438.

Pittock, 2001. Probabilities will help us plan for climate change. *Nature*, 413, p. 249.

Polya, G., 2004. *How to solve it: a new aspect of mathematical method.* Princeton science library. Princeton University Press. Original edition 1954.

Purves, R. A. and Sudderth, W. D., 1976. Some finitely additive probability. *The Annals of Probability*, 4(2), pp. 259–276.

Ramakrishnan, S., 1980. *Finitely additive Markov chains.* Ph.D. thesis, Indian Statistical Institute.

Ramakrishnan, S., 1981. Finitely additive Markov chains. *Transactions of the American Mathematical Society*, 265, pp. 247–272.

Ramsey, F. P., 1926. Truth and Probability. In: R. B. Braithwaite (ed.), *Foundations of Mathematics and other Essays*, pp. 156–198. Routledge & P. Kegan. Reprinted in Studies in Subjective Probability, H. E. Kyburg, Jr. and H. E. Smokler (eds.), 2nd ed., R. E. Krieger Publishing Company, 1980, 23–52.

Red Cross/Red Crescent, 2008. Climate Guide.
URL `http://www.climatecentre.org/`

Rényi, A., 1964. Mathematics, a Socratic dialogue. *Physics Today*, 24, pp. 24–36.

Rényi, A., 1965. Dialogue on the applications of Mathematics. *Simon Stevin*, 39, pp. 3–17.

Rényi, A., 1967. *Dialogues on Mathematics.* Holden Day Publishers, San Francisco.

Rényi, A., 1969. *Briefe über die Wahrscheinlichkeit.* Birkhäuser Verlag, Basel und Stuttgart.

Rényi, A., 1970. *Foundations of Probability.* Holden-Day, Inc., San Francisco.

Rockafellar, R., 1970. *Convex analysis.* Princeton University Press, Princeton, NJ.

Salmon, W., 1966. *The Foundations of Scientific Inference.* University of Pittsburgh Press, Pittsburgh.

Sarukkai, S., 2001. Mathematics, Language and Translation. *Meta*, XLVI(4), pp. 664–674.

Sarukkai, S., 2003. Perspectives on Mathematics. *Economic and Political Weekly*, 38(35), pp. 3648–3649.

Sarukkai, S., 2005. Revisiting the 'unreasonable effectiveness' of mathematics. *Current Science*, 88(3), pp. 415–423.

Savage, L. J., 1954. *The Foundations of Statistics.* Advanced Textbooks in Economics. John Wiley & Sons, New York.

Schneider, S. H., 2001. What is 'dangerous' climate change? *Nature*, 411, pp. 17–19.

Schneider, S. H., 2002. Can we estimate the likelihood of climatic changes at 2100? *Climatic Change*, 52, pp. 414–451.

Seidenfeld, T., 2001. Remarks on the theory of conditional probability: some issues of finite versus countable additivity. PhilSciArchive.
URL `http://philsci-archive.pitt.edu/archive/00000092/`

Seidenfeld, T., Shervish, M., and Kadane, J., 1990. When fair betting odds are not degrees of belief. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, Vol. 1990 (Volume One: Contributed Papers), pp. 517–524.

Shafer, G., 1996. The significance of Jacob Bernoulli's Ars Conjectandi for the philosophy of probability today. *Journal of Econometrics*, 75, pp. 15–32. Prepublicated version.
URL `http://www.glennshafer.com/assets/downloads/articles/article55.pdf`

Šikić, H., 1991. Riemann Integral vs. Lebesgue Integral. *Real Analysis Exchange*, 17, pp. 622–632.

Sill, H.-D., 1993. Zum Zufallsbegriff in der stochastischen Allgemeinbildung. *Zentralblatt für Didaktik der Mathematik*, 25, pp. 84–88.

Singh, S., 1997. *Fermat's Last Theorem.* Fourth Estate, London.

Soanes, C. and Stevenson, A. (eds.), 2005. *Oxford Dictionary of English.* Oxford University Press, (revised) second edition.

Spizzichino, F., 2009. Surprising Coincidences and Some Misunderstandings About "Rare" Events. In: M. Emmer (ed.), *Mathematics and Culture VI*, pp. 99–120. Springer, Berlin, Heidelberg.

Suppes, P., 1968. The Desirability of Formalization in Science. *The Journal of Philosophy*, 65(20), pp. 651–664.

Thywissen, K., 2006. Components of Risk, A Comparative Glossary. *SOURCE - Studies Of the University: Research, Counsel, Education*, 2.

Timmerman, P., 1981. *Vulnerability, Resilience and the Collapse of Society. A Review of Models and Possible Climatic Applications.* Number 1 in Environmental Monograph. Institute for Environmental Studies, University of Toronto, Toronto.

Todesco, G. M., 2009. Mathematics and Animated Cartoons. In: M. Emmer (ed.), *Mathematics and Culture VI*, pp. 159–170. Springer, Berlin, Heidelberg.

Todhunter, I. (ed.), 1949. *A History of the Mathematical Theory of Probability from the Time of Pascal to that of Laplace.* Chelsea Publishing Company. Reprint of the 1865 edition.

Turner II, B. L., Kasperson, R. E., Matson, P. A., McCarthy, J. J., Corell, R. W., Christensen, L., Eckley, N., Kasperson, J. X., Luers, A., Martello, M. L., Polsky, C., Pulsipher, A., and Schiller, A., 2003. A framework for vulnerability analysis in sustainability science. *Proceedings of the National Academy of Sciences of the United States of America*, 100(14), pp. 8074–8079.

Twigg, J., 2004. Disaster Risk Reduction - Mitigation and preparedness in development and emergency programming. *Good Practice Review*, 9.

Ulam, S., 1930. Zur Masstheorie in der allgemeinen Mengenlehre. *Fundamenta Mathematicae*, 16, pp. 140–150.

Ulam, S., 1943. What is Measure? *The American Mathematical Monthly*, 50(10), pp. 597–602.

UNDP, 2004. Reducing Disaster Risk. A Challenge for Development. A Global Report.

UNDRO, 1982. Shelter after Disaster: Guidelines for Assistance.

UNFCCC, 1992. UNITED NATIONS FRAMEWORK CONVENTION ON CLIMATE CHANGE. FCCC/INFORMAL/84 GE.05-62220 (E) 200705.
URL `http://unfccc.int/resource/docs/convkp/conveng.pdf`

UN/ISDR, 2007. Terminology of disaster risk reduction. Visited Sep. 28, 2007.
URL `http://www.unisdr.org/eng/library/lib-terminology-eng home.htm`

Weichselberger, K., 2001. *Elementare Grundbegriffe einer allgemeineren Wahrscheinlichkeitsrechnung I.* Physica Verlag, Heidelberg.

Williams, D., 1991. *Probability with Martingales.* Cambridge University Press, Cambridge.

Wolf, S., 2009. Vulnerability to Climate Change - Mathematics as a Language to Clarify Concepts. In: M. Emmer and A. Quarteroni (eds.), *MATHKNOW2008*, pp. 253–263. Springer, Milano.

Wolf, S., Ionescu, C., Lincke, D., Bisaro, S., Hinkel, J., and Reckien, D., 2007. A Formal Framework of Vulnerability. Deliverable to the ADAM Project. *FAVAIA Working Paper 6, Potsdam Institute for Climate Impact Research.*
URL `http://www.pik-potsdam.de/favaia/pubs/favaiaworkingpaper6.pdf`

Wolf, S., Lincke, D., Hinkel, J., Ionescu, C., and Bisaro, S., 2008. Concept Clarification and Computational Tools – A Formal Framework of Vulnerability. *FAVAIA Working Paper 8, Potsdam Institute for Climate Impact Research.*
URL `http://www.pik-potsdam.de/favaia/pubs/favaiaworkingpaper8.pdf`

Zusammenfassung

Die vorliegende Arbeit gliedert sich in zwei Teile. Diese sind dadurch verbunden, dass der erste Teil die mathematischen Fragen, die im zweiten Teil behandelt werden, motiviert.

Gegenstand des ersten Teils sind die Begriffe Vulnerabilität und Wahrscheinlichkeit sowie die Methode der Formalisierung. Formalisierung wird hier als Übersetzung von Begriffen aus der Alltagssprache oder aus wissenschaftlichen Fachsprachen in mathematische Begriffe aufgefasst. Das Ergebnis einer Formalisierung ist ein mathematisches Modell des jeweiligen Begriffs. Ein Vorteil mathematischer Begriffe ist, dass diese keine impliziten Bedeutungen haben.

Als Beispiel für Formalisierung wurde der Begriff Wahrscheinlichkeit mit seinen verschiedenen Interpretationen und mathematischen Modellen betrachtet. Die subjektive Interpretation ist für die Klimafolgenforschung interessant, da sie sich im Gegensatz zur frequentistischen Interpretation auch anwenden lässt, wenn keine langen Messreihen von Daten zur Verfügung stehen. Diese Interpretation wird besonders im Modell der kohärenten Wahrscheinlichkeitsbewertungen von de Finetti dargestellt. Die verschiedenen Modelle von Wahrscheinlichkeit verhalten sich unterschiedlich zum Begriff selbst: de Finettis Modell erklärt den Begriff und gibt eine Messvorschrift; Kolmogorovs Modell ist vorrangig Basis für die mathematische Theorie, ohne genauer auf Bedeutung und Messung einzugehen. Insgesamt wurde deutlich, dass Formalisierung kein Allheilmittel bei begrifflicher Unklarheit darstellt aber zur Klärung beitragen kann.

Solch eine Klärung war ein Ziel einer Formalisierung von Vulnerabilität und verwandten Begriffen, die am Potsdam-Institut für Klimafolgenforschung entwickelt wurde. Diese Arbeit stellt die Formalisierung vor und verwendet sie zur Begriffsklärung in einer Fachterminologie, für die die Literatur eine Babylonische Verwirrung konstatiert. Es existieren viele ähnliche theoretische Definitionen, aber verschiedene Interpretationen und Arten der Messung von Vulnerabilität. Mit Hilfe genauer und trotzdem allgemeiner mathematischer Definitionen von Vulnerabilititität wurden die gemeinsame Struktur vieler Definitionen aus der Literatur nachgewiesen und zwei Typen von Vulnerabilitätsassessments unterschieden. Mehrere Paare von Interpretationen des Begriffs aus der Literatur, sowie Typen von Risikoassessments aus dem Bereich der Naturgefahren, wurden auf die Unterscheidung zwischen diesen Typen zurückgeführt. Die begriffliche Verwirrung wurde unter anderem dadurch erklärt, dass zwischen Fachbegriffen und Arten von Messungen eine mehr-mehr-deutige Beziehung besteht. Die Idee, die Mathematik als Lingua franca für Forscher verschiedener Disziplinen zu verwenden, zeigte sich auf dem sozialwissenschaftlich geprägten Gebiet der Vulnerabilität nur begrenzt umsetzbar. Durch Diagramme wurde versucht, die Formalisierung für Nichtmathematiker leichter zugänglich zu machen.

Der erste Teil dieser Arbeit betrachtet die Mathematik als Sprache, diskutiert Formalisierung als Übersetzung in Mathematik (anhand des Begriffs Wahrscheinlichkeit), übersetzt Vulnerabilität in diese Sprache und schafft dadurch Klarheit. Schließlich diskutiert er wie man Mathematik allgemeinverträglich präsentieren kann. Damit ist dieser erste Teil eine Arbeit über Mathematik.

Der zweite Teil hingegen ist eine mathematische Arbeit im eigentlichen Sinne: er verbindet Elemente aus der Kategorientheorie und der Wahrscheinlichkeitstheorie. Das mathematische Vulnerabilitäts-Modell von Ionescu [2009] verwendet die kategorientheoretischen Begriffe des Funktors und der Monade zur allgemeinen Beschreibung der ungewissen Entwicklung eines Systems. Ungewissheit wird mathematisch unter anderem durch Wahrscheinlichkeit beschrieben. Für Wahrscheinlichkeitsmaße im Sinne von Kolmogorovs Axiomen hat Giry [1981] nachgewiesen, dass diese Funktor und Monade bilden. Hier wurde gezeigt, dass auch die allgemeineren mathematischen Modelle der endlich additiven Wahrscheinlichkeit und der kohärenten Wahrscheinlichkeitsbewertungen Funktoren über der Kategorie von Mengen und Funktionen sind. Für endlich additive Wahrscheinlichkeit wurden außerdem mehrere Monaden identifiziert. Dies bildet einen ersten Schritt zur Anwendbarkeit der allgemeineren Modelle (subjektiver) Wahrscheinlichkeit auf dem Gebiet der Vulnerabilität.

Erklärung

Ich versichere, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 31.07.2009 Sarah Wolf

Acknowledgements – Dank

I would like to thank Prof. Rupert Klein for providing supervision of this thesis in the very sense of the word: while I was free to work on what I found interesting, when I lacked overview, our conversations helped me get a lot further.

I thank my colleagues from the FAVAIA group at PIK for inspiring work on vulnerability and formalization in a nice atmosphere: Sandy Bisaro, Jochen Hinkel, Mareen Hofmann, Cezar Ionescu, Richard Klein, and Daniel Lincke. Further thanks go to the same people for different things. Cezar has closely watched and commented the emergence of the monads presented in this work. Jochen has provided many useful hints for writing its first part. Daniel helped with computer problems of all kinds, Mareen and Sandy were partners for interesting talks on mathematics, on vulnerability and on much else. Further, Mareen, Jochen, Daniel and Sandy have read and commented an earlier draft of this work. Thank you all very much for this.

My views on probability profited greatly from work in two different groups at PIK: a seminar on De Finetti's Theorem together with Daniela Eskelson, Markus Fucik, Armin Haas, Carlo Jaeger, Jette Krause and Rainald Ötsch, as well as sessions of the Cartesian Seminar on coherent conditional probability together with Nicola Botta, Steffen Fürst, Mareen Hofmann, Cezar Ionescu, Carlo Jaeger, Daniel Lincke and Antoine Mandel. Thanks for helping me explore a whole new probability.

Ich danke meinen Eltern Susanne Hanslik und Helge Wolf für sämtliche Unterstützung, die auf alle möglichen Arten auch zu dieser Arbeit geführt hat. Ganz besonders für die Vererbung der Familienkrankheit Mathematik. Und für den Spaß daran, zum Beispiel die Schwertkämpfer in der Analysis. An diese sei beim Durchblättern dieser Arbeit erinnert. Meinem Bruder Tonio Danke!Danke!Danke! für die Tipps zum Austauschen von ertkämpfern durch ärme. Meinen Schwestern: Lelia, danke für die kleinen Ablenkungen zwischendurch, Elgin, ohne weitere Worte, für einfach alles. Grazie di tutto (!!) anche ad Andrea Buonaguidi. Nicht zuletzt danke ich allen Freunden, die immer wieder zugehört haben, mitdiskutiert, etwas gelesen, und, und, und. Besonders erwähnt seien hier Lutz Faulhaber, Kirsten Hilgenböcker, Maria Remesat und Christine Twittmann.